



Electronic journals: a user's experience

Michelangelo Mangano (*)

Abstract

I collect here some thoughts on the value and usability of the electronic versions of commercial Physics Journals. These critical remarks are the result of my own experience with their use, and were provoked by the sudden disappearance of several important journals from the CERN library. These thoughts are intended to promote a closer interaction between publishers and users, in order to ensure that user's needs are fulfilled before the complete (and irreversible!) transition to a fully automated journal.

The First Shock

On a hot afternoon of August 1999 I retreat to the cool underground sector of CERN's library, to consult an issue of Physical Review D (PRD). On my way downstairs, I pay almost no attention to an apparently innocent sign posted on the wall, indicating that as of the beginning of August all issues after 1985 of Phys. Rev. and Phys. Rev. Letters (PRL) have been archived, as they are available from the Web site of the American Physical Society. Evidently my brain is not ready to receive this message, and presumably decodes its contents in a different way, as I continue my walk to the location where I confidently expect to find the PRD volumes. The surprise of not finding them there turns rapidly into anger, as I finally realize the meaning of the message upstairs. The following 20 minutes were spent in frustration, walking back upstairs to find an available terminal in the library, surfing my way through the PRD E-archives, trying (unsuccessfully) to print the paper on the local printer, browsing with mouse clicks through the paper in search of the relevant sections, and finally finding out that what I really needed was to be found in a Zeitschrift für Physik article quoted in the references. This implied logging out of the terminal and walking back downstairs, at the risk of having to repeat the whole exercise once more should the Zeitschrift für Physik reference have quoted some other interesting paper from PRD or PRL, as it surely did. A journal consultation which 2 weeks before would have taken 3 minutes, ended up taking over 20. Is this progress?

After that frustrating experience, I felt as if my car's maker had recalled all vehicles to replace the steering wheel with a mouse: click on the left button to steer left, click on the right one to go right. Memories of long afternoons spent in the library of my University surfaced: as a student, one of my most rewarding experiences was to randomly browse through old issues of physics journals, searching for to-me-unknown, but possibly interesting, articles. I then felt sorry for the future generations of students, who may not be given the pleasure of skimming through a full volume of PRD while holding it in their hands. How much time will be wasted consulting issues electronically, waiting for the PDF file to pop up so that we can gauge whether the promise issued in the title or in the abstract is held?

In spite of my own personal frustration, it is quite clear even to me that the electronic media revolution will affect publishing strategies, the way people access scientific information, and the way libraries function. What is, however, less clear to me is whether the changes we are seeing right now in the commercial publishing arena are improving our way of working, and justify an immediate transition away from the library "*as we know it*". It is important that library managers don't get tempted by the potential for change which the new technologies open to them, and maintain contact with the users' needs, to ensure that the transition to the "*virtual*" library will fulfil everybody's dreams.

Potential Benefits

Neglecting the advantages open to the library staff and administrators, let me consider here the main potential benefits for the users of a virtual library:

1. Scientists working in Institutes with limited resources and poor libraries may in principle have access to the same material as scientists working in richer environments.
2. Electronic access to journal issues will allow us to consult articles from home or while travelling, connecting our laptops from the lounges of airports or from hotels.
3. "Active" documents, namely articles with internal links from text to equations, and external links from bibliographic items to their electronic counterpart, may make electronic consultation competitive with a live session in the library.

In order to access the electronic version of physics journals, physicists must use validated computer accounts. These are available only to researchers whose library subscribes to the hard-copy version of the journal. Scientists from poorer institutions gain nothing from this situation, as their libraries cannot afford the subscription to the hard copies anyway. There is no indication that this will change in the future, or that publishing companies will significantly reduce their subscription fees allowing for electronic-only subscriptions, making them affordable to everybody. I expect an increased pressure on the international laboratories such as CERN to grant computer accounts to people from remote Institutes, to allow them to gain access to these resources. It is not clear to me whether this demand can be met.

Also point 2 is only partially addressed by the current developments. Physicists from subscribing Institutions may not have access to the electronic archives while travelling, or working from home, if using commercial Internet providers. Using their home-institution account may be impractical in certain circumstances.

Point 3 is still a long way from being realized. Internal links within the text still do not exist. Links to all quoted papers require all quoted journals to have web-based archives, and this is clearly not the case. Many consultations of the e-versions of articles still require trips to the library, and this is particularly true of the articles published before the advent of the Los Alamos archives (as I mentioned above, PR and PRL are available in e-versions since 1985). As for the articles which have appeared in recent years, electronic access to references can be done directly and very quickly through the Los Alamos archive and Spire, completely by passing the publishers' electronic archives.

I should perhaps notice that points 1 and 2 are actually solved by the fully-electronic journal JHEP (Journal of High Energy Physics), managed and produced by SISSA. The presence of an outstanding editorial board, and of a serious refereeing process, ensures a quality control equal to

that of the commercial journals. On the other hand, the absence of subscription fees makes it universally available. The implications of electronic-only journals go well beyond the subject of this note, which is limited to the easyness of access and of use, and I chose not to let them divert my main focus here.

One of the difficulties in making scientific papers easily readable on a computer screen is related to the way we read them. Very rarely does one go through them sequentially, line after line and page after page, as one would do with a normal written text. Forward and backward referencing to text, tables, figures and equations is very frequent. Several pages at once need to be under our eyes to compare equations, to follow the development of a calculation across pages, to compare the contents of different tables or figures. In spite of all bookmarking tools made available by the current software, nothing matches the power of our fingers to flip through real pages back and forth and to keep track of where we are. It will still be sometime before the temptation to send to the printer every article popping upon the computer screen will disappear. Until then, an enormous amount of paper is wasted every day. The time taken by an average printer to print e-versions of journal articles is also a bothersome limitation, at this stage.

With the much-reduced burden of preparing papers for publication, given that most of the typesetting is done by the authors, publishing companies should at least earn their fees by producing more versatile electronic versions of the articles. To the occasional user, even locating the reference is an unnerving multi-step operation. If one just logs on to a library computer with as only reference the standard volume number, year, and page, it will be a while before arriving at the paper, especially if useful bookmarks are lacking. In the case of papers stored in the Los Alamos archives, access is instead immediate, as there is a standard, easy-to-remember URL, to which it is sufficient to add the document number, given in a fixed, standard format directly related to the reference itself (e.g. <http://arXiv.org/ps/hep-ph/0008001> for the first paper of September 2000 in the hep-ph archive).

The Reality

While preparing this article, I made some experiments from a terminal in the CERN library. I tried to access papers for which I had a precise reference, and I tried to follow a search path for works of which I knew only the author name and the journal. I started by using the tool suggested by the CERN library, a search engine which points directly to a paper provided the exact reference is known (http://library.cern.ch/electronic_journals/ej.html). This turned out to be very effective in quickly reaching the documents. The problem came with looking at the documents. In the case of Nucl. Phys., B : 537 (1999) 443, it took 65 seconds for the 15-page document to appear. In the case of Nucl. Phys., B : 485 (1997) 291 I quit after a minute, with my browser indicating that 17 minutes and 29 seconds were still remaining to download the document (130 pages, admittedly long; however, when I look in a journal the time it takes to find the paper in the issue usually gets shorter the longer the paper!).

For comparison, downloading the postscript version of this same paper from the Los Alamos archives (<http://arXiv.org/ps/hep-ph/9605323>) took only 25 seconds, which is acceptable.

As a final exercise, I tried the PRD server, accessing Phys. Rev., D : 50 (1995) 2966 (the "Top quark evidence" paper from CDF, a classic of recent experimental literature). When the PDF file started being downloaded, I was warned that it would be another 1 hour and 30 minutes before I could see it on the screen (5.7 MByte document, for 60 pages of text). Unfortunately this paper was not submitted to the Los Alamos archives, and the link to the CDF preprint server (available through Spire) does not work. So if this volume of PRD is removed from the library, the only way to get the paper (perhaps to just check a table, a number, or a reference) is to wait 1 hour and 30 minutes for the PDF image to pop up on the screen.

Is this progress?

In the case of searches when the precise reference is not available, things get even worse. I tried the Nucl. Phys. B server.

I started from the Elsevier server, clicked on Nucl Phys B, on "Published" and then on "search". Each of these clicks took 25 to 35 seconds to move to the following windows. When I got to the "search" page, I discovered that no Boolean author search is possible (so if your name is Wang or Xu, forget about it). It took over 40 seconds to return the papers published under my name. Similar searches done using Spire take between 2 and 10 seconds, they have no overhead, they allow searches across all journals and preprint archives, and allow rather detailed Boolean searches involving name, title, date range, etc.

I repeated this exercise, with the help of the Library staff, over the following days and at different times. In the case of the PRD server, we noticed similar performances during the day times, but much faster response after 7:30pm (CERN time). Phys. Rev., D : 50 (1995) 2966 was downloaded in about 5 minutes at this time of the day. In the case of Nuclear Physics, the behaviour was more random, with peak performances of one minute to download Nucl. Phys., B : 485 (1997) 291.

Suggestions for Improvements

Aside from the obvious advice to invest in better servers and better connection lines, here are just some suggestions for features which would make the use of the e-versions of published articles competitive with the hard copies, and with the versions available from the LANL arXiv:

1. Internal links to equations, tables, figures. The links would ideally generate a pop-up window with the required information, keeping the page from which the link originate available.
2. Possibility to repaginate the document (e.g. to collect a series of equations or a set of figures into a single page) or to easily and quickly display multiple pages at the same time, so that cross-comparisons can be carried out by looking at a single screen.
3. Ability to extend the above to several papers, so that one can work on more than one paper at the same time (as we often do with the paper versions).
4. Fast location of a paper starting from its volume and page numbers. Each journal should have an easy-to-remember root for the URL of the paper archives, to which to add volume and page with a simple syntax (e.g. <http://npb.com/500-111> for the article appearing on page 111 of volume 500 of Nucl Phys B.)

Until this progress is made, the hard-copy issues should not disappear from libraries. Debugging and improvement to the electronic consultation facilities will happen anyway, as we shall always try to reach articles directly from our office, if we know exactly what we are looking for. But we cannot be deprived of access to hard copies, which in many instances provide easier access, and are better suited for study and research while in the library.

As a final remark, I want to stress the cultural value of a visit to the library: this gives the opportunity to consult books, to take a look at recent issues of journals in fields other than our own specific professional interests, and opens the door to unexpected interesting discoveries. Going to the library is good for us, and it would be an ecological disaster if the dominance of electronically available journals were to discourage this habit!

PS .I am pleased to mention, at this point, that all removed volumes of PRD and PRL have been, one year later, restored to the shelves. I thank the personnel of the Library for understanding our needs and accepting this request, in spite of the major logistics trouble caused by the return to the status quo.

Author Details

Michelangelo Mangano

Theory Division

CERN, CH 1211 Geneva 23,

Switzerland

e-mail: <http://library.cern.ch/HEPLW/2/papers/2/Michelangelo.Mangano@cern.ch>