Problems and Solutions in EP Technologies

V. Nechitailenko

Geophysical Center, Russian Academy of Sciences, Moscow, Russia

Abstract

The paper contains a brief comparative analysis of different tools for electronic publishing including publishing of an electronic clone of a POP journal and an electronic journal with a POP version. The analysis is based on the experience gained by the Electronic Publishing Group of Geophysical Center, RAS which was involved in the creation of these types of journals (seven titles with nearly 2000 papers) under a grant agreement with the American Geophysical Union.

Introduction

It has become commonplace to state that electronic publishing provides many advantages to authors and publishers, as it allows us to pass over limitations of traditional printed-on-paper (POP) journals. At the same time some of the tasks which are solved easily in traditional publications can produce problems in electronic publishing, especially in the screen presentation of mathematical texts.

The potential power of the most advanced technologies which are being developed now including further extensions of HTML/SGML, Java and JavaScript programming languages, can result in very comprehensive and exciting scientific online publications, but it is very doubtful that these tools could be widely used by both scientists and publishers.

Authors' and publishers' approaches to the technology of electronic publishing are quite different. Authors' primary interest is to use those tools and technologies which could help them to present their scientific concepts, analysis and conclusions in the best way. Economic aspects of scientific publications are out of the scientist's primary interest, while these are in the first rank of the publisher's interests.

It seems that widely used term 'electronic publishing technology' is not informative enough. There are great number of different tools and possibilities for sending data and information over the network, and presenting them on screen and/or on computer-based media. Publishers select and use those tools which they consider most effective (the last term does not mean the same for different publishers), and most convenient for the perception by the potential reader community.

Online electronic journals

Among a great variety of electronic publications used for scientific communications, electronic online journals have a special interest. In our discussion we consider the electronic journal as a tool for scientific communication based on computer networks, and using hypertext (SGML in a wider sense) and multimedia *as principal components of document architecture and contents.*

While developing a technology chain of preparing, publishing and distributing a serial online journal a publisher must define the following:

- guidelines for authors (limitations and/or recommendations on volume, style, and formats of the basic and ancillary components of paper);

- a type of electronic journal, i.e. an electronic clone of a POP journal, an electronic journal with POP copy, or an electronic-only journal;

- a model of the reader community (what is the so-called 'typical reader', tools used by him for interacting with electronic journal, etc.).

There is a number of reasons why publishers of electronic journals accompanying them with POP versions:

1. *Author's position.* As *Coles* [1996] noted, while nearly 75% of scientists would like access to electronic journals only slightly more than 50% are prepared to submit their works to electronic journals. The number of scientists in academic institutions who have regular access to electronic information services is about half the number of people in industry.

2. *Tradition.* Existing publications depositories (archives, libraries), as well as indexing and refereeing systems are highly oriented towards traditional publications. Though we can see great activity in the development of principles and concepts of an electronic scientific archive by both international scientific bodies and intergovernmental organizations [*Russon*, 1997], this system is far from being completed.

3. *Copyright.* While existing international and national legislation protects author's and publisher's rights, the latter can be easily destroyed in the case of electronic journal, because we have no reliable system for long-term saving of the original electronic paper which can be used as a basic reference item.

4. *Subscriptions.* This is a very important issue. Subscription income can provide full or partial coverage of production cost. Some estimates show that accumulated subscription income will be higher in the case of *per-paper* or *per-use* subscriptions compared to the traditional *per-journal* subscription. This is due to an expected essential increase in individual online subscribers. Unfortunately there is a great obstacle in the way, i.e. the unreliability and the limitations of existing electronic money transfer systems.

These and some other considerations stimulate publishers to publish POP versions of their electronic journal (very often concise versions). This copy is not a POP clone of the electronic journal but it can help to resolve the problems mentioned above. As one of the most successful examples of this kind, we can mention *NSSDC News*, published by the NASA's National Space Science Data Center. The POP version of *NDSSDC News* presents concise versions of all papers published in the full-scale electronic journal (*http://nssdc.gsfc.nasa.gov/nssdc_news/*), and provides readers with basic information such as titles, authors, abstracts and hyperlinks to original full text papers. Another example, a recently started electronic journal DGGGMS Gerald (*http://www.uipe.ru/*), looks less successful. The POP version of this electronic journal is published as a full set of HTML pages, as they are seen on the screen, and thus the advantages of POP journal, in terms of producing high quality printouts, are not used.

The present author is far from being a guru able to give recommendations of a general type. Instead, he will provide a short description of experience gained from publishing electronic clones of POP journals as well as an electronic journal with a POP version.

Example 1. Electronic clone of a POP journal

Electronic clones of five translated geophysical journals of Russian Academy of Sciences were created under the grant agreement between the American Geophysical Union (AGU) and the Geophysical Center, Russian Academy of Sciences (GCRAS) [*Hedlund et al.*, 1998].

The AGU along with other learned societies which are involved in publishing activities, uses the TeX publishing package and its subsets like $LaTeX$, $AMSTeX$, etc., and widely recommends authors of AGU's journals to use the $LaTeX$ package while preparing manuscripts. AGU provides authors with all necessary information on 'how-to' and style files (templates) developed for each AGU journal. As it was pointed out by the AGU Publications Committee [*Editorial*, 1998], '*... we intend to stay with $LaTeX$ as the first choice for electronic submission of journal articles until there is a well-accepted, international standard that easily and reliably can handle math across different platforms.*' That is why we will not consider here in details all pros and cons of the TeX package; let us mention only that (i) it is the best publishing package today for maths presentation, (ii) the package is a public domain type and is easily acceptable, and (iii) it has a powerful tools for generating templates. The lack of graphics tools in the system is easily compensated for by variety of extensions.

The technological chain for producing camera-ready copy of each issue and its electronic clone included:

1. Marking up initial ASCII files following the standard developed for galleys, including headers, maths, tables, bibliography and other components of papers. A special galley template, containing a number of TeX macros, was developed for this step. Authors did not participate in the preparation of $LaTeX$ galleys because initial ASCII files were provided by translators of Russian original papers. This resulted in high homogeneity of initial materials and made our task much easier, while total production cost became higher compared to other AGU journals.

2. Preparing, in parallel to step 1, graphics files containing figures and some tables with incorporated graphics.

3. Composing page layouts and an issue as a whole following recommendations developed for each translated journal. The template ($LaTeX$ style file) developed for this step made it possible to do the following:
   - to compose the paper's title page with one- and two-column parts,
   - to put 'floating' objects (figures and tables) in the place following the journal style,
   - to generate PostScript and HPGL files of each paper for final printing of camera-ready copy.

4. Producing components of each journal's electronic clone, including HTML versions of abstracts, tables of contents, and associated information. The latter includes data on volumes of original and compressed PS and HPGL files, some statistics describing papers, hyperlinks to other components of the online database. This information is extracted from ancillary files created by special macro definitions at step 3.

The detailed description of the technology chain used is published on an Electronic Publishing Experimental Site created under the AGU/GCRAS Agreement (*http://eos.wdcb.rssi.ru/agu/technol.htm*), and includes the full set of macro definitions, batch files and programs used in this chain.

Sample 2. Electronic journal with a POP version

Geomagnetism and Aeronomy International *(GMAI)* (*http://eos.wdcb.rssi.ru/gmai/gmai.shtml*) is a new electronic journal recently launched by AGU and published in cooperation with the Electronic Publishing Group of GCRAS. As is pointed out by *Hedlund et al.* [1998], according to the *GMAI* concept the electronic version is a primary while the POP version is secondary and is shortened due to inherent limitations, such as lack of colour, interactive components, real floats, etc. At the same time the POP version is prepared as a usual professional edition which meets AGU's standards.

When developing and adapting software and the technological chain for preparation of this journal, we used the positive experience of publishing electronic clones, described above. Basic parts of technological chain include:

- preparation of a paper's manuscript, accepted by the Editorial Board for consideration, and converting it to the $LaTeX$ format. This is 'done' either by the author or by the technical editor,
- translation of the $LaTeX$ paper under the template developed for presentation a paper in the so-called 'review format,'
- converting the $LaTeX$ paper into an HTML version using $TeXWeb$ filter and putting it on the WWW server. Only reviewers and editorial staff have privileges to access papers on this peer-review stage. Reviewers can copy papers for reviewing and write their reviews in online mode, using the Peer-Review Online System (PROS) developed for this electronic journal [*Nechitailenko*, 1998].

- preparation of online and POP versions of the paper after the peer-review stage.

The most difficult task is converting heavy maths into the screen images. Well known packages for translating TEX to HTML , such as LaTEX2HTML [*Drakos*, 1997] or the HTML template in MS Word, replace formulas and separate maths characters with generated images. The TTH filter [*Hutchinson*, 1998b] uses a quite different approach, presenting formulas as regular tables and filling in these tables with math characters which are supported by modern WWW browsers. Both packages are rather universal and produce conversion to HTML of a full set of LaTEX macro definitions (in case of LaTEX2HTML) or plain TEX (in case of TTH). Nevertheless, due to some limitations of these packages we gave the preference to the TEXWeb filter, developed in the frame of this project.

It's most important limitations are as follows:

- LaTEX2HTML is developed for use under the UNIX operating system, while all our technology chain is based on IBM PC computers;
- the process of screen graphics generation is unnecessarily complicated while the latest versions of the emTEX package contains all necessary drivers;
- templates we are currently using contain dozens of macros which are not supported by the LaTEX2HTML package;
- matrix presentation of formulas by TTH is not very compact, though it has a great advantage over LaTEX2HTML, as TTH uses WWW browser settings, when scaling fonts in formulas;
- use of TTH can also produce problems if Cyrillic fonts and diacritical characters are used simultaneously in formulas.

This list must not be considered as a strong judgment. It reflects at first hand our specific needs. For a more comprehensive analysis see [*Hutchinson*, 1998a].

A quite different approach was used in the *IBM Techexplorer Hypermedia Browser* [1997], which is a plug-in viewer for WWW browsers. At least one feature of this viewer seems extremely important. The WWW server sends a document to the client in the TEX format, i.e. there is no necessity to convert a TEX file to HTML at the server's side. This also produces less traffic compared to the traffic produced by PS and graphics files.

At the same time, the existing version 1.2 essentially extends a set of macros for managing screen presentation, while it has no means of interpreting some LaTEX macros or TEX macros outside of LaTEX. This simply means that the publishers need actually to prepare two versions of the initial document in TEX format, the first version for the electronic journal and the second one for the POP version, if they wish to produce POP version at the professional level.

The TEXWeb filter is a Perl 5 script working under OS UNIX as well as under Windows 95/NT Perl interpreter. TEXWeb is not a universal tool and solves rather local problems of converting a LaTEX version of the paper, prepared under the fixed template into HTML format, though it can be adapted to other styles by following macro definitions in the attached configuration file.

For converting maths to HTML it scans the LaTeX file, extracts simple maths macros and compares them with the predefined set of HTML macros. If an extracted element does not match any of HTML macros, it is converted to a graphic file with the scaling defined by the default WWW browser settings. Converting to graphic files is performed using the set of drivers from the standard emTeX package. We are actually using the emTeX version which supports Cyrillic fonts [*Khimenko and Shen'*, 1996].

TeXWeb also provides generation of hyperlinks to the reference list and to floating objects (figures and tables). Floating objects are generated as separate documents which are loaded to the main browser window or to separate individual windows (default option). The latter is very convenient for readers, giving them the possibility of opening and closing floating components when necessary.

The full description of TEXWeb filter is published on the server previously mentioned (*http://eos.wdcb.rssi.ru/texweb*).

Conclusion

Electronic publishing looks like one of the bandwagons of modern information technology. It suggests new possibilities for both readers and publishers, still has some unsolved problems, and gives excellent promises. As players on this scene we need to carry out a twofold task: to be in front of newest ideas and developments, but at the same time to avoid possible gaps between authors, publishers and readers. As our experience shows, it is arealistic task,and there are the ways for smooth transition from printed to electronic publishing world.

References

B. R. Coles, Electronic Publishing in Science: Where are we now? In: *Electronic Publishing in Science*, Paris, 1996 (URL *http://eos.wdcb.rssi.ru/eps/coles.htm*)

N. Drakos, R. Moore, The LaTeX2HTML Translator. (URL *http://www-dscd.llnl.gov/files/programs/unix/latex2html/manual/manual.html*)

Editorial. AGU's Electronic Publishing Strategies, *EOS, 79*, No. 3, p. 29-30, Jan. 1998

J. Hedlund, V. Nechitailenko, J. Sears, *Electronic Publishing in Geophysics* - Presented at Second ICCC/IFIP Conference on Electronic Publishing 98, Budapest, Hungary, April 20-22, 1998.

I. Hutchinson, Approaches to WWW Mathematics Documents. Feb. 1998a, (URL *http://hutchinson.belmont.ma.us/tth/webmath.html*)

I. Hutchinson, TTH: a TeX to HTML translator. Feb. 1998b, (URL *http://hutchinson.belmont.ma.us/tth/tth.html*)

V. Nechitailenko, Peer-Review Online System, 1998 (URL *http://eos.wdcb.rssi.ru/papers/pros.html*)

D. Russon, ICSTI takes action in the development of the electronic scientific archive, *Science International*, No. 66, 1997.