ELECTRONIC COLLOQUIA: IDEA AND PRACTICE

Jochen Bern
bern@uni-trier.de

Christoph Meinel
meinel@uni-trier.de

Harald Sack
sack@uni-trier.de

FB IV — Informatik
Universität Trier
D-54 286 Trier
GERMANY

Abstract

The scientific community, especially the more computer related fields of science like computer science and mathematics, have not only been doing research in information technology, they have simultaneously begun to exploit its advantages for their own purposes. However, scientific publishing is different from other areas where information technology has been applied in that it requires publications to be not only fast, but also comparable to traditional media in terms of quality control and long-term availability. Peer reviewing and paper-based publishing are still the methods of choice to achieve the latter goals, restricting electronic means to more or less “inofficial” communications.

Within the field of scientific communication, we find a number of types of communication with slightly different goal and character: From scientific journals with their extensive and, consequently, slow reviewing mechanisms over conference papers to — nowadays usually electronic — forms of publication by the author. On the other hand, and in spite of the delay incurred, peer review is advantageous: The remaining publications are reasonably filtered for quality, making the series, resp. conference proceedings, worthwhile to read.

The department for computer science at the University of Trier (Trèves), Germany, is researching and implementing solutions in the area of information technology. In this paper, we will present two of our three main projects.

Meanwhile, our university has created the *Center for Scientific Electronic Publishing (WEP) [6],* which promotes the exchange of ideas on information technology and its applications between departments within the university.

The projects presented in this paper are the *Electronic Colloquium on Computational Complexity (ECCC) [7]* and the *STACS Electronic Submission Service [15].* For both, we will introduce the specific situations they were developed for (Sections 2.1 and 3.1), present the specific problems of interest and the chosen solutions (Sections 2.3 and 3.2), and conclude with a report of current status and use (Sections 2.4 and 3.3). Finally, we will outline our view of the future both for the field as a whole and for the specific projects (Section 4).

Published in the proceedings of the *International Symposium of SIGDOC'98,* Quebec (Canada), 1998 (to appear)

Keywords

Information technology; electronic journals; electronic colloquia; electronic submission to conferences; Electronic Colloquium on Computational Complexity (ECCC); Symposium on Theoretical Aspects in Computer Science (STACS)

1 Introduction

Scientific publishing has special requirements that have influenced the choice of media and somewhat hindered exploitation of information technology so far. The most important ones are:

- Research results are being achieved, and hence need to be published, at an ever-increasing rate. While one would expect this to be a strong promoting factor for the use of information technology, it actually resulted in a a special form of publication, usually labelled “personal communications”. This type of publication has promoted the idea that fast means of communication necessarily entail low quality and quotability of the respective material.

- With the increasing amount of publications to evaluate, the readers become more and more dependent on publications with restrictive filtering based on quality and topic. However, this puts the growing burden onto the reviewers and entails increasing delays for these publications.

- The most persuasive reason for paper-based publishing, however, is that researchers need to be able to cite other publications in an unambiguous and lasting way, which is not necessarily guaranteed in the case of a central data repository run by a single person or organization.

On the other hand, there is a substantial need for the use of information technology, at least between publisher and reader, because publications are facing a decreasing number of readers. With the growing cost of the filtering process, paper-based publications are currently on the verge of becoming prohibitively expensive unless cheaper and more focused means of delivery become available.

1.1 Scientific Journals

The prototypical media for the abovementioned filtering processes are journals. In computer science, the delay between submission and publication in a high-ranked journal may well exceed a year. "High-ranked", on the other hand, extends to all means of publications available; publication in a respected journal is often considered the ultimate form of approval in the respective scientific community.

In the recent years, there have been a number of *electronic journals* appearing on the WWW. However, upon closer inspection, most of them turn out not to have changed the filtering process in a significant way, but only the "reader interface", so to speak. The shortcoming of long publication delays hence persists in these.

1.2 Preprints and Preprint Servers

At the other end of the spectrum, (paper-based) preprints are an established means of circulating results in a timely fashion. Quotability is ensured by certain properties of paper-based media, namely the fact that copies are circulated to interested parties. These copies prevent later changes of the publication unless the new version gets properly labeled as such, usually by means of preprint numbers.

Consequently, the direct electronic counterpart of paper-based preprints — usually condensed into *preprint servers* maintained on a departmental, organizational, or even larger basis — is often regarded as an interesting source of information, but lacking in terms of reliability and usability for citations in own publications.

1.3 Conferences

As a compromise between these two extremes, conferences have become more and more important for researchers to stay in touch with the development in their field. While conference papers, which will later be published in the conference proceedings, are subject to a refereeing process, meeting colleagues in person allows researchers to exchange ideas in an informal way. However, refereeing for conferences is subject to tighter time restrictions and does not achieve the same level of quality control as journals.

As a consequence of the timing restrictions, the question whether submissions should be accepted on paper or electronically is not entirely up to the organizer's preferences in the case of conferences. Electronic submission to conferences has become the state-of-the-art method in spite of the existing compatibility problems because of its speed.

Electronic submission services need to fulfill increased security requirements because a disturbance of the service near the deadline can invalidate individual papers: With deadlines for a series of conferences being usually a year apart, resubmitting a paper for next year's conference is problematic at best, and useless at worst.

Another issue of special importance in conference submissions because of timing constraints is proper choice of allowed file formats. Currently, electronic submission most often requests PostScript files to be sent; however, PostScript is not fully portable (to the point of making it unreadable) and does not support postprocessing of the submissions (e.g., to add "watermarks" like copyright notes, or to collect submissions into a single conference proceedings document).

2 The Electronic Colloquium on Computational Complexity (ECCC)

2.1 Environment

The ECCC [7, 3, 4] is mainly intended for use by scientists in the field of *Computational Complexity*. This field is part of theoretical computer sciences, yet, the strong mathematical orientation causes the community to be composed of computer scientists and mathematicians alike. The number of researchers active in Computational Complexity that have registered at ECCC is about 100; for comparison, the number of theoretical computer scientists registered at the Virtual TCS Rolodex [1] is about 1,000. The ECCC newsletter is sent to about 300 subscribers, including several local exploders.

In 1994, a group of top researchers in the field of Computational Complexity discussed the possibility of creating an "online meetingplace" to exchange recent results in a timely, yet still somewhat filtered, manner, and to maintain public online discussions. Supporters of this idea (which later formed the *scientific board*) agreed upon the desired mechanisms, and ECCC was founded at the University of Trier in late 1994.

2.2 General Design

2.2.1 Conceptual Idea

In terms of organization, there are three main groups of participants to the whole system: the *scientific board*, a group of currently 38 researchers in computational complexity from all over the world; the *local office*, a group of persons at the University of Trier (actually the authors of this paper) administering the physical system; and the users, which potentially includes the whole internet. More or less as a snapshot of the current user base, there is a mailinglist to which regular announcements of newly published material will be sent. However due to mail exploders this list may not reflect the real situation.

The basic types of publication units in ECCC are *ECCC Reports, Comments,* and *Revisions.* Besides of these, there are "ECCC Books" — other material published at ECCC that (due to its intended aim, audience, or size) does not quite fit into the concept of a scientific report describing ongoing research. To mention: PhD theses, lecture notes, survey papers, and monographs — the latter is the ECCC internal synonym for mostly voluminous works (e.g., book projects) that describe a certain branch of research to a large extent in greater detail. However, ECCC books do not follow a standardized submission and publication process and therefor we do not further mention this type of publication at ECCC.

ECCC Reports and Comments, between time of submission and time of publication, exist in a *Submission State* so that they are visible only to the group supposed to *screen* them (which is the term we chose to avoid already used terms like "referee"). Report Submissions and ECCC Reports receive unique serial numbers, while Comment Submissions, Comments, and Revisions are numbered with respect to the number of the ECCC Report they are referring to.

2.2.2 Document Submission

The mechanism that allows submission and screening of an ECCC Report is actually the most complex. First, one of the authors has to produce a PostScript (tm) file with the report to submit and email it — along with certain metadata in a specific format — to the address `eccc-submission@eccc.uni-trier.de`. For a detailed description of the required format, see the URL `http://www.eccc.uni-trier.de/eccc/info/how-to-submit.html`.

The ECCC server will receive the submission, analyze and process it, and finally enter it into the list of current Report Submissions as well as emailing a result notice to the given email address. During the processing, the server splits the submission back into its components (email headers for future reference, "description" with the metadata, and the PostScript file), performs sanity checks on the metadata, and attempts to add a title page header to the PostScript so that printouts will unmistakably state status and number of the Report Submission.

The members of the scientific board will, again by email, request the current list of Report Submissions whenever their time permits. Upon finding a submission they're interested in, another emailed command will lock the submission for them so that no other board member will be able to screen it, and optionally mail them the PostScript file. Finally, a third emailed command will accept, reject, or simply unlock the Report Submission. All board-only commands will not be accepted unless proper proof of identity of the sending board member is presented; in addition, every single command is acknowledged via email to a *known good* address of the board member in question.

As soon as a decision is made, another email notification is sent to the submitter detailing the ECCC Report number alotted to the paper and the revision submission mechanism (if accepted) or an optional remark to the author from the board member (if rejected). If the submission has been rejected, it is removed from the current submissions directory; if it has been accepted, an ECCC Report number is alotted, the respective public directory, an overview HTML page, and a PostScript file with the Report Submission header replaced by the proper ECCC Report header will be created, and the remaining data is moved to a nonpublic directory for future reference. The lists of ECCC Reports are created anew in regular intervals, for reasons described below, so there is no need to change them in the process as well.

Finally, the server will regularly recreate the list of current submissions, the per-year lists of reports, the metadata for the bibliographic servers as detailed below, the databases for the search mechanisms, and the `ls-lagR` files in the FTP area (all of this several times per day); create a template for the newsletter to the mailinglist from the metadata of new ECCC Reports (once per month); and check for both submissions and Screening Locks that approach or exceed their respective inactivity timeouts.

2.2.3 Copyright Considerations

One major potential problem for the acceptance of ECCC was that publication of papers in a journal or conference usually requires that it does not get "published" elsewhere. While preprints and preprint servers are usually not considered "publication" in this context, ECCC might well be considered as such, which made users reluctant to actually submit papers to ECCC.

In discussions both with users and publisher companies, ECCC decided not to claim copyrights of any work (so that the authors may later transfer the *complete* copyright to a publisher) and, consequently, allow ECCC Reports to be removed upon author request if necessary. The numbers assigned to the removed papers will *not* be reused so as to ensure proper citability; removed papers will be replaced by a notice that the paper has been removed and, reference to the published version. Copyright issues are the only accepted reason for withdrawal of material published at ECCC.

2.3 Improvements over Previous Models

2.3.1 Filtering Process

The noticeable turnaround time of journals was avoided by using a stripped-down filtering process, tentatively called *screening.* Submissions to ECCC are brought to the attention of all members of the *scientific board.* After selecting a submission (which is done according to personal preferences and on a voluntary basis), one single board member performs the whole screening. Namely, he acquires an exclusive lock on the submission, decides over acceptance or rejection, and sends his decision, which will be effective immediately, via email to the *local office*.

The server will send appropriate notifications if submissions stay locked by a single board member, or unrequested, for a long time. In particular, there is an enforced timeout to the screening process for the case that no member of the scientific board is interested in judging the paper, so that the decision for a submission will be taken two months after submission at the latest.

Because of the fact that the scientific board includes a large part of the top researchers in Computational Complexity, and comprises the members of most editorial boards of journals in the field, the quality of the screening is not much inferior to a respected journal's refereeing.

2.3.2 Ongoing Discussions

Apart from such submissions, which become *ECCC Reports* when accepted, ECCC supports *Comments* on and *Revisions* of existing ECCC reports. Comments may be submitted by everyone, but will be screened; revisions, on the other hand, can only be submitted by the original author, but are not subject to screening.

For every ECCC report, an overview page — meant to be the "official entry point" of the report — gives the basic information about the report. It then proceeds to list the comments and submissions received for the report. The intent is to list not only the report, but also the ensuing discussion of the report and corrected or updated results from the authors at a central point. Unlike paper-based publications, where special publications *(citation databases)* are required to complement the backward references in the publications themselves with forward references to later contributions, this mechanism will display the whole discussion no matter which document the reader got referred to.

2.3.3 Quotability

ECCC guarantees quotability by restricting the possibility of having material removed from the server to one single case — when papers included in ECCC have been published elsewhere in a quotable fashion *and* the publisher's copyright regulations mandate that said paper may not be published elsewhere. In that case, ECCC would still list the report but refer readers to the new source. Other than that, all ECCC reports, including all their revisions, and their respective comments will stay on the server and can be freely retrieved. Including references to the ECCC has become quite usual in other publications in the field of Computational Complexity.

2.4 Implementation and Statistics

2.4.1 Reader Interface

Most readers will access the ECCC content with a web browser, using the HTTP [7] and FTP servers set up for ECCC. However, as a consequence from being set up in a time when a considerable part of the scientific board did not have reliable access to the Internet, the information is available purely via FTP and even from a dedicated FTPmailer as well. (Likewise, the `eccc/` part of the URL is a relic from the days when HTTP servers did not support virtual hosts.)

In addition to the static content, search mechanisms have been implemented for the HTTP interface which allow to search for keywords in author names, titles, and abstracts.

2.4.2 Submitter and Board Interface

Submissions and commands from the scientific board are received via email only, as we use (in the case of the board, registered) email addresses as our primary means of identification. Submitters are required to compose their submission email according to a template, which includes several lines of singled-out metadata and a PostScript file representing the paper (resp. comment) itself. Board members will use predefined email subjects for commands, and receive PostScript submissions for screening by email.

2.4.3 Specific PostScript Issues

An ubiquitous task in handling submitted papers that are stepped through several states by several people in remote collaboration is to keep track of the current state and proper identification of the papers. For example, submissions to the ECCC receive *submission numbers* according to the date and time of their *arrival at ECCC*. These are replaced by a *report number* reflecting date and time of their *acceptance* in later stages. However, it is highly desirable to have the paper, resp. any printouts, show the submission number to the screening board member, as he needs the number to submit his decision.

While PostScript is not modifiable in the general case, we were able to develop tools that will add small parts of text or graphics to PostScript files as long as they have been created by *LaTeX [8]*, more precisely, `dvips`. We use these tools to "watermark" ECCC documents with additional information, especially their identification numbers.

2.4.4 Manually Edited Contents

Besides the contents that are administered in the automated way described above, the WWW server of ECCC contains a large amount of material of interest to the community collected by the local office. Since all members of the local office have accounts on the machines running ECCC, there is no need to implement mechanisms for remote control in this area; in other words, the respective web pages are maintained manually.

Such pages, to name only the most popular ones, contain links to homepages of fellow researchers and their departments, lists of upcoming events *(ECCC calendar),* links to other material related to Computational Complexity, and advance versions of book manuscripts made available to ECCC by the authors.

2.4.5 Statistics

As of the time of this writing, ECCC contains 256 reports and their respective discussions, representing a little more than three and a half year of activity. The scientific board has 38 members. Readers of ECCC are located in the USA (15+%), Germany (10+%) and about 75 other nations.

The number of report downloads from ECCC (i.e., counting only downloads of PostScript files) has risen from 3,734 (1995) over 4,938 (1996) and 11,289 (1997) to 5,261 (first half of 1998), while the total number of WWW hits amounted to 53,771, 47,902, 53,256, and 38,219, respectively.

2.4.6 Relation to Paper-Based Publications

The local office of the ECCC is producing printed copies of ECCC material both for its personal use and for archiving in the library of the University of Trier. In addition, ECCC has been assigned the ISSN 1433-8092.

However, ECCC will likely not be widely published by other means than its centralized servers because of the conflict with the possible removal of reports after their publication elsewhere.

3 Spin-Off: The STACS Submission Service

3.1 History

The Symposium on Theoretical Aspects in Computer Science (STACS) [15, 5] is an international conference covering all aspects of Theoretical Computer Science that german and french sites take turns hosting. It has proven to be — together with the ICALP conference — the main European conference for exchanging ideas in this area. The committee members are internationally respected, the number of submissions is high (typically in the range of 100 to 140 papers), and the acceptance rate is low (about 40 papers). Researchers come from all over the globe to attend the conference.

STACS '97 took place at the *Medizinische Universität Lübeck,* Germany, which at that time had a rather unstable Internet connection. Having acquired experience with submission mechanisms for PostScript papers with ECCC, we volunteered to handle electronic submission for them. Since the electronic submission software provided by ACM's SIGACT [14] featured some properties that we deemed both undesirable and easy to remedy, we started porting and rewriting ECCC software into another electronic submission suite oriented towards the specific problems of conference submissions.

3.2 Specific Issues in Conference Submission Services

The main problem in handling conference submissions is timeliness of the process. Unlike preprint servers and electronic colloquia like ECCC, which accept submissions basically at any time, and journals, which will repeat submission deadlines for single issues in rather rapid succession, success or failure in complying with the given deadline of a conference is very important for submitters. This results in a series of prerequisites that a conference submission service must fulfill:

- In order to automate (and, hence, speed up) the process as much as possible, if the submission format allows for portability problems at all (PostScript does), the submission service should *avoid possibilities for corruption* in the transfer of data and *provide reproducable and understandable error messages* if errors occur.

- In order to avoid confusion, the process of electronic submission should be not only simple, but also resemble the usual workflow for paper submissions where possible. For example, it is not intuitive that a PostScript file needs to be transmitted twice to be first checked out and then submitted, as is the case with the SIGACT software.

- Besides of ensuring a level of privacy that measures up to the privacy of surface mail, the security measures taken must *prevent attackers from blocking the server* (denial of service attacks) as far as can be achieved with reasonable overhead. This is especially true considering that conferences and their submission mechanisms nowadays get announced via WWW pages, which, in turn, get indexed by the common web search engines. SIGACT's basic assumption that only the respective scientific community will access the service, hence the service is not at danger because "the community is honorable", doesn't hold anymore.

- Since the submission service will not usually reside at the conference organizer's site, extra provisions have to be made to ensure timely, unaltered, and confidential delivery of submissions to the organizers, resp. the referees. As a side issue, any specific requirements that the organizers and referees may have on accompanying metadata need to be communicated to the submission service administrators in time because they don't use the data themselves, and won't realize such needs on their own.

3.3 Implementation and Use

As in the case of ECCC, we based the user interface of the STACS submission service on email with included PostScript files. Since the most popular and stable PostScript interpreter available is `ghostscript` [2], we chose to have incoming submissions processed automatically as follows:

- The PostScript file is processed by `ghostscript` such that it produces a single image file (currently GIFs) per (successfully interpreted) page of the paper.
- The submitter receives a basic analysis in reply, including the results of some simple integrity checks (checksums, size, first and last lines of the PostScript file extracted) and the number of pages produced by `ghostscript`.
- If desired (and unless prevented by specific security control commands issued by the submitter), the submitter can retrieve the output log of `ghostscript` and the images of single pages to further verify the portability of his PostScript.
- After verifying the results to his satisfaction, he may send a `SUBMIT` or `DELETE` command without retransmitting the paper.
- Every such PostScript file transmitted is identified by a "ticket", a string with a large random portion that prevents abusers from guessing other submitter's tickets.
- In order to avoid the possibility of tickets being snooped from email in transit, we *optionally* offer PGP [13] encryption in both directions. (Please note that with the current legislation in France [11], people there may not use encryption without a special permit from the *préfecture du département*.)
- In order to reduce the risk of service denial attacks, submitters need to successfully retrieve a ticket (thus proving that the email address the ticket is sent to is "live" and email to it is read by them) before submitting the paper (i.e., claiming a substantial amount of disk space on the server).

The resulting service has been used with outstanding success to handle electronic submissions to the STACS '97 (Lübeck, Germany) and STACS '98 (Paris, France) conferences, and will be updated and reused for STACS '99 (Trier (Trèves), Germany) and all following STACS conferences.

4 Future Developments

The most central point in future development of scientific electronic documents in general, as well as of ECCC and the STACS submission service in particular, is proper handling of metadata. This requires that more metadata will be collected from authors, and in more versatile formats than currently used. This data then needs to be converted into the desired target formats, most notably the input formats of *bibliography databases* like DBLP [9, 10] in the case of ECCC.

Other development work currently underway for ECCC is implementing automated means to update the *calendar* offered on the ECCC site. The goal is to free the maintainers from tasks of format conversion and creation of overview lists when they add calls for papers and other material received from sources like mailinglists.

Finally, we would like to improve the issue of portable and reusable file formats. With the spread of other authoring systems than LaTeX [8] in the mathematics and computer science communities, portability of submitted PostScript files gets worse and worse. We will evaluate the use of PDF instead of PostScript. We expect it to be superior both in terms of portability and reusability, at least for the current versions of the respective standards.

Acknowledgments

We are grateful to the University of Trier (Trèves) for supporting the technical facilities necessary to run both ECCC and the STACS submission service. We would like to thank our colleagues at the *Lehrstuhl für Komplexitätstheorie und VLSI-Design* for fruitful discussions on various aspects of the concept of ECCC. In particular, thanks to Stasys Jukna for his great commitment to the ECCC project, and to Jens Doppelhamer and Johannes K. Lehnert for their skillful assistance.

References

[1] ACM Special Interest Group on Algorithms and Computation Theory (SIGACT). The virtual TCS rolodex. Available HTTP: URL: http://sigact.acm.org/tcs-rolodex /woody.html

[2] Aladdin Enterprises (1989). Aladdin ghostscript [Computer software]. Menlo Park, CA: Available FTP.

[3] Bern, J., Damm, C., & Meinel, Ch. (1996). ECCC — ein elektronisches Kolloquium im Internet. Informatik-Spektrum, 19, issue 4/1996, 230–231.

[4] Bern, J., Damm, C., & Meinel, Ch. (1997). The electronic colloquium on computational complexity (ECCC): A digital library in use. Proceedings of ECDL '97 (Pisa), LNCS vol. 1324, 405–421.
Available FTP: Hostname: ftp.informatik. uni-trier.de Directory: /pub/Users-Root/reports File: 97-16.ps

[5] Bern, J., Meinel, Ch., & Sack, H. (1998). The STACS electronic submission service. To appear.

[6] Center for Scientific Electronic Publishing (WEP), University of Trier [On-line material]. Available HTTP: URL: http://www.wep.uni-trier.de/

[7] Electronic Colloquium on Computational Complexity (ECCC) [On-line material]. Available HTTP: URL: http://www.eccc.uni-trier.de/eccc/

[8] Lamport, L. (1985). LaTeX — a document preparation system. Addison-Wesley.

[9] Ley, M.. Digital Bibliography & Library Project (DBLP) [On-line material]. Available HTTP: URL: http://www.informatik.uni-trier.de/~ley/db/

[10] Ley, M. (1997). Die Trierer Informatik-Bibliographie DBLP. Proceedings of GI Jahrestagung INFORMATIK '97.

[11] Loi de réglementation des télécommunications, chapter 17 (1996). Journal Officiel, 27. July 1996, no. 96-659. Available HTTP: URL: http://www.telecom.gouv.fr/francais/activ /telecom/nloi17.htm

[12] Meinel, Ch. (1998). Elektronisches Publizieren im World Wide Web — Eine neue Dimension in der wissenschaftlichen Kommunikation. Forschung&Lehre, no. 3, 122–125.

[13] Network Associates (1990). Pretty good privacy (PGP) [Computer software]. Available FTP.

[14] The SIGACT electronic publishing board (1995). Information about FOCS '95 electronic submissions. SIGACT News, 26, no. 1, March 1995, 58–62.
Newer versions of this material (user instructions):

- Available HTTP: URL: http://sigact.csci.unt.edu/~scg98tt/SoCG98tt.html
- Available HTTP: URL: http://sigact.acm.org/~stoc98/STOC98.html

Complementary material (organizer instructions): Available HTTP: URL: http://sigact.csci.unt.edu/sigact/esubserve.html

[15] Symposium on Theoretical Aspects in Computer Science (STACS) [On-line material]. Available HTTP: URL: http://stacs.uni-trier.de/