Electronic Publishing in Science, Paris, Feb. 1996

Participants and topics

The joint ICSU Press/UNESCO conference Electronic Publishing in Science[1] was held in Paris from 19–23 Feb. 1996 and attended by *ca.* 150 participants drawn from learned societies and academies, scientific associations, commercial publishing houses and librarians. The scientists present were predominantly from the physical sciences. A series of speakers presented material for the most connected with primary (*i.e.* journal) publication grouped in the sections

- Where are we now?,
- Electronic data storage, access and archiving,
- Tools and standards for protection, control and presentation of data,
- Legal and ethical issues in electronic publishing,
- The Scientist's view of electronic publishing,
- Views of learned societies and professional scientific publishers,
- Economics and organization of primary electronic publishing, and
- Options for the future.

The working groups discussed matters under the topics

- Real costs,
- The Electronic Archive,
- Scientists in Developing Countries,
- Implications for training and Work–styles, and
- Peer review,

and the conference produced a set of general consensus recommendations[2] for transmission to UNESCO (and consequently member country governments) and ICSU (and consequently member unions, academies and learned societies).

The current author was present at the EPS conference as a delegate of the International Union of Crystallography (IUCr). The IUCr is a specialist union within ICSU, the International Council of Scientific Unions. IUCr is a publisher producing 6 primary journals in crystallography and other reference works and books of immense scientific and practical value such the *International Tables of Crystallography*. Necessarily this report dwells particularly on those issues judged by the author to be of particular concern to the IUCr. This is unfortunate since the papers presented at the conference contain a wealth of information on the way copyright, libraries and librarians, archives and historical records are responding to the requirements and challenges of the electronic publishing in science.

Blackground

Scientific publication as such has entered a state of crisis that the advent of electronic technology is exacerbating. Governmental budgets world–wide are in deficit, the world economy is running slow and funding for science (and education) has decreased. The impact has been a net decrease in library subscriptions to scientific journals. With high–fixed overheads and a decreasing client base, many journal publishers have little choice but to make proportional subscription rate increases to cover lost revenue. Scientific publication is not exactly subject to the rules of a market–driven economy since for the most part, the producers and consumers are the same people. Further the number of papers published continues to increase and the migration of authors to journals published in the USA remains a dominant feature.

If not exactly on the lips of every speaker, it was frequent to hear suggestions that the number of papers 'published' should be reduced. 'Published' here is taken to mean peer–reviewed documents where matters of language and layout or typography are undertaken professionally.

[1] http://www.lmcp.jussieu.fr/icsu/Information/Proc_0296/
[2] http://www.lmcp.jussieu.fr/icsu/Information/Proc_0296/Recommendations.html

Analysis led a significant proportion of participants to the conclusion that scientific peer−reviewed publication is a product which mostly benefits the author who should thus bear the publication costs through page charges.

The technology of print−on−paper, due to its very nature, leads to the necessity of multiple local libraries in order for information to be readily accessible to a user. Due to the limited content of library catalogues, search capacity is poor. In the STM domain one has hence seen the appearance of a whole series of 'secondary services' (index, abstracting, citation, current−awareness bibliographic and other specialized databases) offering search capacities far greater than an ordinary library catalogue. In general these secondary services are operated by organizations separate from the primary journal publishers.

Eground

By contrast information diffusion by electronic means offers a rather different picture. On the one hand, there is the opportunity for more publication. It is not difficult or expensive for an individual or a association to set up a home page or a rudimentarily−indexed repository of text articles. Such a technique has proved a tremendous success in the 'Ginsparg' repository of scientific works in high energy physics. Many wonder whether this success can be repeated elsewhere or is only due to the social and technical characteristics of scientists carrying out high−energy physics research (small closed circle of people, highly computerized, all working together in centralized experimental facilities, all using the same text processing system and there being no commercial interest in the scientific results). Astronomy has a similar structure. A further novel facet of the individualized web publication is its fluidity, meaning that a document can be updated, modified and amended by its author without restraint. It also allows greater scope in style and language compared to formal scientific primary publication.

On the other hand a more top−down approach to electronic publishing finds its motivation not only in electronic delivery but also in the increased efficiency offered to the scientist by providing integrated secondary services: indexing, abstracting, forward referencing, keyword and full text searching, data abstracting. This view point finds support from the existing publishing and library establishment. In the test bed scientific Digital Library (DL) at the University of Illinois at Urbana−Champaign, primary scientific articles are prepared in a standardized semantically marked−up form (see SGML below[3]). Such a mark up allows the automatic extraction of identification information for the integrated secondary services. Users access the search capacity and the full−text articles across the network using standard WWW browsers to which an external SGML viewer is added. The Illinois DL is serving students and faculty from a group of universities (5 to 8) covering a student population of 500'000 for an area with a civilian population of 9.5 million people. The DL contracted site licences with the publishers of the electronic journals who delivered the articles in electronic form as SGML documents marked up with various DTDs. Those responsible for the Illinois DL are convinced that scientific publishing will evolve into a system of federated full−text and specialized databases operated by a diversity of organizations and that for the economic operation of these systems institutions will have to form together into consortia. This view was also supported by the other librarian who presented a talk at the conference.

Clearly a prime necessity for electronic publishing and communication in the sciences is the need for suitable bandwidth on the network, adequate hardware and software for the scientific user coupled with the requisite training.

Several levels and approaches

In several comments and presentations at the conference the notion of electronic scientific information distribution at different levels or in several tiers was apparent. On one tier, at a low cost and in high number, one would find rapidly−distributed fluid articles in various file formats with limited indexing and in a wide variety of languages. On another tier, a costly one but with few publications, fully peer−reviewed articles, presumably all in English, tagged in a standardized way

[3] #sgml

allowing very powerful globally−integrated secondary services.

APS, the American Physical Society is implementing a plan to restructure its publishing activities. Amongst other actions, the number of publications will be reduced, journals will be made available electronically, review and action is to be taken to speed up the non−technological aspects of publishing and an eprint (electronic pre−print) service is to be implemented. The AAS, the American Astrophysical Society in its plan is devoted to an integrated publishing activity with a large emphasis on powerful secondary services. A project in discussion by IUPAP, the International Union of Pure and Applied Physics, is based on a two−tier approach around regional centres of eprint repositories from which articles are published by existing journals.

Several media

Studies have shown that the use of electronic means for information diffusion is better and more widely accepted in the USA than in Europe. So in producing documents at the present time and in the near future, it will be necessary to do so both in electronic form and as print−on−paper. This double format is a factor in increasing costs. Further this attachment to the traditional support is a major factor limiting more widespread authoring of truly distributed interactive hypertext and virtual reality.

SGML and ISO 12083

SGML − Standard Generalized Markup Language is an ISO−standard for defining tags to undertake semantic markup of text. The actual tags are defined by the user in a dictionary called a DTD. SGML is rich and complex, and writing a DTD is a specialist's work. General SGML text systems are expensive. Nevertheless SGML is being increasingly used in the electronic publishing industry. A recent ISO standard (12083), designed by physicists, provides a set of DTDs for the specific needs for electronic publishing in science and technology. An EPSIG, Electronic Publishing Special Interest Group, has been formed as an association and provides a newsletter, discussion groups and tutorials in its effort to promote ISO 12083. HTML, the hypertext markup language used over the World−Wide Web is based on a different SGML−compliant DTD.

SGML and ISO 12083 have many advantages for text. Consistent use of the same DTD over a wide community of users was requested by several speakers at the conference. This would help considerably in the extraction of information from the SGML files of the primary journal publications to fuel the secondary services: *e.g.* indexes, abstracts, citations and forward referencing *etc*.

However there are difficulties in the application of SGML to tables, mathematics and chemical formulæ. I have the impression that this is due to the conflicting structures of these data forms with the data model of SGML. One can not hope for a rapid solution to these problems. For example, for a mathematical formula to be of any use in an analytical mathematics package it needs to be semantically correct. I understand that a recent working group came to the conclusion it is not possible to define an SGML DTD which will allow the semantically−correct markup of all mathematical formulæ. One is thus forced to use a graphical or rather typographical markup for mathematics. Another curiosity of ISO 12083 is that it does not, at least at present, allow for the markup of distributed hypertext so successful with the World−Wide Web. It seems that ISO 12083 dates from the time when the principle objective of 'electronic publishing' was the production of print−on−paper by electronic means!

Viewers for SGML are commercially available. In the Illinois DL, the Panorama SGML viewer of SoftQuad was selected. Authoring tools suitable for a scientific user are, in the words of one speaker, still around the corner so production of SGML−marked documents is undertaken at present by professionals.

16th July 1996 – © Howard Flack[4] – **Not to be copied or reproduced without permission**

[4] Howard.Flack.html