

New and changing research publication practices due to open access publication initiatives

Erik Sandewall

**Professor of Computer Science, Linköping University
And Chairperson, ICSU Committee on the Dissemination of
Scientific Information**

Introduction

Modern information technology, and in particular the WWW (World-Wide Web) technology has spurred the development of open access publication, where research publications are made freely available on the Internet. The reasons for these initiatives are usually stated in economic terms and with reference to how costs are split between the original producer of a document (the author or her institution), the middleman (the publisher), and the consumer, as follows:

- much of the traditional costs for the middleman are automatically transferred to the consumer, namely those for computers, paper and printing, as electronic copies of articles are read on the computer screen and sometimes (not always) printed on the computer printer;
- other traditional middleman costs have already been transferred to the producer, namely those for typesetting as authors are required to submit their manuscripts in computer-based form;
- yet other traditional middleman services and associated costs are being eliminated by many publishers, namely the service of language checking and correction;
- additional middleman costs are becoming split between all three parties as distribution of paper copies of journal issues by ordinary mail is replaced by electronic distribution over the Internet.

It is then argued that the remaining services and costs of the middleman are so marginal that it is not worthwhile to charge for them using e.g. subscriptions, because of all the hassle with collecting subscriptions and restricting access so that only subscribers can obtain the publications. It is better, it is argued, that articles are made accessible free of charge and the remaining 'middleman' costs are covered either by the producer through institutional repositories, or by national or international repositories that are financed by research funding agencies.

Objections against this line of reasoning are of two kinds. Some object to the economic analysis, and argue that the actual middleman costs are grossly underestimated in the cited analysis, and that the proposed new arrangement is not economically viable in the long term, in particular if it is supposed to scale up to the full volume of research publication today.

Other objections refer to the *effects* of the open access publication on the current system for scientific publication as a means of communication between researchers. It is argued that open access is likely to erode the current quality control scheme which is based on peer review as well as quality ranking of journals using impact factors and other bibliometric tools.

In this memorandum I will address the question of scientific communication in a broader perspective. Instead of assuming that the present system is the best of all possible, I will argue that the present system is itself determined by the available technologies during earlier periods of time, and that there are good reasons to change it for something better now that modern information technology makes those changes possible.

This memo is based on a presentation that was made at the ICSU conference in March, 2003 in Paris, and in particular it is based on the powerpoint slides for that presentation. Consequently some of the text is fairly concise and often written as item lists. I hope this will not be seen as a problem by the readers.

1. A summary of the rules of the present 'publication game'

- Quality control (peer review) of publications
- Highly inappropriate to publish the ideas or results of another researcher
- A journal will not republish a previously published article
- Correct reference shall be made to the first report of a result that the present work builds on
- Priority of results is important in evaluation the performance of a researcher
- Reviewers are anonymous
- Reviewers shall only evaluate the objective quality of an article, not guard their own interests
- Articles are considered confidential while reviewed
- Priority is counted from the day of publication of an article, after review and acceptance
- There are also conventions for author assignment in group efforts, or between student and principal advisor

This system of rules is fairly complex, widely accepted, and important for the proper functioning of the system. Notice, however, that these rules are *not* based on legal systems, nor on economic considerations; they are by and large a combination of social rules in the scientific community, and rulings by scientific organizations including journal editorial boards.

The *purpose* of establishing and maintaining those rules are also clear:

- Efficient dissemination and preservation of scientific information
- Provide efficient and fair incentives for researchers and research groups

2. Technology dependence

Contrary to what is often claimed, the contemporary publication scheme as summarized above is not based on centuries of traditions. Journal based on international peer review started to appear only in the mid-twentieth century. Notice that:

- Peer review by peers located anywhere in the world, and not only those that are present at the site of the publisher, requires that several copies of the manuscript are available.

This became possible with the advent of typewriters and carbon paper, but it was not possible before.

- In the technology of the mid-twentieth century, there was effectively only two ways of preparing copies of a manuscript: typewriter and carbon paper (a few copies), and typesetting in lead and printing (expensive). This directly implies the use of peer review in order to decide which articles are worth the high cost of printing.

3. Reasons for reconsidering the rules of the publication game

We are no longer restricted by technology in this way, and we ought to ask the question which publication and communication schemes are optimal given the present technology. The following sections discuss a number of reasons why it may be worthwhile to reconsider the rules and conventions listed in section 1.

A. Maybe rules that worked well in the previous era lead to bizarre consequences when applied for modern information technology.

Example: The continued insistence on the rule “do not publish previously published results” has led to (1) not using preprints, or (2) not considering preprints as having been “published”.

In fact, the following sequence of events illustrate how quickly a rule may change from reasonable to obsolete:

- The New England Journal of Medicine ruled in the early 1950's that a previously published article could not be republished in their journal. This rule was quickly followed by other journals.

- Easy reproduction of 'preprints' and 'departmental technical reports' became technically feasible in the early 1960's.

- In order to accommodate the existing journal rule with the new technology, it was ruled that a preprint was not to be considered as 'published', a use of language that contradicts common sense and that would not be accepted by, for example, a patent bureau

- Today, by extrapolation of the same course of events, an article may be available on the Internet and accessible for everyone by the click of the mouse, but it is anyway not considered as 'published' (in some disciplines) if it has not been peer reviewed. Other disciplines consider it as published, therefore not acceptable to a journal, and thereby the effective use of the new medium is thwarted.

The technology dependence of the original rule is a reason to reconsider its relevance.

B. Maybe there is undesirable behavior that was not possible before but which has become possible with the new technology.

Example: Removing published articles from electronic access because their contents are considered objectionable by some groups (This has already happened).

C. Maybe the continued application of established rules prevents what would otherwise have been a good use of new technology.

Example: Open discussion with the peer community as part of the reviewing process is not compatible with the idea that priority counts from the date of publication and publication occurs after reviewing. (Some experimental journals, such as the Electronic Transactions on Artificial Intelligence, ETAI, have reversed this procedure so that an article is first published, then peer reviewed by open discussion, and finally included in the journal if it passes the quality test).

D. Maybe the good use of new technology requires additional rules or generalization of existing rules.

Example: In some disciplines, a research result is primarily represented as a contribution to a data base. How is the identity of the originator of the result represented in such databases? How is reference to earlier work represented? Should it be? If not, how will the scientific community assign credit to such work?

E. Maybe new or extended rules are needed in order to strengthen the negotiating position of the scientific community visavis its commercial partners

In the area of scientific publication, the scientific community as a whole is 'outsourcing' a large part of its internal communication to commercial partners. This is becoming increasingly expensive, and we need to deal with the problem.

Consider for example the following scenario: the editorial board of journal J is dissatisfied with the pricing policies of its commercial publisher, and decides to set up a new journal N with a new publisher. The old publisher recruits an entirely new editorial board for J.

Question: shall the current impact factor and perceived prestige of J then belong to J or to N?

Note: the assumption that it stays with J is one of the reasons why prestigious journals are not very price-sensitive.

F. Maybe reduced outsourcing and more do-it-yourself publication necessitates additional rules.

Example: One possible development is that universities and institutes publish their own work, and that 'journals' are defined as selections from the literature ('overlay journals'). Then there must be credible safeguards against the possibility that an author makes changes retroactively into his/her previously published works. This requires an awareness in the scientific community about what are adequate safeguards.

G. Maybe new technology destroys the basis for some existing, useful activities, and new incentives are necessary in order to obtain replacements.

Example (this example was actually mentioned in a previous talk at the ICSU conference): If it is no longer a viable business to publish and sell tables of integrals for a particular community of scientists, that community has to assign professional credit to researchers that do that job instead.

Maybe creating digests of earlier results should be considered as valuable as creating 'original new results' ?

4. Is it possible to change the rules of the publication game

I have now reviewed a number of reasons why it may be worthwhile for the scientific community to review and to change the rules of the publication game. The immediate question is - is this possible? Well, by definition any change must happen in the same way as those rules were formed in the first place, that is, as a combination of social rules in the scientific community, and formal rulings by professional societies or journal editorial boards. What is needed at this point is therefore:

- Groups that bring these issues (short-term and long-term) to the attention of the scientific community and ignite the discussion.
- Broad discussion among researchers in various fields (different disciplines have different needs and different practices).

- Policy recommendations by a major organization. It would be appropriate that ICSU did this.

- Policy decisions by journals, universities, funding agencies, academies, etc.

Notice that all these changes must take a long time perspective. In particular, the key reason why the present system is so expensive for the research community is that price competition does not work because the author, who chooses which journal he is going to submit an article to, does not carry the subscription costs that are implied by his choice of that journal. However, authors choose journals partly based on their expectation of what will be the most valuable for them in their continued career, which means a time perspective of 20 years or more.

Proposed changes of priority must be credible over that timescale if they are going to actually influence author behavior, assuming of course that the author continues to have a free choice about which journal to submit to.

5. Long-term direction: Research Knowledge Management

The example in section 3 showed how a rule that was originally formulated in the context of one technology, can rapidly become obsolete due to technological change. If the scientific community is going to change the mechanisms of its internal communication system, then it is important not to fall into the same trap again. An analysis of likely and desirable directions on the scale of decades into the future is important, while also of course very difficult.

I propose that the natural direction is towards Knowledge Management systems, using a term that is popular in industry, but now knowledge management for the totality of knowledge that is built up in research enterprises. In principle, this includes:

- conventional research publications
- research background material: laboratory notes from experiments and observations; software documentation, etc
- sources, for example documents in historical archives
- research administrative documents: project proposals, formal reports from research projects
- structured data that express research results, contributions to databases
- presentations of research results using alternative media or multimedia: audio recordings, video recordings, visualizations in virtual reality
- pedagogical presentations of research results, for example 'powerpoint' presentations at conferences and seminars
- discussions about research work or about research results, for example as e-mail communication between researchers

At present these different kinds of information are maintained and archived (some of them, sometimes) in very different ways.

There are good arguments for taking an integrated view of this entire information structure. Doing so is a relatively long-term undertaking, but it has to be done because it is better that the scientific community discusses this explicitly, than that we should slide in this direction without proper analysis of the changes in information handling and the effects of those changes on the research communication system.