

Integrated Electronic Publishing Environment: A Case Study

Vitaly Nechitailenko

Geophysical Center RAS, 3 Molodezhnaya Str., Moscow 117296, Russia
vitaly@irtysh.wdcb.rssi.ru

Abstract

The state-of-the-art of electronic publishing in science and its nearest perspectives with the emphasis on technical issues are discussed. The discussion is based on the real-life activities of the author and his colleagues in publishing online journals (<http://eos.wdcb.rssi.ru/>) and the creation of an integrated electronic publishing environment. The EP environment includes a full set of technologies and tools for preparation of papers from manuscript to print and online versions, including online submissions, peer-review process, updating and revising papers, back- and forward referring, TeX to HTML conversion with hyperlinking based on automatic syntax analysis, etc.

Introduction

Electronic publishing technologies based on modern computers and computer networks have become very effective tools for data and information exchange in all areas of human activity, but especially in the scientific fields, which deal with large numbers of multidimensional data sets resulting from unique experiments initiated by researchers or found in nature. Planetary Geophysics is one of these fields, which far exceeds many others in terms of the diversity of disciplines under study, observational systems and networks, as well as techniques of analysis and interpretation of data.

This high level of diversity produces an interest in electronic publishing in the geophysical scientific community, and stimulates the exponential growth of electronic publications in geophysics. Today, the general picture in this area varies greatly. Along with such publications as *Earth Interactions* (<http://EarthInteractions.org>) there are also many less complex publications, including electronic versions of traditional journals, collections of abstracts and bibliographical indexes, as well as numerous non-serials (books, reviews, papers).

Typical papers and books in Planetary Geophysics can be compared with mathematical publications and painter's books at the same time. Original results of observations and analyses are usually presented in various formats on different media. Very often scientist needs to spend a lot of time to put their results into the Procrustean bed of rules used by traditional printed-on-paper (POP) journals and books. Electronic publishing technologies give us possibilities to overcome these POP limitations enhancing papers with multimedia and other electronic extensions.

Electronic Publishing Experimental Site

One of the most actively developed Web sites (<http://eos.wdcb.rssi.ru>) was created in 1995 by the Geophysical Center RAS in cooperation with the American Geophysical Union (AGU). This project succeeded the Translation Project started in 1993 and aimed to publish five Russian geophysical journals in English: *Atmospheric and Oceanic Physics*, *Geomagnetism and*

Aeronomy, Geotectonics, Oceanology, and Physics of the Solid Earth.

Running the Translation Project we strictly followed two principles from the beginning:

- using only computer technologies in all stages of camera-ready copy preparations, including typesetting, composing math and graphics, as well as composing papers and issues as a whole, and
- saving and archiving all final and intermediate products on computer media.

These principles provided the foundation for creating, in a short time and with minimal extra cost, electronic clones of POP versions of translated journals, soon after Joint ICSU Press/ UNESCO Expert Conference on Electronic Publishing in Science (Paris, 1996; <http://eos.wdcb.rssi.ru/eps/>), which gave a great impetus to the project completed in 1997. The database is about 3 GB (~1,500 full text papers, ~6,000 documents) and includes electronic clones of five mentioned above journals.

Next step was taken with publishing the online full-text journals Geomagnetism and Aeronomy International (GMAI), launched by AGU in 1997 and published in English (<http://eos.wdcb.rssi.ru/gmai/gmai.shtml>), and Russian Journal of Earth Sciences (RJES), launched in 1998 and published in Russian and English (<http://eos.wdcb.rssi.ru/rjes/>). The technology developed for these journals covers all stages of preparation of manuscripts, math, graphics, online and POP versions of the journals. Both journals are primarily online journals but accompanied with POP versions.

Why do we need POP versions of the the electronic journal? There are a number of reasons why publishers of electronic journals are accompanying them with POP versions:

1. *Author's preference.* Many (if not most) of the academic authors still prefer POP journal as a basic tool for dissemination of scientific results.
2. *Archiving system.* Existing publication depositories (archives, libraries) as well as indexing and refereeing systems are highly oriented towards traditional publications. Though we can see great activity in development of principles and concepts of electronic scientific archives by both international scientific bodies and intergovernmental organizations [Russon, 1997], this system is far from being completed.
3. *Visibility.* Currently online journals are deposited and assigned primarily to URLs of publishers' home sites. Universal systems of identification, such as the Digital Object Identifier (DOI) System, (<http://www.doi.org>), and URI (<http://www.ics.uci.edu/pub/ietf/uri/>)/ URN (<http://www.ietf.org/html.charters/urn-charter.html>), are still in the study phase.
4. *Copyright.* While existing international and national legislation protect author's and publisher's rights, the latter can be easily damaged in the case of electronic journals, because we have no proven, agreed upon, reliable system for long-term saving of the original electronic paper which can be used as a basic reference item.
5. *Subscription.* This is another 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 the per-paper or per-use subscription compared to the traditional per-journal subscription. On our Web site each paper published online is accompanied with Postscript versions of the same paper composed according to standards of POP journal.

Main Phases of Technological Chain

The unified technological chain is used for both journals. Differences of journal styles are minimal and covered within used templates and software configuration files. Main phases of the used technological chain looks as follows:

Phase 1. Preparation of initial version of paper. According to current practice a paper can be

submitted either in ASCII text format or in LaTeX format. In the first case technical staff convert ASCII text into LaTeX version of the paper. The author is also requested to submit text and graphics in hard copy and encouraged to submit corresponding graphics files in one of graphics formats listed below. The author may also submit extra files (EF) containing movie, sound, animated graphics, etc.

Phase 2. Publisher's technical staff verifies a completeness of all submitted materials and produce initial LaTeX version of the paper under manuscript templates developed for different journals.

Phase 3. Postscript and HTML versions of the initial paper are produced from the LaTeX version. Corresponding links are included into HTML version for existing multimedia and other extensions. After Phase 3 the paper is included into a Peer-Review Online System.

Phase 4. Both considered journals are peer-reviewed. Depending on reviewers' preferences they can use either old paper technology, writing their reviews and sending them to the Editor, or they can use the Peer-Review Online System, which supports writing online reviews and provides reviewers and managers with a set of other options (see below).

Phase 5. Papers accepted after the peer-review phase are updated and corrected following the author's updates made in accordance with reviewers and editor recommendations. For final composition page templates are used instead of manuscript templates. Final Postscript and HTML files are produced at this stage and the paper is included in the current issue of the online version of the journal. Normally online version of each particular paper become accessible to readers within a few months before the hard copy of the corresponding issue of the journal appears.

Author and publisher views on the technology issues

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

Author and publisher approaches to the technology of electronic publishing are rather different. The author's primary interest is to use those tools and technologies, which could help him to present his scientific concept, analysis and conclusions in the best way. Economical aspects of scientific publications are out of the scientist's primary interest, while these are in the forefront of the publisher's interests. The side effect of the author's approach is the usage of software tools, which is badly compatible with those used by the publisher.

The current EP technologies and tools, and those which will appear in the next few years, produce strong influence on authors' and publishers' roles and vice versa. In this respect an emphasis is given to hot and cool technologies and tools, a problem of their harmonization, and their influence on the economics of electronic publishing.

While selecting tools and technologies for the mentioned serial electronic publication, we must take the following into account:

- compatibility of these tools with already used technology chain components
- necessity to have special training for both authors and technical staff
- production cost, if it increases, what other advantages does new tools give us?

- possibility to use this technology at the author's end,
- dissemination and subscription issues, etc.

The rapidly changing electronic publishing environment obviously suggests more and more possibilities for both author and publisher. Estimates of applicability of any new technology vary in wide range. The progress in this field sometimes turns negative, as new hot tools and software used by authors do not match the technology used by publishers and vice versa. In opposite to practices and standards developed in connection with traditional print publication, practices and standards for electronic publishing in science are still in progress. To fully use the positive and to eliminate some negative aspects of electronic publishing technologies as they grow one need to define at least temporal "Rules of Road" for both authors and publishers. Such general rules were recently published by the ICSU's Committee on Dissemination of Scientific Information (ICSU Press)[*ICSU*, 1999].

Recommendations to Authors

Though both online journals are oriented on international authors' community, currently most of the authors are from Russia or from FSU countries. The average standard of computer equipment in Russian research institutes is not high enough, so we try to avoid to give too restrictive recommendations to authors. Along with the set of recommendations to authors concerning traditional print publications, GMAI and RJES authors are requested to use the following formats while preparing their papers:

Basic text. The first choice for text format is the LaTeX (LaTeX 2e is recommended though LaTeX 2.09 version is also accepted). It is strongly recommended to use templates (LaTeX style files) developed for these journals they are available at <http://eos.wdcb.rssi.ru/tools/>. Authors must avoid using any macros except those defined in mentioned LaTeX versions, or in the mentioned templates. A LaTeX file of each paper must be accompanied with the manuscript printout containing all necessary elements (math, tables, headers, captions, etc.) and proofread by author. Those authors who are unable to submit paper text in LaTeX format may submit an ASCII file of the paper and the manuscript printout with math and chemical formulas clearly written by hand containing all necessary mark-ups requested by journal style.

Graphics for printed copy. The first choice for graphics format is EPS. Recommended resolution 300dpi. For one-bit graphics (black & white) PCX, MSPn, and BMP formats may be used. Though necessary scaling can be easily produced at the final stage, authors are recommended to prepare graphics files in 1:1 scale, to avoid reducing the quality of the graphics.

Graphics for online copy. Preferred are GIF and JPEG formats. It is recommended to use color graphics for online copies in all cases when this is applicable. Authors may submit only color version of graphics but must be aware of that formal conversion of color image to grayscale image may result in losing readability. If, for example, a set of indexed colors is used as a scale, after conversing this set to grades of gray, using standard procedures of some popular graphics editors, the resulting scale can be ugly. It is recommended further that an image for screen presentation must have a resolution no higher than 100dpi and a physical size within 800x600 pixels (which will be good for most online terminals). With higher resolution or size only a part (sometimes small) can be seen on screen. As a rule paper must be accompanied with two sets of graphics files, for printed copy and for screen presentation.

Tables. It is recommended to avoid tables bigger than one printed page. Very often a big table is in fact small relational database. In this case it is recommended to arrange them as an electronic supplement to the paper rather than as a part of the paper.

Sound and video. These components must be self-consistent and must be able to call up from WWW browsers, like Netscape Navigator, v. 3.0 and higher or Microsoft Explorer, v. 4.0 and higher. Currently no editorial work is done with these files.

Demo software and interfaces to databases. These components can be of two types. The first type of software does not work in real time and need to be loaded and installed on the user's computer. The second type can be activated during the session, like HTML documents enhanced with JavaScript or Java applets. Using server side includes (SSI) components in the paper are possible after thorough analysis made by the server's Webmaster. It is necessary to avoid loading malicious codes which can occur in submitted files. For the same reason we don't use currently online loading of manuscripts using PUT method of the CGI protocol.

EP Technologies

Technology used for electronic clones of POP journals

While developing technology for producing camera ready copies under the Translated Journals project (<http://eos.wdcb.rssi.ru/transl/>) we followed the AGU standards and traditions. As many other learned societies, which are involved in publishing activities, AGU widely uses the TeX publishing package and its subsets like LaTeX, AMSTeX, etc. and recommends to 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.' We will not consider here in details all pros and cons of the TeX package. More details can be found [*Nechitailenko*, 1998a, 1998b] at <http://eos.wdcb.rssi.ru/tools/>.

Technology used for online journals

The most difficult task is converting heavy math into the screen images. Well known packages for translation TeX to HTML, such as LaTeX2HTML [*Drakos and Moore*, 1997] or the HTML template in the MS Word program, replace formulas and separate math characters with generated images. The TtH [*Hutchinson*, 1998a, 1998b] realize 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 converting to HTML 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, which is under development now with the beta version in use.

The TeXWeb filter is the Perl 5 script working under OS UNIX as well as under the Perl for Win32 interpreter. It was developed as an attempt to solve a two-fold task:

- to jump from direct interpretation of TeX instructions to installing 1-1 correspondence between TeX and HTML macros, and
- to have a mechanism of easy updating which could follow the HTML evolution and the transition to XML.

The TeXWeb is not an universal tool and solves rather local problems of converting LaTeX version of the paper, prepared under the fixed template, into HTML format, though it can be adapted to other styles following macro definitions in the attached configuration file.

For converting math to HTML it scans the LaTeX file, extracts simple math macros and compares them with the predefined set of HTML macros. In a case when the extracted element is not matching any HTML macros it is conversion to the image with the scaling defined by the default WWW browser settings. Converting to graphic files is made using the set of drivers

from the standard emTeX package. For Russian language journal we are using the emTeX version which supports Cyrillic fonts [Khimenko and Shen', 1996].

TeXWeb provides also on-fly composing of the paper's contents, generation of hyperlinks to the referenced items and to floating objects (figures and tables). Floating objects are generated as separate documents, which are loaded to the main browser window, or to the separate individual windows (default option). The latter is very convenient for readers giving them a possibility to open and close floating components when necessary.

Although, in the above described technological chain, the POP version of the journal is secondary, it is prepared as a normal professional publication. The total production cost of the electronic journals are higher – up to 20-25% extra costs compared to an equivalent POP journal and approximately the same for the journals with no POP versions, mainly due to the necessity to produce different graphics for the screen and paper presentation.

The technologies and software used for the two mentioned journals are published in more details on our server (<http://eos.wdcb.rssi.ru/tools/>).

Peer-review Process

Among the tools which were developed for online journals there is a special Peer-Review Online System (PROS), which has a high enough level of security, and supports on-the-fly generation of all documents in the system, depending on confirmed privileges of users (reviewers) and IP addresses of computers at user's end.

Only authorized access to the system is permitted. The system distinguishes two types of users: reviewers, those who are registered in the system and have valid usernames and passwords, and managers, which are the registered users accessing the system from the predefined computers (identified by IP addresses).

The PROS has two main branches, Reviewer's Page and Manager's Page. After entering valid usernames and passwords the system brings the user to the Peer-Review Page, which is composed in a different manner for reviewer and manager. The reviewer can see and work with papers assigned to him by the manager and can go to the Reviewer's Page, while the manager can see and work with all papers loaded for the review process and can go to the Reviewer's Page and to the Manager's Page as well. Except for the initial page with the authentication form and Info page called from the latter, all other pages are generated on-the-fly by a set of Perl scripts depending on current scenario and user's privileges.

The following options are provided by the system for reviewers:

- changing user's password
- confirming acceptance of the selected paper for review
- writing review in online mode

The system composes two-frame windows and loads the paper under review into the top window and Review Form in the bottom window. The reviewer can copy and paste necessary fragments from the paper's text in the top window. Online versions of the manuscript contains HTML 3.2 extensions and/or inline graphics to display math and special characters. When cutting and pasting such fragments from the paper into the review, the reviewer may loose some characters. Necessary recommendations on the matter are provided online. After closing the current session reviewer can continue editing during following sessions.

The managers can also do the following:

- add user's account to the list of registered users
- remove user's account from the list; this operation removes one or more marked users from the List
- add paper(s) to the Peer-Review Page; this operation accepts filled in forms, containing the paper's code, title, and author's names, and creates links to corresponding directories assuming that these directories are already installed and loaded
- remove paper(s) from the Peer-Review Page
- define reviewer(s) for the selected paper; this option provides mechanism for assigning reviewers for selected paper
- read newly submitted and already accepted reviews
- view the list of registered users, and event log list
- generate reports and reviews, which are sent to authors

More detailed description of the Peer-Review Online System can be found at <http://eos.wdcb.rssi.ru/tools>. There is also possibility to load a non-commercial version of the system, which can be configured for any arbitrary journal.

Conclusion

Electronic publishing looks like one of the bandwagons of the modern information technology. It suggests new possibilities for authors, publishers, and readers and gives excellent promises, but still has some unsolved problems. As players on this scene we need to solve a two-folded task; to be in frontline of new ideas and developments, but at the same time avoid possible gaps between authors, publishers and readers. As our experience shows it is a realistic task. There are ways for smooth transition from the printed to the electronic publishing world.

References

Drakos N. and Moore R. The LaTeX2HTML Translator. <http://www-dscd.llnl.gov/files/programs/unix/latex2html/manual/manual.html>, 1997.

Editorial. AGU's Electronic Publishing Strategies. EOS 1998: 79 (3) January, 29-30.

Hutchinson I. Approaches to WWW Mathematics Documents. <http://hutchinson.belmont.ma.us/tth/webmath.html> 1998a.

Hutchinson I. TtH: A TeX to HTML translator. <http://hutchinson.belmont.ma.us/tth/tth.html>, 1998b.

ICSU Guidelines for Scientific Publishing. Third Edition. ICSU Press, 1999, 33 pp.

Khimenko V. and A. Shen'. Vokrug TeX'a (Around the TeX), 1996, 60 p. (in Russian, distributed as a part of Russian LaTeX 2e package)

Nechitailenko V. Technology of Electronic Publishing in Geophysics, ICSU Press Workshop, Oxford, March 1998a, <http://www.bodley.ox.ac.uk/icsu/nechitaippr.htm>

Nechitailenko V. Problems and Solutions in EP Technologies, Proceedings of the Second ICC/IFIP Conference on Electronic Publishing 98, ICC Press, 1998b, 73-8.

Russon D. ICSTI takes action in the development of the electronic scientific archive, Science International 1997, 66, 1-3.