

# Digital Proceedings: Making CAAD-Knowledge Widely Available

Bob Martens  
Vienna University of Technology, Austria

Ziga Turk  
University of Ljubljana, Slovenia

## **Abstract**

This paper deals with the efficiency of publication means within the scientific community. Scientists spend a substantial part of their total time in visiting meetings and conferences, reading publications, writing emails, etc. Thus any improvements in the communication processes, in particular technologies which minimize non-value adding activities such as tedious retrieving of source material from library shelves, could have far-reaching (positive) consequences. In this respect a customized solution for the digitalization of paper-based conference proceedings is presented. The process of creating electronic copies of papers - disseminating knowledge on CAAD - in pdf-format is described as well as the respective decisions which were taken in this context.

## **Keywords**

CAAD-related Publications, Web-based Bibliographic Database, Searchable Index, Retrospective CAAD Research

## 1 Introduction

Pioneering experiments with free scholarly publishing on the Internet can primarily be traced back within the US scientific community obviously large enough to have a critical mass of scientists to run their own journals, in many cases independently of other foreign communities. Due to the size of the country also the funding of such development in terms of grants has been made available. Long before the advent of the Internet, in the late 1970s the US Congress - based on earlier research efforts funded by the National Science Foundation - was considering legislation for a National Periodicals Center (Biggs, 1984), a central mega-repository of refereed journal articles from which separate copies of articles could easily have been ordered. This initiative supported by many researchers was ultimately defeated through heavy lobbying by some publishers and large libraries. A possible counterpart to the US in this domain could be the European Union, since scientists from the EU taken as a whole produce roughly as many scientific papers as their US colleagues. By means of the so called "Fifth Framework Programme" the EU is strongly focusing on making the R&D process in Europe more efficient.

## 2 (Free) Electronic Publishing Repositories

The scientific publication procedure has been so far only marginally affected by the possibilities offered by the Internet. The reason is a lack of sound business models and pilots to demonstrate the ultimate benefits of time- and cost-efficient scientific publication procedures. Scientists write papers and review publications of fellow scientists for free. They, however, pay substantial subscription fees to paper-based journals, even though these generally only deliver the printed publications some 1-3 years after they were written. The establishment of a repository consisting of electronic publications (project reports, theses, conference and working papers, etc.) would make a significant difference especially concerning availability and accessibility. Regarding the technical developments a large scale migration to (free) electronic publishing is feasible and favorable compared to the current procedure. In some

fields of science, an initiative is being approached stopping scientists from subscribing to and writing for journals not available on the Web for free (Anon, 2001).

Further research and even more so demonstration work will be required to point out all resulting benefits and to re-engineer parts of the scientific publication process by establishing a fully functioning on-line service, where scientific output would be available almost for free and where a virtual on-line community of authors and readers would meet. Among a set of technological tasks the following research agenda could be defined:

- study of the life-cycle economics;
- investigation of social, psychological and legal barriers;
- re-use of available technologies to develop and implement the repository;
- setup of an on-line virtual community infrastructure;
- filling of the repository with a critical mass of scientific output;
- implementation of technologies for the automatic maintenance of the repository;
- development of adaptive, intelligent, agent-based user interfaces;
- creation of add-on services;
- extended development of an existing peer-reviewed journal;
- monitoring of user requirements and satisfaction by means of regular surveys.

A recent book by Tenopir and King (2000) provides an excellent survey of models and empirical studies from the 70's onwards, but has received quite mixed reviews (see <http://www-jime.open.ac.uk/book-reviews/tenopir.html>) since it indirectly defends the current commercial publishers' case, based on empirical evidence collected before the world wide web started to have a significant impact. The overall situation has thoroughly changed and studies which are more than 2-3 years old, although interesting, are only of marginal value.

A clear distinction in the web literature is noticeable between authors believing the free publishing model to be the future and those thinking that web publishing is only a little cheaper (10-20 %) than traditional paper publishing and that there is no need for fundamental changes in the business processes. Unfortunately the arguments on

both sides are often restricted to anecdotal evidence, and the individual's background strongly influences his opinion. Thus it is unlikely for persons with a traditional publisher or library background to be very supportive of the free publishing model, since it questions the entire basis of existence of their "industry".

### 3 Dissemination of CAAD-Knowledge

In the eighties first considerations regarding the implementation of CAAD occurred at many architecture schools, soon resulting in a need for exchange of ideas and experiences. Platforms such as ACADIA, CAADRIA, eCAADe and CAAD Futures were established to fill that gap. The (bi-) annual conference can be regarded as the major event and various other events were to follow. Up to the second half of the nineties paper-based proceedings were published, generally with a small circulation only. What happens to smaller editions is that they are safely stored away in the studies of participants of the conferences and university members, but hardly become available to the public. Library networks have so far not been systematically supplied with copies, as far as could be traced back.

CUMINCAD, an acronym for "Cumulative Index on CAD", is a bibliographic index compiling papers related to computer-aided architectural design (Martens and Turk, 1999 resp. 2000; <http://itc.fgg.uni-lj.si/cumincad/>). This work started in 1998 and helps focusing on future CAAD education and research activities. Implemented with a data-base, it allows searching and browsing in the ways usual on the Web. It provides a "historical evolution" to learn from previous efforts and draws attention to older original works that could have been ignored because they could not be found otherwise. CUMINCAD thus supports the search and the dissemination of CAAD-related publications.

Presently, all available CAAD-conference Proceedings are recorded in CUMINCAD and apart from general bibliographical data also a summary is presented. If, however, a record is of interest, it might be rather difficult to retrieve the full-paper. I.e.: the already mentioned problems regarding "gray literature" results in the fact that pro-

curing e.g. via the international "inter-library loan" is principally possible, but very time-consuming and rather uncertain as to its outcome. Studies are showing (Bjoerk and Turk, 2000) that particularly the younger researchers do not wish to bother anymore using any literature but the one that is conformably and available to them via Internet for free. Perhaps one succeeds in contacting the specific author directly via the registered e-mail address, the possibility of offering the digital or digitalized full papers, however, surely is to be regarded the required completion of CUMINCAD. A further aspect in favor of digitalizing is that a complete collection of Proceedings would become "transportable" by means of the CD-Rom medium granting a clear overview of the scientific conference performance of an association. At the time of writing more than 3.000 records are available and approx. a third of them were imported from the CADLine-Database (Kalay, 1991).

### 4 Digital eCAADe Proceedings (1983-2000)

Up to the 1996 Annual Conference, paper-based eCAADe proceedings were published in a relatively small number. Starting with the 1997 conference also an annual CD-Rom was produced, which made this kind of information indeed widely available. There are hardly any individual teachers, researchers, librarians, etc., disposing of a complete set of eCAADe proceedings. The small number of printed proceedings as well as the policy of changing conference hosts lead to a situation that an ordering afterwards is quite uncertain in terms of having success. However, eCAADe holds the copyright on those publications and in fact this set of publications is to be regarded a valuable "capital" of the association. Making nearly two decades of teaching and research work in this specialized field of CAAD available would therefore be of great importance (approx. 3.600 pages). For this reason the eCAADe-council decided to finance this project by own funding. This decision results from the offer of the Viennese company *Mediatecture* (<http://www.mediatecture.at>). The practical experience of this enterprise in this area, - the 1991- Proceedings serving as a test set in this framework

with approx. 300 pages - was particularly convincing.

In order to keep the documents efficiently small for the internet and also providing a full-text search, not only a scanning job (page by page), was performed. Furthermore, presenting the original layout, traced back to the original proceedings, was aimed at. The main item of performance was the conversion to pdf-files. OCR-scanning and conversion to MSWord was offered as a general feature, and moreover editing and eliminating scanning mistakes resp. finally creating Adobe-pdf-files. The abbreviation "pdf" stands for "portable document format" which can be implemented on practically all current platforms (<http://www.adobe.com>). Pdf-files have become popular and aim mainly at producing a printing format in its original form on any available computer by means of a reader (free of charge) or a plug-in, which means that complex software-installations become obsolete and using is available without any "ifs and buts".

Text recording and conversion in MS Word-documents was achieved by means of *Omnipage* (<http://www.caere.com>). Using the original fonts and type sizes so possible was attempted. Working with the produced Word-document soon showed that the handling of the given file size might become tricky due to the number of pictures. Thus dividing into several (part-) documents seemed meaningful. Moreover, a footer was inserted in order to indicate digital processing as such and subsequently, entering page numbering wherever missing. The main work, however, was to eliminate mistakes resulting in the course of text recording. The submitted printing quality of the individual Proceedings played a decisive role in this context: i.e. unsatisfactory printing performance leads to time-consuming corrections.

The Word-documents produced herewith still remain very memory-consuming in their cor-

rected form. By means of pdf-conversion effective compriming was achieved. And again experimenting was necessary due to unintended occurring of pixels leading to illegibility.

## 5 On- and Off-line Access

The possibility of producing html-documents has so far not been documented. Wishing to be able to create the best-possible match of the original this variation was not pursued, as html-presentations are not necessarily made for printing-out purposes (problems with page break and page references) and this is what the users are actually interested in. On the other hand the eCAADe 1997-Proceedings available exclusively in html-format were converted to pdf, not e.g. loosing any links specified.

As soon as the complete "rough material" was available in the pdf-format the next step was to create two different final versions:

- a complete indexed file for off-line use (CD-ROM)
- a collection of stand-alone files for on-line use (connected with CUMINCAD)

What has not been considered so far is that pdf-based publication provides the possibility of options and restrictions. If so required document printout can be prohibited, read-only rights being granted, whereas the addition of annotations or even text editing may be available provided "Adobe Distiller" (involving charges) is installed. Any restriction may be canceled by entering a password.

For the production of the CD-Rom-version the individual documents were compiled and indexed according to year of appearance. Definitive decisions concerning the production of a hybrid CD-Rom (Mac/PC) have been issued by the eCAADe council and this product will be available in the course of spring 2001. It was also agreed to add the already digitally produced proceedings (1997-2000) to this CD-Rom.

Approx. 400 individual contributions from the rough material were separated by means of the "Distiller" for the off-line-version. The Annual General Meeting of eCAADe-members in Weimar 2000 agreed to the proposal of the eCAADe-Council to have restricted access to

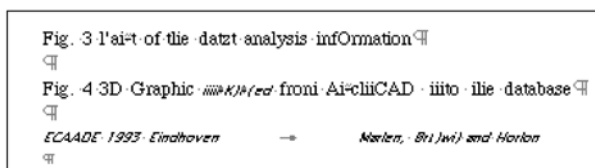


Figure 1. Example of an extremely "distorted" OCR-document.

those full pdf-papers, i.e. exclusively limited to eCAADe-members. CUMINCAD was adapted to meet those requirements, i.e. that the e-mail address is to be entered prior to transmission a pdf-contribution. Then a check is made, and provided the specific e-mail address is registered in the member database, the desired document is furnished as attachment.

## 6 Further Extension Plans of CUMINCAD

Due to the current expansion three stages of availability will be provided: anonymous access will remain, only issuing access to bibliographic data, however. E-mail address entry makes the user a "friend" also having access to the summaries. At present full access is only available to eCAADe-members and grants access to pdf-documents. This can possibly be understood as a contradiction to the idea of a free publishing model. The membership fee, however, is affordable and finances this kind of republication activities.

During the period of 22 months the CUMINCAD site served about 330.000 requests (500 per day). Two thirds thereof came from robots or from the two authoring institutions. The remaining 110.000 requests were from 20.000 different workstations in 30.000 different sessions (45 sessions per day). The majority of these people seem to be just passing by not exploring the site. 3900 visitors looked at more than 5 CUMINCAD pages and this is the number of users we can claim that CUMINCAD has. While keeping the service free, since late January 2001 it is required that users who want to see an abstract should register and provide his/her email address. In the last 4 months 150 have registered - about 1.2 per day.

Additionally, an exchange of ideas with the CAAD-associations (ACADIA, CAADRIA and SIGRADI) is to be encouraged possibly leading to a mutual access for members. ACADIA like eCAADe has a comprehensive collection of paper-based Proceedings. Subsequent procuring seems readily possible apart from some few exceptions (1998 resp. 1993-1999 still available). Anyhow, also a (partial) digitalization could seem meaningful. The situations of CAADRIA and SIGRADI might prove better as the present Proceedings probably are available in a digital for-

mat and thus mere simplified conversion or indexing work would be necessary.

## 7 Conclusion

This contribution issues an extensive description of the frame conditions for the digitalization of Conference Proceedings representing in its entirety a great asset for a CAAD-association. The wide availability of eCAADe-Proceedings in the period of 1983 to 2000 was accomplished by means of an off-line-version (CD-Rom) and an on-line-version (CUMINCAD), financially absolutely affordable. Considering the working effort involved it could have been rather tempting only to scan-in the total package without text recording, then, however, the file-sizes in on-line use would have proved quite problematic and, furthermore, a full-text search would not have been possible. Regarding this aspect the Digital Proceedings differ considerably from their paper-based originals. Doubtlessly, a digital publication can be designed according to specific appearance requirements from the very beginning provided the paper-based form is only to act as a mere byproduct. Apart from the more than 2000 pages 3.600 additional follow-up pages conference material from the various sources have been made digitally available. Thus the stage for further CAAD research work is well prepared (Turk et.al., 2001) and spectacular devices to this end will not have to be invented over and over again.

## References

- Anon (2001). Should the record of scientific research be privately owned and controlled?: Public Library of Science, <http://www.publiclibraryofscience.org/>, accessed 10.6.2001.
- Biggs, M. (1984). The Proposed National Periodicals Center, 1973-1980: Study, dissension, and retreat, in: *Resource Sharing and Information Networks*, Spring-Summer 1984.
- Martens, B. and Z. Turk. (1999). Working Experiences with a Cumulative Index on CAD: "CUMINCAD", in: *Architectural Computing from Turing to 2000* [eCAADe Conference Proceedings] Liverpool (UK), 327-333.
- Martens, B. and Z. Turk. (2000). The Creation of a Cumulative Index on CAD: "CUMINCAD" in *ACADIA Quarterly* 19:3, 18-19.
- Martens, B. et.al (forthcoming 2001). Digital Proceedings: Experiences regarding Creating and Using, in: *eCAADe 2001 Conference Proceedings*.

- Kalay, Y.E. (1991). *CADLINE, A Bibliographic Database on CAAD-Related Publications*. Buffalo: State University of New York.
- Tenopir, C. and King, D. (2000). *Towards Electronic Journals: Realities For Scientists, Librarians, And Publishers*. Special Libraries Association.
- Turk, Z. et.al (forthcoming 2001). The Topics of CAAD - A Machine's Perspective, in: *CAAD Futures 2001 Conference Proceedings*.