

# Repositories for Scientific Information Exchange

## *An Overview on SciX-Pilots Related to CAAD*

*Bob Martens*

*Vienna University of Technology*

*<http://www.tuwien.ac.at>*

*Academic publishing started in the hands and under the control of learned societies. Open access to digital publications has the potential of returning control to these associations again. In this paper CAAD-related digital library pilots, which were developed in the framework of the SciX-project (“Open, Self Organising Repository for Scientific Information Exchange”) will be elaborated. SciX focused its work on the infrastructure and business models for subject-specific repositories. Electronic publishing can re-establish the central role of scientific communities and associations in the scientific publishing process.*

**Keywords:** *Database Systems; Scientific knowledge management; Retrospective CAAD research; Web-based bibliographic Database*

### **Introduction**

On the occasion of the 20th eCAADe-Conference (Martens et.al, 2002), the SciX-project was presented. SciX stands for “Open, Self Organising Repository for Scientific Information Exchange” and is financially supported by the European Union (IST-project 2002-2004). The goal of SciX is to analyse the business processes of scientific publishing, to invent new publication models and through a series of pilots to demonstrate how this should work. In the envisioned scenarios, professional associations such as eCAADe play an important role. Their members are the potential users of SciX’s platforms, authors and readers of the papers.

CUMINCAD - Cumulative Index on CAD -, which was initiated in 1998, served as one of the pilots

for this project. Currently over 6,400 CAAD-related publications are recorded and from these nearly 4,000 full papers in pdf-format have been made available. All in all over 2,000 users have registered for this service since 1998. The vision of “Open Access” has not yet fully been realized. Registration for CUMINCAD is free, but download of full pdf-papers is restricted to members of the contributing CAAD-associations (“added value of membership” for nearly 500 individuals). However, negotiations with the CAAD-associations have started, regarding the matching of added value for members and Open Access to the scientific community. Discussions are ongoing and aim at restricted access to recent papers (two years old at a maximum) for the CAAD-

membership.

eCAADe has delivered a substantial contribution in terms of content to CUMINCAD, which is highly valued by the users. In a timeframe of two decades (1983-2003) nearly a thousand conference papers have been produced in the framework of the eCAADe-conference series. The retrospective digitalisation of eCAADe conference proceedings can be represented here as a major effort and besides a recording of the e-papers, also a CD-rom with Digital Proceedings for the period 1983-2000 has been produced. Doubtless, an important source of “eCAADe-capital” has been made available in this way to the scientific community.

The current plans for further development of the CUMINCAD-repository have already been presented in the International Journal for Architectural Computing (Martens and Turk, 2003a). For example the extension to include theses and dissertations was discussed. Furthermore, the procedure for setting up a keywording system for automated classification of entries was explained, as this is planned to be another milestone in the future (see also Martens and Turk, 2003b). It has to be stated, that the functionality and the concepts behind CUMINCAD have been demonstrated at several conferences and the aim of this paper is not to duplicate these efforts again. However, the main scope of this contribution is to give an update on newly developed and implemented features and underutilized functions. Furthermore, a branch of less well-known self-organizing repositories has been set up, and this also in non-English language regions. Finally, the reader may become stimulated to organize further subject-specific repositories with Open Access by means of SOPS (SciX-Open Publishing Services). The setup and handling can be arranged on a shoestring budget and is, as a matter of principle, defined by an efficient use of (human) resources. Currently SciX is hosting digital libraries in the field of CAAD, Construction Informatics, Electronic Publishing and Environmental Studies.

## Options for Scientific Information Exchange

For a long time, scientific publishing remained largely in the hands of learned societies and similar, science-driven associations (Guedon, 2001). Publishing houses have been entering the market since the mid 19th century, but their role has been marginal and profits negligible until the 1960s, when the Science Citation Index (<http://www.isinet.com/>) was introduced and the number of participating universities around the developed world grew quickly. The business model of the publishing houses is a rather fascinating one. Scientists perform research activities, write papers, review their peers' work and edit scientific journals. The copyright on this work has to be given away for free to a publishing party that has not been taking part in the value-chain until the moment of publishing comes up. Within the SciX-consortium there is a strong belief, that waiving the right to copy (copyright) and distribute results of scientific work to commercial publishers hinders the efficient exchange of this information and makes scientific results harder and more expensive to obtain. At this point an emerging role for learned societies and professional associations can be identified, especially towards subject-specific archives.

The idea to use the Internet for scientific information exchange is not new. Existing solutions are of the following types:

- Preprint archives offer drafts of papers that have been submitted to publication in paper based journals. No quality control is provided. Often, the papers are quite similar to the final works published. Perhaps the best known archives in this context are the Los Alamos or the arXiv preprints archive.
- Electronic journals (eJournals) and magazines (eZines) which provide quality control mechanisms similar to paper based publications. 400 such journals are believed to have existed in 1999, including a Journal on Electronic Publish-

ing. Today this number is estimated at over 1000.

- On-line bibliographies are collections of papers (usually without full text) from a certain discipline. After having been published as a booklet for a number of years the abstracts are currently freely available through a database on the web. A well known example is the CiteSeer service offering full texts of some 2.5 million papers related to computer science. CiteSeer is accumulating the papers from the Web and copying them from authors' websites to one central location where they are classified, indexed and cross-referenced

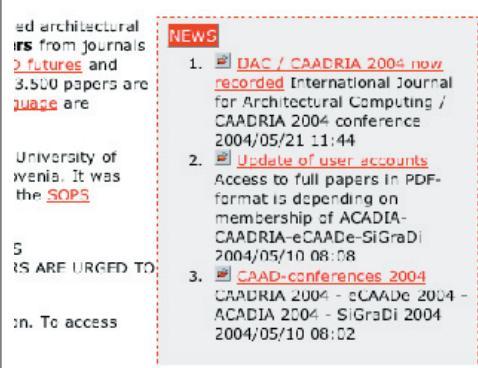
In the framework of the SciX project SOPS ("SciX Open Publishing Services" - <http://www.scix.net/sops.htm>) was created – a set of services from which various electronic publishing media, such as e-journals; conference review support; personal, subject-specific and institutional repositories etc., can be setup and efficient collaboration as well as knowledge management around it is being built. SOPS is multilingual and exists in English, German and Slovenian languages. Upon agreement of the software license, SOPS can be used for free. However, the customisation of a specialized SOPS-solution can not be realized for free, as manpower is needed. In most cases setting up a new service is not asking for specific input other than filling out a web form. Due to OAI-compatibility (OAI=Open-Archives-Initiative) data can be transported to any other service platform, if so required. The entry, manipulation and consolidation of data is handled in the sense of self-organization: activists take care, for example, of the recording of annual proceedings, whereas minimal training is necessary to get this operational. But also the contribution from individuals is useful (classical example: dissertation or thesis). In this respect publications are directly submitted in a mask by the proposer; the owner of the repository simply approves (or not) and takes care of the integrity. All these measures aim at the so-called "snowball-effect" as nobody will visit a repository without a critical mass of content again.

## CUMINCAD: "New" Features and Undertiled Functions

In the framework of the SciX-project, substantial extensions to CUMINCAD could be realized as this repository was run only on a shoestring budget before the start of the SciX-project. Early in 2004 both the database engine and the graphical web interface were renewed. For example the news-section was fully redesigned (fig. 1, <http://cumincad.scix.net/cgi-bin/news/>).

Also CUMINCAD.refs - a related Citation Index - (fig. 2a-b, <http://cumincad.scix.net/cgi-bin/refs/>), was linked up more closely to CUMINCAD. More than 20,000 references from previous CAAD conference proceedings have been recorded here. This means that users can follow the "impact" of a specific publication entry in terms of citations in other publications. Further continuing analysis, such as ranking according to the authors and publications most quoted would also prove useful and feasible (see for example Cerovsek and Martens, 2004). Repeating research already undertaken by others and the failure to evaluate is nothing new. The Citation Index hasn't been promoted too extensively as yet, but may be utilized as an important source for the measurement of impact in the future within CAAD-related research.

The search-engine within the SOPS-services is as



The screenshot shows a news section with a red dashed border. On the left, there is a vertical list of text: "ed architectural", "rs from journals", "2 futures and", "3.500 papers are", "image are", "University of", "venia. It was", "the SOPS", "5", "TS ARE URGED TO", "in. To access". The main content is a list of news items under a red "NEWS" header:

1. [DAC / CAADRIA 2004 now recorded](#) International Journal for Architectural Computing / CAADRIA 2004 conference 2004/05/21 11:44
2. [Update of user accounts](#) Access to full papers in PDF-format is depending on membership of ACADIA-CAADRIA-eCAADe-SiGraDi 2004/05/10 08:08
3. [CAAD-conferences 2004](#) CAADRIA 2004 - eCAADe 2004 - ACADIA 2004 - SiGraDi 2004 2004/05/10 08:02

Figure 1  
News section after CUMINCAD-relaunch.

easy to handle as for example the search-mask in “google”. If, for example, a user enters <design> over 4.000 records would be found. However, <title: design> would narrow this number down to roughly 2.000. But this is still a large number of entries and a further expression to narrow this down would necessarily have to be entered: <authors:brown title:design> (leads to 20 records from the authors named “Brown”). If a user wants to find out how many papers the author “Brown” has published within the series of eCAADe-conferences, he would enter <authors:brown series:ecaade>. The following expression (fig. 3) would show all the eCAADe-papers in the year 2003: <series:ecaade year:2003> However, with a growing numbers of papers, the use of AdvancedSearch-routines makes sense. The pull-down menus are self-explanatory; if someone is

searching the eCAADe-papers from 2000-2003 he/she would finally see the search expression “series: ecaade {year} > 2000 and {year} < 2004”, which is created by the AdvancedSearch-routine (fig. 4). Of course a user who is experienced with the search syntax can enter this directly in the “Search for:”-frame. In terms of output the “formatting” pull-down menu offers a variety of alternatives. For example the result of a search can be formatted directly into an endnote-format or in Harvard-style.

The interest in the creation of personal bibliographies, reading lists etc. may grow in the future. On the occasion of the retirement of Professor Tom Maver a large part of his work has been collected and stored into CUMINCAD. In some cases only a printed copy was archived, therefore this material was scanned and converted into a digital format.

Figure 2a-b  
Direct link to CUMINCAD.refs  
and example of a search in the  
Citation Index.

| #  | details | full text | citation  | check to select          |
|----|---------|-----------|---|--------------------------|
| 1. |         |           | Achten, Henri and Joosen, Gijs (2004) <b>The Digital Design Process - Reflections on a Single Design Case</b> , Digital Design [21th eCAADe Conference Proceedings / ISBN 0-9541183-1-6] Graz (Austria) 17-20 September 2003, pp. 269-274 | <input type="checkbox"/> |
| 2. |         |           | Akgun, Yenal (2003) <b>An Interactive Database (HizmO) for Reconstructing Lost Modernist Izmir</b> , Digital Design [21th eCAADe Conference Proceedings / ISBN 0-9541183-1-6] Graz (Austria) 17-20 September 2003, pp. 359-372            | <input type="checkbox"/> |

Figure 3  
Searching for eCAADe-entries  
in 2003.

The procedure for setting up this bibliography has been presented in the International Journal for Architectural Computing (Martens, 2004).

## Development of Multilingual Repositories

So far, CUMINCAD has been the centre of attention, but complementary repositories in other languages have also been set up in the framework of the SciX-project. While English may be regarded as the lingua franca of scientific communication, publication activities in other languages are important to strengthen local cultures, invent professional terminology, enable technology transfer to those who are not very proficient in English, such as practitioners

and students and to provide an incentive for authors to publish in their native languages.

CUMINCAD.ES (Spanish CAAD-related publications <http://cumincades.scix.net>) was set up in 1999 and serves the Spanish reading CAAD-community as an important source of archived scientific information. A first stage of development concentrated on a complete recording of CAAD-related Conference Proceedings (SiGraDi and ConVeACA), which included also summaries and full-text versions. Especially doctoral degree students benefit very much from this repository. At the time of writing approx. 700 records have been created and nearly 100% of these entries provide also a full-text version in pdf-format. In the sense of a self-organizing repository, the development of CUMINCAD.ES relies on the

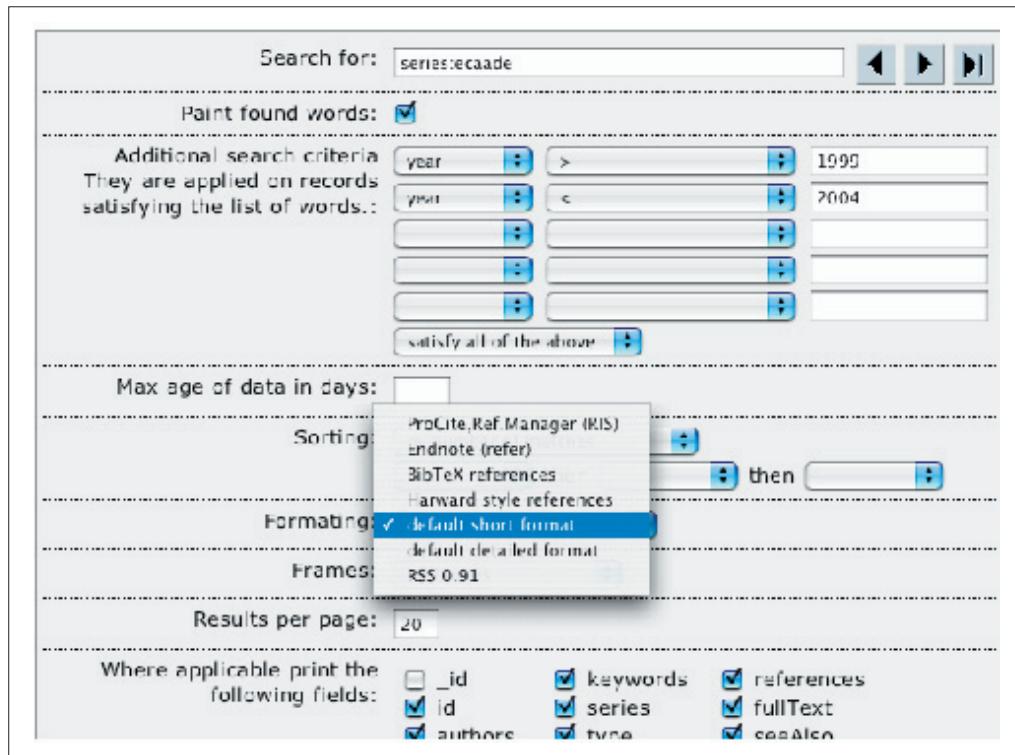


Figure 4  
Advanced Search for  
eCAADe-entries in 2000-2004  
and formatting options.

input from both individuals and associations. This repository is steadily growing and individual users are invited to record their bibliographies in the corresponding area (Martens and Turk, 2003c).

In a similar way Architektur-Informatik.SciX.net (fig. 5, German language publications in the area of CAAD - <http://www.architektur-informatik.org/>) and Raumplanung.SciX.net (dedicated to urban and regional planning) were created. Now where the SOPS-repositories start to grow like mushrooms, an overview was created: <http://www.scix.net/db/use/libraries/>. At the time of writing 12 repositories are listed.

## Conclusions

Thousands and thousands of papers have been published in conference proceedings. The reach of these publications and therefore the potential impact of the work reported is limited to the conference participants. Most of these proceedings are not available electronically or archived in libraries. It is in the interest of the associations and societies as well as of an individual author to make sure that

publications reach as many readers as possible.

What was clear from the very beginning was that the success of a digital library strongly depend on its contents, i.e. the availability of a critical mass. What was clear from the very beginning was that the success of a digital library strongly depend on its contents, i.e. the availability of a critical mass. This was realized pretty soon as far as CUMINCAD was concerned and publications were provided by the CAAD-associations (ACADIA, CAADRIA,eCAADe, SiGraDi), basically furnishing digital datasets of the most recent conferences, but also including retrospective digitalization (even way back to the first conferences held). Principally, electronic data procurement for the “younger” associations turned out to be easier, as datasets have been archived ever since the mid-nineties, being, however, dependent on the individuals concerned in many cases with their differing habits regarding data archiving. Some data carriers and software packages are no longer in use and thus corresponding computer equipment has also been eliminated.

A strategy to create a subject-specific digital library depends on good relationships with the associa-

The screenshot shows the 'ak.ai' digital library interface. At the top, it says 'Architektur-informatik : Digitale Bibliothek : Verarbeitung'. Below that is the search results section 'Suche Ergebnisse'. A search bar contains the text 'sonath'. Below the search bar, there are navigation icons and a 'Suchen' button. A small text box below the search bar reads: 'Titel-Ergebnisse - gebraucht erste Fahrzeug um nur Aufzeichnungen zu finden, welche das Wort 'sona', oder auch 'Fahrzeug' aber nicht 'gebraucht' enthalten. alle Titel sind HTML-orientierte Sucher'. Below this, it says 'Treffer 1 zu 17 von 17'. The search results are displayed in a table with columns: '#', 'details', 'full text', 'citation', and 'check to select'. There are three results listed:

| #  | details | full text | citation   | check to select          |
|----|---------|-----------|--|--------------------------|
| 1. |         |           | Beetz, Jakob (2003) <b>EIN PARAMETRISCHER ANSATZ FÜR DIE ECHTZEIT-INTERPRETATION VON HANDSKIZZEN IN DEN FRÜHEN PHASEN DES ARCHITEKTUR-ENTWURFS</b> , Diplomarbeit, Bauhaus-Universität Weimar Prof. Dr.-Ing. Dirk Dorath (Beurteiler/Gutachter) Jur. Prof. Dr.-Ing. Frank Pelzold (Zweitgutachter) | <input type="checkbox"/> |
| 2. |         |           | Dorath, D., Thurov, T. (2002) <b>VISION EINES MITWACHSENDEN GEOMETRIEMODELLS FÜR DIE BAUAUFNAHME</b> , VDI-Gesellschaft, Bautechnik(ed.): VDI - Bericht Forum Bauinformatik 2002 - Junge Wissenschaftler forschen, VDI Verlag, Düsseldorf, 2002, pp. 9 - 15  | <input type="checkbox"/> |
| 3. |         |           | Dorath, D., Thurov, T. (2002) <b>BAUPLANUNG IM KONTEXT DER DIGITALEN BESTANDSERFASSUNG - PLANUNGSSYSTEME IM BESTAND &amp; TECHNIKEN DER BAUAUFNAHME</b> , VDI-Gesellschaft Bautechnik(ed.): VDI - Bericht 1638-Bauen mit Computern Kooperation in IT Netzwerken, Düsseldorf, 2002, pp. 481 - 492   | <input type="checkbox"/> |

Figure 5  
Visual representation of the  
Architektur-Informatik  
repository.

tions, whose members are involved in teaching and research work. It is of the utmost importance that these associations consider such a repository as "their own". Such an approach is also contributing to the reversal of the process that took scientific publishing away from the professional societies and associations in the early 20th century.

#### Acknowledgements

The presented work has been conducted in the context of the SciX project, funded by the European commission under the contract IST - 2001 - 33127. SciX is a 24-month project with an EU funding of 1.000.000. Co-ordinated by the University of Ljubljana (Slovenia), the partners include the Swedish Business School of Finland, the Icelandic Building Research Institute, an e-Business company Indra/Atlante (Spain), Vienna University of Technology (Austria), the FGG Institute (Slovenia) and the University of Salford (UK). The home page of the SciX project is at <http://www.scix.net/>.

## References

- Cerovsek, T. and Martens, B.: 2004, On the Extended Use of Citations in CAAD. Making the network between related publications visible, in: Proceedings of the 7th DDSS conference [forthcoming]
- Guedon, J.C.: 2001, In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing, Proceedings, Association of Research Libraries, in Proceedings of the 138th Annual Meeting, Toronto, Ontario, May 2001 - <http://www.arl.org/arl/proceedings/138/guedon.html>
- Martens, B., Bjoerk, B.-C., Turk, Z.: 2002, Open, Self Organising Repository for Scientific Information Exchange - The SciX Project, in: eCAADe 2002 Conference Proceedings, Warsaw (Poland), pp. 222-229
- Martens, B. and Turk, Z.: 2003a, Cumulative Index of CAAD: Current Status and Future Directions, in: Int. Journal of Architectural Computing - Volume 1: - Issue 02: July 2003, pp. 220-231
- Martens, B., Brown, A., Turk, Z.: 2003b, Automated Classification of CAAD-related Publications: Conditions for Setting-Up a Keywording System, in: ACADIA 2003 Conference Proceedings, Indianapolis (USA), pp. 364-369
- Martens, B. and Turk, Z.: 2003c, CUMINCAD.ES: A First Step Towards Multilingual Digital Libraries in CAAD, in: SiGraDi 2003 Conference Proceedings, Rosario (Argentina), pp. 48-50
- Martens, B.: 2004, On the Archiving of Tom Mavers' Bibliography (1969- ), in: Int. Journal of Architectural Computing - Volume 2: - Issue 01: March 2004, pp. 77-88