The Electronic Colloquium on Computational Complexity (ECCC): A Digital Library in Use*

J. Bern, C. Damm, Ch. Meinel**

FB IV — Informatik
Universität Trier
54286 Trier
Germany

Abstract. The Electronic Colloquium on Computational Complexity (ECCC) is a digital library that specifically addresses the current problem of scientific publishing, more precisely, the problem of presenting suitably filtered work to other researchers, for the field of computational complexity. Developing the detailed concepts in discussions with a *scientific board* of researchers in this field, ECCC now fills the gap between author controlled electronic publication *(preprint servers,* very fast but lacking content filtering) and conventional journal or conference proceedings publication (currently taking months, if not over a year, from submission to publication). Additionally, like a real colloquium, ECCC supports ongoing discussions through the publication of comments to already published material. Further authors have the possibility to present improved versions of their publications while maintaining bibliographic consistency by version control.
In this paper, we will first describe the situation ECCC is meant to remedy (Sections 1 and 2) and then detail the setup with respect to organization (3.1), basic functionality (3.2 through 3.4), cooperation with other services (3.5) and plans for the future (3.6).

1 Introduction

The advent of the World Wide Web brought new possibilities to the scientific community: quick access to up to date information and quick dissemination of own results. However, the internet is a huge sea of informations and (apart, of course, from the underlying protocols) it is largely unstructured. Its structure can perhaps be compared to the associative memory of the human brain: it is often hard work to remember something, once you lost the association chain to it. Attempts to change this situation with the internet usually cannot compete with the immense growth of the net and number of information providers.

---

* In Proc. ECDL'97, LNCS 1324, 1997 405–421.
Work on ECCC is supported by "Forschungsfonds der Universität Trier"
** {bern,damm,meinel}@uni-trier.de;
http://www.informatik.uni-trier.de/~{bern,damm,meinel}/

One remedy of this situation (at least for researchers) could be digital libraries that collect publications in one or several fields in a way that allows to uniquely identify publications, allows access to them over years and decades. One usually has to choose between 1) largely unfiltered publication of research reports or 2) peer reviewed electronic scientific journals. As scholarly publication requires a scheme of quality control, 1) leads to publicly accessible documents, which are nevertheless considered to be in a status of prepublication. Possibility 2) still has the usual drawbacks of large publication delays in spite of the fast "electronic way" of handling.

The Electronic Colloquium on Computational Complexity (ECCC) started in late 1994 and is somewhere in between 1) and 2) (more like 2)). An international scientific board ensures scope and quality control of submissions in a way similar to the selection process of scientific symposia. The standards for this selection are high and (due to the persons involved) comparable to the ones applied at the top conferences in the field. Submissions that succeeded are published in the ECCC Report Series. Usually this happens 1–2 months after submission. Beyond being a mixture of 1) and 2) above, ECCC uses further possibilities of electronic media. After an ECCC-Report is presented to the community, everybody working in the field can contribute to the discussion on the result (like in a real colloquium) by either turning directly to the authors (ECCC provides convenient support) or by turning to the author *and* the "audience" (the readers) by sending comments that are published together with the original report. Further, authors have the possibility to present most recent versions of their work at the Electronic Colloquium. Version control ensures bibliographic consistency.

2 Situation prior to ECCC

2.1 Scientific Community in Question

The scope of ECCC is Computational Complexity — a central field in theoretical computer science with strong mathematical orientation and many different facets. Potential contributors (and members of the scientific board) are mathematicians and computer scientists. Most of them use computers mainly for communication purposes. The spectrum of users (readers) is larger: it includes theoreticians as well as computer scientists with background interest in theory but not actively working in the field.

It is hard to give estimates for the size of the community. About 100 scientists actively contributing to various aspects in the field have "registered" at ECCC by providing the URL of their Web-pages. The actual number however is certainly much larger. The "Virtual TCS Rolodex" ([9] — a list of home pages of people working in theoretical computer science maintained for the corresponding special interest group of the Association for Computing Machinery) stores about 1000 addresses. Again the actual number is certainly larger. This shows, that computational complexity is a strong community within theoretical computer science though certainly not the strongest. The list of subscribers of ECCC currently has about 300 entries (including several mail exploders).

2.2 Traditional Publication Media

In order to "really" publish a research result in computer science, researchers submit ready-made reports either to conferences or journals, where they will be read, refereed, selected, and, finally, printed. While this process involves selection and filtering, i.e., quality control (which is the main reason why this type of publishing is commonly regarded as "real"), it has become very time consuming. In addition, printed conference proceedings and journals have the usual disadvantages of classic media in general, like availability (unless bought by the local library), lack of search mechanisms, or non-correctability. The evolving electronic journals are faster of course, since printing is eliminated. However, publication delay is mainly governed by the refereeing process and the time required by iterated polishing of the paper.

2.3 Online Preprint Servers

In order to circumvent the problems and especially the delay in the classic publication process, a lot of authors started publishing research results by electronic means over the internet. Since these services are under the control of the authors' organizations, if not of the authors themselves, they allow instantaneous publication and correction, while availability is ensured by the existing internet protocols. However, the problem of searching, resp. gathering of bibliographic data, has not been finally solved — though there are several attempts to create collected bibliographic databases from distributed sources [6, 4, 3]. As a prize for correctability and complete local control, this approach is additionally subject to the citability problem; a publication which is not guaranteed to remain stable, accessible, or even in existence cannot be reliably cited in other publications.

3 Concept of ECCC

3.1 Organization

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 mailing list to which regular announcements of newly published material will be sent. However due to mail exploders this list may not reflect the real situation.

3.2 General Concept

The basic types of publication units in ECCC are *ECCC Reports, Comments,* and *Revisions.* Beside of these there are "ECCC Books" — other material published at ECCC that (due to its intended aim, audience, and 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 therefore 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.

The screening procedure is less formal than peer refereeing. It is much like the screening of submitted papers for conference proceedings. Report submissions get screened by members of the scientific board. They check whether the submissions are in the scope of ECCC, present new and interesting ideas in a readable way, and are nevertheless in strict mathematical form (contain full proofs). The board members are acknowledged leaders in the field, which ensures high standards for submissions to successfully undergo the screening procedure. There is an additional timeout rejection (today 2 months) to ensure that submissions will be decided upon in a limited time frame. Comment Submissions are currently screened by the local office, with the intent to change this over to a similar mechanism as for Report Submissions once the desired features and proceedings have been pinpointed. Revisions are not screened, as they are supposed to be corrected versions of already accepted ECCC Reports; however, it is ensured that only the author(s) of the ECCC Report can submit Revisions for it.

While the public access to ECCC allows several types of services to retrieve "published" ECCC data, from HTML down to FTPmail, there is only one service that ECCC will take input from — electronic mail. This service is necessary, due to the fact that several board members lack network access that reliably offers synchronous connections; since the received input is either endangered by overly complex instructions (esp. for submissions of any kind) or of very low bandwidth, we decided not to implement other types of service at this time.

3.3 Input Processes

Submission / Report Mechanism 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 meta data 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 meta data, and the PostScript file), performs sanity checks on the meta data, and attempts to add a title page header to the PostScript so that printouts will unmistakably state status and number of the Report Submission.

```
penthesilea:/pub/eccc/submissions/19970429.352# ls
Description     Identification  Remarks
Headers         Original.ps     paper.ps
```

Fig. 1. *Files for a current Report Submission.* `Headers`, `Description` *and* `Original.ps` *form the original submission email;* `paper.ps` *has the ECCC header added. Note the typical Report Submission number in the path; it starts with the date the submission was received. Since there is no file* `LOCK`, *the submission is not currently locked by a board member.*

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 allotted 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 allotted, 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 non public 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 meta data 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 mailing list from the meta data of new ECCC Reports (once per month); and check for both submissions and Screening Locks that approach or exceed their respective inactivity timeouts.

```
penthesilea:/pub/anonFTP/pub/eccc/reports/1995/TR95-044# ls
Abstract.txt  Paper.ps     README       index.html
```

Fig. 2. *Publicly accessible files of an ECCC Report. This directory is actually accessible by anonymous FTP, server* `ftp.eccc.uni-trier.de`*, path* `/pub/eccc/reports/1995/TR95-044`*.* `index.html` *collects all information about this report, including references to comments and revisions.*

```
penthesilea:/pub/eccc/report_ctl/1995/TR95-044# ls
Accept         Description    Identification  Remarks
Authors        Headers        Original.ps     Title
```

Fig. 3. *Non public files of an ECCC Report.* `Accept` *records the board member, time stamps etc. of the acceptation process.* `Authors` *and* `Title` *are singled out for technical reasons.*

Comment Mechanism While Comment Submissions look quite like Report Submissions from the submitter's point of view, they are not currently screened by the scientific board but by the local office; in spite of our efforts to promote use of Comments, the volume does not yet warrant a full copy of the mechanisms. This may be a matter of scientific ethics — it is a serious act to publicly comment on a colleagues research work in a scholarly-like environment. However, adding comments to own ECCC reports is more often used by the authors.

On the other hand the (still) low volume of incoming comments simplifies the identification of authorized screeners and makes part of the screening dialogue unnecessary (all office members have local shell access). Basically, the board members read Comment Submissions directly from files and send a single simple email to accept or reject them.

```
penthesilea:/pub/eccc/comments/TR96-055.001# ls
Description    Identification  Remarks
Headers        Original.ps     comment.ps
```

Fig. 4. *Files for a current Comment Submission; Note the analogy to Report Submissions, as well as the number now referring to an existing ECCC Report.*

Apart from this, the Comment Submission mechanism has only one major difference from the Report Submission mechanism in that Comments always refer to an existing ECCC Report. This gets reflected in the organization of the files, the assigned numbers, notification emails to the author of the respective report, and, most importantly, in the fact that Comments are inserted into lists of whatever kind directly adjacent to this report. Thus, even the per-year lists of previous years have to be recreated periodically, as the Reports may still have ongoing discussions.

```
penthesilea:/pub/anonFTP/pub/eccc/reports/1996/TR96-065# ls
Abstract.txt  README       commt01.txt   revisn01.ps
Paper.ps      commt01.ps   index.html    revisn01.txt
```

Fig. 5. *Public files of an ECCC Report which has one Comment and one Revision. Note that the references to both have been inserted into* `index.html` *and* `README`.

```
penthesilea:/pub/eccc/report_ctl/1996/TR96-065/Comments# ls -R
01/         LastNumber

01:
Accept          Headers         Original.ps
Description     Identification  Remarks
```

Fig. 6. *Non public files of a Comment. Note that this Comment may not have been Comment* Submission *TR96-065.001.*

There is an experimental setup to allow ASCII and/or HTML Comments, rather than forcing all Comments, no matter how short, to use PostScript.

Revision Mechanism As mentioned before, Revisions are not subject to screening; they are, however, required to come from the authors of the original ECCC Report. Hence, all screening mechanisms are replaced by a simple identity checking comparison, while the remaining mechanisms of a Comment Submission are reproduced.

Copyright Considerations and Report Retraction 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.

```
penthesilea:/pub/eccc/report_ctl/1996/TR96-065/Revisions# ls -R
01/

01:
Description    Headers     Identification  Original.ps    Remarks
```

Fig. 7. *Non public files of a Revision. Note the absence of* `01/Accept` *and* `LastNumber` *(last* submission *number).*

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.

Miscellaneous Information Apart from all the presented automated processing, the local office maintains other HTML pages manually in order to provide additional information to the users. Particularly popular pages list WWW home pages of researchers in the field of computational complexity (see also section 2.1, home pages of their departments, lists of upcoming events, and links to related information on other web servers.

3.4 Public Interfaces

Standard Interface: HTTP and FTP A recurring problem in ECCC, even before actual software was devised, has been the fact that a noticeable number of board members is *not* able to rely on their LANs having a usable synchronous connection to the internet, reducing the services available to them to services that can be used, directly or via gateways, with email. Since WWW-to-mail gateways were rather rare when ECCC got implemented in late 1994, not to mention board members being used to them, it was decided to make most of the data accessible in a format retrievable by anonymous FTP (and, hence, the existing FTPmailers).

Another advantage of this approach is that the FTP protocol is better suited for transfer of large files — PostScript files created from LaTeX sources include pixmaps for the fonts used, rather than relying on any PostScript fonts — and conversion of file formats between different systems.

The URL to access ECCC via WWW (more precisely: HTTP) is `http://www.eccc.uni-trier.de/eccc/`, while the topmost directory for the anonymous FTP access is `/pub/eccc` on the server `ftp.eccc.uni-trier.de`. In terms of hierarchy, the web pages

– continue to be served off a HTTP server down to the per-year lists and search engines,
– consist of a special (HTML) file `index.html` served off the FTP server for the per-Report overview page, and
– finally refer to the same (PostScript) files as the FTP access.

In the appendix we have some screenshots to demonstrate the users view on ECCC while accessing a certain report via HTTP.

Special Considerations for Users with Poor Network Access: FTP-mail Soon after the first versions of the ECCC software were running, we were confronted with the fact that a fraction of the board members were not used to FTPmailers, or even plain unable to receive emails as large as a complete PostScript file. On the other hand, running a standard FTPmailer software on the ECCC server was undesirable since it would most probably be (ab)used for requests to other FTP servers than `ftp.eccc.uni-trier.de` because it basically allowed the full range of commands the Unix (tm) `ftp` program has. Thus, we implemented a simplified FTPmail software which basically accepts lists of filenames to retrieve and directories to list, a set of options allowing to control the output format (notably options to `gzip` and split files prior to transmission), and has the server to use hardcoded — in fact, it directly accesses the local file system.

Further details on the FTPmailer can be obtained by sending an email with subject `HELP ECCC` to `ftpmail@ftp.eccc.uni-trier.de`.

Interactive Content: Web-based Search Mechanisms Following the popular demand for the ability to search for interesting information with CGI-bin's, we have added a set of specific ECCC search engines to the web pages. Unlike the usual concept of search engines, which do a full-text search over all available files — not a very successful approach in LaTeX-generated PostScript files — the ECCC search engines match entire ECCC Reports, represented by their `index.html`'s, by scanning the list of authors, the title, or the ASCII abstract, respectively.

Some Statistics Currently there are 185 papers published in the ECCC reports series — about 60 every year. We observe a continuously growing number of remote report downloads from ECCC (PostScript files), which is probably the most reliable number, as it usually indicates really interested users: 1995 we had a total of 3734 downloads. For 1996 and (the first half of) 1997 the figures are 4938 and 7499, respectively. This amounts to an increase from 10 over 13 to about 30–40 downloads per day on the average. The number of HTTP accesses are clearly much larger: the figures are 53.771, 47.902, and 60.338 for these 2.5 years. These data have to be taken with care, however, since on the one side they include incidental "visitors" as well as certain search engines; on the other side they do not count requests cached by local proxy servers. By far the most of all requests come from US domains (more than 30 per cent) and German domains (more than 10 per cent). These are followed by requests from Japan, France, Spain, Canada, United Kingdom, Italy, Israel (between 2 and 5 per cent) and requests from about 70 other countries.

3.5 Interfacing with Bibliographic Services and Printed Libraries

One of the problems in both journals or proceedings and preprint servers is that there is usually no bibliographic service maintained anywhere, unlike in books. With electronic publication, it can actually be a problem to find a paper even if authors and title are known, but not the affiliation (former or current). There are currently several attempts underway to alleviate this by collecting bibliographic data in defined formats from the author's individual server into an accumulated database on a bibliography server [6, 4, 3].

The major obstacle with these is that there is currently no generally accepted format for bibliographic data to adhere to. As an example, one might compare the format used by the *DB&LP* bibliography server [6] — heavily influenced by HTML, thus insisting on a hyperlink, but not requiring any more content description than the title, see `ftp://ftp.eccc.uni-trier.de/pub/eccc/db-bibliography-input/current` — to the one Hypatia [4] asks for, see `http://hypatia.dcs.qmw.ac.uk/html/bibliography.html`.

For the time being, we support the *DB&LP* bibliography server with appropriately formatted files and are working on support for NCSTRL [3].

In terms of printed libraries, we are currently archiving printed copies of ECCC ourselves. Further, ECCC has now an own ISSN number (1433-8092) which ensures unique bibliographic identification.

3.6 Future Work

More CGI Mechanisms Up to now, the list of upcoming events is maintained manually, with no automatic archival of the outdated announcements. Other software is available that implements an archive, but still requires manual maintenance of the lists and entry of new announcements [1]. Work is underway to implement automated creation of up-to-date lists, archival and archive searches, and file format conversions so that announcements only need to be fed into a proper HTML form. Parts of this system will be implemented along the lines of the existing DAPHNE system [5].

Better Searching: A General Format for Bibliographic Data As a consequence from the mentioned incompatibility of bibliography services with respect to their input formats, we concluded that the only viable approach for the time being is to implement a "master" database from which other formats can be extracted. In order to reduce problems with format conversion, this master format is required to

- allow to enter all items of information that may be requested by specific formats (i.e., the format needs to be expandable, and there has to be a fair number of items from the beginning to avoid later addition of required fields to lots of existing entries),
- explicitly allow multiple basic formats (like ASCII, LaTeX, pseudo-HTML, ...) *per field*, and
- automatic conversion between these basic formats where needed.

Once the master database is implemented, unique bibliographic information will be generated for all publications at ECCC. The need for offering such information in a standardized format is urgent. We observe more and more references to ECCC-Reports in printed media.

Mirrors Because of the decision to allow public access via several types of service, mirroring ECCC is nontrivial; e.g., the hyperlink from a per-year Reports list (served via HTTP) to a Report's overview page (served via anonymous FTP) necessarily has to spell out the name of the FTP server and an absolute path, and a mirror simply copying the list would refer users back to the main site rather than to the respective mirror of the FTP data. Thus, a certain part of the files has to be converted for every mirror of ECCC. In addition, it is unlikely that mirror sites will be willing to run nonstandard software — like the ECCC search engines or the FTPmailer — either.

4 Conclusion

We described the main design goals and the current status of implementation of the Electronic Colloquium on Computational Complexity. We see the novelty of the approach in an implementation of a means of publication of high standard research reports, that combines filtering and quality control with fast dissemination. The implementation to a certain degree models the holding of scientific conferences: there is a submission and screening procedure and there is support to comment on published reports.

The community welcomed ECCC as a contribution to considerably improve the scientific infrastructure.

5 Acknowledgment

We are thankful to the university of Trier for supporting the technical facilities necessary to run ECCC. We would like to thank our colleagues at 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 engagement in the ECCC project, and to Jens Doppelhamer and Johannes Lehnert for their skillful assistance.

References

1. *Complexity Special Interest Group of the GI (GI-FG 0.1.4) Web Pages,* Jochen Bern, Christoph Meinel, Universität Trier, `http://www.informatik.uni-trier.de/GI/FG-014/`
2. Jochen Bern, Carsten Damm, Christoph Meinel, *ECCC — ein elektronisches Kolloquium im Internet,* Informatik-Spektrum, Band 19, Heft 4/1996, S.230–231
3. *Networked Computer Science Technical Reports Library (NCSTRL),* Cornell University, `http://www.ncstrl.org/`
4. *Hypatia* Computer Science and Pure Mathematics Electronic Library, Mark Dawson, Paul Taylor, Queen Mary and Westfield College, `http://hypatia.dcs.qmw.ac.uk/`
5. *DAPHNE Authoring System*, ITWM-Trier (`http://www.itwm- trier.fhg.de/`)
6. *Databases & Logic Programming Bibliography Server (DB&LP),* Michael Ley, Universität Trier, `http://www.informatik.uni-trier. de/~ley/db/`
7. Michael Ley, *Die Trierer Informatik-Bibliographie DBLP,* GI Jahrestagung INFORMATIK'97 to appear, Springer-Verlag, 1997
8. *Ariadne — the Red Thread Through the Web,* MeDOC Consortium, `http://ariadne.inf.fu-berlin.de:8000/`
9. *Virtual TCS Rolodex,* ACM Special Interest Group on Algorithms and Computation Theory, `http://sigact.acm.org/tcs-rolodex/woody.html`

A Appendix

On the following pages we included some screenshots to demonstrate the users view on ECCC while retrieving an ECCC Report with comments using a WWW browser.