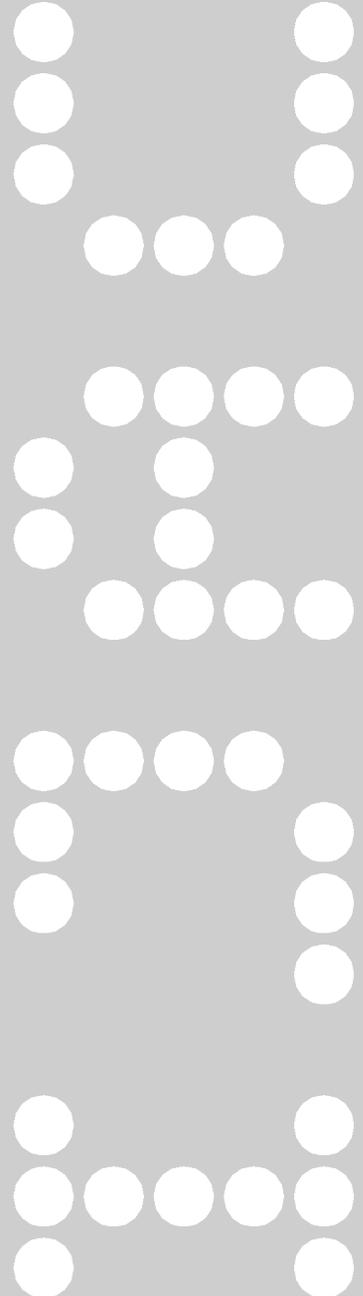# Cumulative Index of CAAD: Current Status and Future Directions

Bob Martens and Ziga Turk

Cumulative Index of CAAD: Current Status and Future Directions

Bob Martens and Ziga Turk

This article discusses the Cumulative Index of CAAD (CUMINCAD.SciX.net) – a digital library set up in 1998 serving the CAAD-community as an important source of scientific information. During the first stage, the metadata of CAAD-related conference proceedings were compiled and published on-line, including all abstracts and approximately 50% of the full-texts. In a subsequent step a Citation Index was created. Currently, theses and dissertations are being added to the library. Furthermore, a hierarchical topic structure was developed for automated classification of publications in the future, with topics being defined by keywords and characteristic papers. The next version of CUMINCAD, expected to be released later this year, will also feature a discussion forum, an event calendar, an option for commenting on and ranking publications as well as creating an on-line personal bibliographic review. CUMINCAD is a unique digital library in the field of CAAD serving a growing user-community. Younger doctoral degree students and junior researchers will benefit most from this edited, structured collection freely available via Internet.

1. Introduction

CUMINCAD – an acronym for CUMulative INdex of CAD / cumincad.scix.net – was created in 1998 by the authors, as response to limited, difficult access to scientific information in the field of CAAD. At this point associations such as ACADIA, CAAD Futures and eCAADe had already been in their second decade of existence whereas CAADRIA and SiGraDi had only been newly founded to serve the regions of Asia and Latin America more effectively. While the conference proceedings can be regarded as the only tangible result of an association's activity, the previous paper-based publications, with fewer than 250 copies, were practically unavailable. They became "gray literature".

CAAD Futures was the only exception due to the fact that publishing and distribution of proceedings were handled by a renowned publishing house, involving high purchasing costs of up to 600 Euro per conference volume. Thus access to a complete collection was only available in rare instances, as university libraries were not systematically collecting copies. Even if so, the interlibrary loan system, known to be considerably time-consuming and costly, would have applied.

► Figure 1. A screenshot of a search result in the CUMINCAD digital library

What was clear from the very beginning was that the success of a library like CUMINCAD would strongly depend on its contents, i.e. the availability of a critical mass. Publications were provided by the mentioned associations, who could furnish digital datasets of the most recent conferences, and retrospective digitalization (even way back to the first conferences) was attempted. Principally, electronic data procurement for the "younger" associations turned out to be easier, as datasets have been archived ever since the mid-nineties. However, archiving has been dependent on the individuals concerned, in many cases with differing habits regarding data archiving. Some data formats and software packages are no longer in use and corresponding computer equipment has been eliminated.

CUMINCAD: Paper 728f:
Grids in Design and CAD

home | search | advanced | announcements | new papers | authors | keywords | references | basket

authors Gross. Mark D.
year 1991
title Grids in Design and CAD
source Reality and Virtual Reality [ACADIA Conference Proceedings / ISBN 1-880250-00-4] Los Angeles (California - USA) October 1991, pp. 33-43
summary The grid is a useful device for expressing design rules about the placement of elements in a layout. By expressing position rules for elements in relation to a grid. a designer can systematically organize decisions in a layout design problem. Grids and placement rules offer a discipline that can help a designer work effectively to lay out complex designs, and it can also facilitate group design work. Unfortunately, computer supported drawing systems often cannot support this way of working because they lack a sufficiently rich implementation of grids. The Grid Manager module of the CoDraw program shows enhancements useful for architectural Computer Assisted Design. These enhancements would enable more effective ways of using the computer as a design tool.
series ACADIA
email [illegible]

check to select

details citation

Bier, E. (1990) Snap-dragging in three dimensions . Computer Graphics 24 (2): 193-204
Bier, E.A. and Stone, M.C. (1986) SnapDragging , Computer Graphics (SIGGRAPH 86) 20 (4): 233-240
Flemming, U. (1987) The role of shape grammars in the analysis and creation of designs , The Computability of Design. Edited by Y. Kalay. New York: Wiley
Gross, M. (1990) Relational Modeling . The Electronic Design Studio. Edited by M. Mc Cullough, W. Mitchell and P. Purcell. Cambridge, MA: Wr Press

◄ Figure 2. Screenshot of a specific record (including the citations it is referring to)

2. Related work

CUMINCAD is not the first attempt to provide indexed information related to CAD. Doug Noble and Karen Kensek indexed articles that appeared in ACADIA publications since its founding in 1981. Although the "ACADIA 15 Year Index 1981-1996" (http://www-rcf.usc.edu/~dnoble/author.html) is no more than a plain text list, it issues an overview and is often found by search-engines.

CADline is a bibliographic service developed by Yehuda Kalay [1] in cooperation with librarians at a time, where no web-based services were available. A proprietary system does not attract users in the same way as a web-based interface and furthermore, involves more preparation. The collected information, however, consisted of nearly 1.000 entries subsequently donated to CUMINCAD.

3. Scope of Information and Access Possibilities

Filling of the CUMINCAD-repository was performed on a shoestring budget. The associations' annual conferences set the stage for the marketing campaign to this end. Developments were covered in various media, such as a two-part report in ACADIA Quarterly [2, 3]. Thus the fundamental framework and the development directions had been determined.

Within the European Union's SciX project ("Open, self organising repository for scientific information exchange", http://www.scix.net, 2002-2004) [4], there are high expectations regarding CUMINCAD. The next version of CUMINCAD will act as a significant indicator in terms of digital library technology to be developed in SciX as well as of the re-engineering of the scientific publishing process. SciX is also providing hosting for digital libraries of other fields of science. Several communities such as the ELPUB

Conference series, The International Association for People-Environment Studies (IAPS), etc. are already using the SciX platform to implement their digital archives. Ironically, ELPUB stands for ELectronic PUBlishing, however the ELPUB-proceedings have not been published electronically. The "history" of ELPUB started in 1997. Nearly all papers from the early days of ELPUB are archived (on paper) and have recently been digitized to produce e-prints. Digitization from paper-based materials by means of scanning and OCR is much more labor-intensive than "reusing" archived digital data sources. The latter also allow for a more compact and high-quality reproduction including full color, which is not available in print form.

ELPUB Digital Library: works

Paper 9931:
PROLEGOMENA TO ANY FUTURE E-PUBLISHING MODEL

id	9931
authors	John W. T. Smith
year	1999
title	PROLEGOMENA TO ANY FUTURE E-PUBLISHING MODEL
source	Electronic Publishing '99. Redefining the Information Chain - New Ways and Voices : Proceedings of an ICCC/IFIP conference held at the University of Karlskrona/Ronneby, Sweden 10-12 May 1999/Edited by John W.T.Smith, Anders Ardö, Peter Linde. Washington D.C.: ICCC Press, 1999. 326 p. ISBN: 1-891365-04-5
summary	Considers why, after nearly twenty years of development, the electronic publishing model has not replaced the paper based model for academic journals. Gives three insights that attempt to explain this: the first is the 'means-end' confusion of commercial publishers, the second is the failure to realise it is the purpose not the form that is important about the current journal model, and third is the failure to recognise that a net-based replacement for the journal does not necessarily need a publisher. Finally some 'ground rules' to be used when thinking about, or designing, any future electronic publishing model are proposed.
keywords	New publishing models, electronic journals
series	ELPUB:1999
type	symposium
email	J.W.T.smith@ukc.ac.uk
content	file.pdf (179,429 bytes)
discussion	No discussions. Post discussion ...
ratings	very good(4) Rate it!
changed from	128.130.109.131
last changed	2003/04/11 15:44

ELPUB Digital Library: works

powered by WODA

▲ Figure 3. Screenshot of an ELPUB-record including links to discussions and ratings.

The previous situation of ELPUB regarding availability of e-papers situation is not unique [5] and basically applies to many other academic associations, which function mainly on the basis of volunteer work. IAPS, for example, has a much longer history than ELPUB as its activities started more than three decades ago. IAPS is aware of the fact that their presence particularly in Eastern-Europe could be improved and a digital library could serve as strong accompanying measure. The kick-off with a self-organizing repository has already been accomplished (http://iaps.scix.net) and the situation concerning full (back) digitization is being studied presently. In a first attempt, scientific contributions from 1996 onwards have been recorded, as far back as data in an electronic format existed to be gathered. Instead of searching individual proceedings one by one, an overview search can be performed by the end-user more conveniently.

Home | Publications | Students | Networks | Join IAPS | Links

IAPS Digital Library Menu

▸ home
▸ search
▸ new papers
▸ authors
▸ keywords
▸ announcements

About IAPS digital library

IAPS Digital Library

Search IAPS for: [ ] [Search]

How to contribute: Contributions are most welcome!

- Contributions of full text versions of the papers already in the library. To do so follow the link "correct" next to the paper in the library.
- Contributions of new individual papers. To do so, fill in this form. Note that full text must be uploaded as well and that these **contributions will need to be approved**, before they are adopted into the library.
- Mass contributions of new papers. Please read these instructions.
- Editorial interface and direct input (password protected).

Acknowledgements: IAPS was made possible by SciX project ... and ... and ... IAPS was inspired by CUMINCAD.

▲ Figure 4. Screenshot of the IAPS Digital Library

On completion of the above mentioned SciX-project, the commenced repositories are to be continued. CUMINCAD will be carried on in a similar way as in its very infancy (1998-2001). Thus integration of the new conference proceedings always plays a major role. As the required metadata usually are issued by the conference organizations in a predefined quality, this will only amount to slight efforts regarding subsequent editing and indexing the full-papers in pdf-format (portable document format). Subsequent editing of the references to be linked with the bibliographic entries can be more time-consuming, as datasets are to be divided into four information modules (author, title, source and year of publication).

3.1. CUMINCAD: What is Collected?

First a few figures as to quantity might be interesting. At present CUMINCAD contains 5.409 entries. This amounts to 6.8 megabytes of metadata excluding citations. 95% of all papers have an abstract. An increasing number of papers is available in full-text (total 2 gigabytes – currently 2.506 pdf-files). In the course of the time, we expect most entries will provide of a full-text version. The annual production of conference papers has been rising substantially in the last couple of years. Section 4 will deal with the collection process of relevant theses works in more detail. In 2001, 410 papers were published in conference proceedings, a number probably not growing substantially in the near future. The current growth on an annual basis – without the biannual CAAD Futures – is approximately 350 papers per year.

As a repository continuously grows, individual users are interested in having their bibliographies recorded. In some cases, perfect collections are already available in the web where the information required can be retrieved by download. The information offered may be also very sketchy, so that only directly contacting the authors proves effective.

The inclusion of e-prints and preprints from individual researchers makes sense and the storage has to be arranged in different categories, so that users of these information packages are well informed about the status of a publication. The minimum prerequisite for every submittal is the delivery of a summary; the preferred submittal also includes the corresponding full-paper in form of a pdf-file. A certain level of pre-structuring of the metadata can furthermore be regarded as useful. First attempts to this end have already been initiated and renowned authors have accepted the invitation for cooperation. We have to be aware of the fact that the "pioneers" from the early days of Computer Aided Architectural Design may still be active, but some are already considering retirement and the inclusion of materials should be secured.

3.2. CUMINCAD and Citation Index

The Citation Index (itc.fgg.uni-lj.si/db/use/cumincadref) was based on the archived full-papers of the conference proceedings. Each paper reference entry was divided into four parts and equipped with the CUMINCAD-ID of the specific publication, in order to provide a link to its display. Therefore, a database-user can see the authorship references on the present publication. By immediately accessing the Citation Index, publications quoted elsewhere can be spotted, which is also interesting for authors wanting to find out where their work is being quoted.

An evaluation of the approximately 20.000 references demonstrates that approximately 30% concern citations dating from 1996 or earlier. Slightly more than the half date from the period 1981-1995, while approximately 10% were created in the period of 1971-1980, with fewer than 10% of the entries were published before 1971. Further assessment, such as to ranking according to authors and publications most quoted would also prove wise and feasible.

The entries in CUMINCAD are comparably "younger" – no big surprise. Only 3% of the CUMINCAD entries concern publications prior to 1980. Two-thirds were published before 1995. The remaining 30% refer to publications dating from 1981-1995.

As a next step the entries of both databases were tuned to match each other, as importing of citations into the CUMINCAD-database might prove to be an effective extension of its information scope. Following various filtering processes, an expansion potential amounting to approximately 3.000 entries was specified. Nevertheless, these entries only contain bibliographic basics and the addition of abstracts and/or full-texts would surely be an

enhancement regarding the scope of information. Such an expansion would probably only cover a quarter of the specified 3.000 entries. A CUMINCAD-extension would not only add to quantity, but particularly to the quality of a successful search.

3.3. Access Policy

Even though the "open-access"-idea remained the authors' main focus, three classes of users were determined for CUMINCAD. Entering anonymously, only bibliographic information is disclosed, this being an incentive for registration as "friend" free of charge, as this status entitles one to viewing summaries and advanced research features. Retrieval of full-text versions – as available – had been restricted to the members of the CAAD associations (700 persons) so far. Thus the continuing contribution of full-papers in pdf-format was granted, and the associations could offer this as an additional service for valid memberships. At the time of writing over 1.300 "friends" are registered and we can assume that a significant segment of the scientific community throughout the field of CAAD will be reached this way.

Changing access towards open access (regarding the full-text versions) seems promising and could comprise all material being e.g. older than one or two years. Thus the CAAD-associations could offer their members exclusive access to recently issued publications as an additional bonus. Free-of-charge registration as "friend" is meaningful for recording of user behavior and for getting in touch with "key-clients". This all does not involve much additional effort on behalf of the users in their search for scientific publications.

3.4. Technical Aspects

The CUMINCAD digital library is a Web Application developed using the WODA (Web Oriented Database) rapid application development system. Nowadays several digital library software packages are available as open source (e.g. ePrints, DSPACE). However, at the time CUMINCAD started working, none of them existed. Presently, CUMINCAD aims at offering much more than just a digital repository. The relaunched CUMINCAD will be built on top of a series of Web services which will be combined in different ways to achieve different kinds of applications. A digital library is just one of them; others are an electronic journal, a personal archive and support of a conference or workshop organization.

WODA is an open source tool (www.ddatabase.com), developed at the University of Ljubljana. Applications like CUMINCAD are generated by WODA automatically, based on an extensive set of parameters defining *data structures, business logic* and the *graphical user interface*. Contrary to the other three-tier applications, WODA uses an object-oriented approach to capture all relevant information from all three tiers at one site. For example, the type of a field like "authors", its validation criteria and the way this field is rendered on various screens, is defined at one location.

Based on such definitions regarding information to be managed by an application and by a definition of common parameters(such as access rights, users, queries, formats etc.), pages responding to approximately one hundred different requests are generated on the fly. The most popular requests are *Show* (to display a record), *SearchForm* and *Search*, to find records.

WODA uses a search syntax similar to the industry-standard search engines and is quite fast. It stores data in a proprietary format directly on the host's file system and does not use an underlying database engine. The consequence is that the search speeds decreases linearly with the amount of data. However, as we have shown in this section, the quantity of CUMINCAD data is increasing linearly while hardware speeds are increasing exponentially.

As a web-oriented database, WODA is not a relational database. It works just like a relational database, for example, in defining a linked relationship between a paper and its references. Its Web characteristics come handy when some relations have not been pre-defined by a human or when it would prove difficult to do so – such as in Figure 2 showing paper 728f. Here, the references do not match with the paper being referenced, because these links have not been set up manually. Consequently, the small red search icon allows the user to search CUMINCAD for a paper with a title most similar to the one listed.

The SciX project will, as shown in the ELPUB screenshot (Fig. 3), extend the digital library concept with numerous add-ons such as discussion forums, ratings, annotations and reading lists. It will also enhance the search with full-text searching, similarity measurements, user profiling and automatic clustering and classification. The basis for the latter is described in Section 5.

4. Extension of Cumincad with Theses and Dissertations

By its very nature a thesis or a dissertation is produced in limited quantities, and in many cases the only copy available is the archival copy deposited in the library. In the course of a request, cost and delay factors are a significant deterrent. The lack of usage depends on a number of factors, such as knowledge of the existence, the contents, or the availability of the thesis or the dissertation. The relatively restricted access to print-theses is the main reason for their under-utilization. Making the full-text available from any computer desktop across the web would greatly increase knowledge, access and availability of such a meaningful resource.

Enterprises such as *University Microfilms* provide of a large collection, but mainly collect US-dissertations. On the other hand individual libraries have different ways to make these academic "products" visible, but the outcome is like a widespread, mosaic landscape of knowledge. Archiving and distribution of Electronic Thesis or Dissertations (so-called ETD's, e-Thesis or e-Dissertation's) can be regarded as a chance to make this scientific output available.

The situation in Europe concerning repositories with dissertations and theses lacks any similar (central) "Union Catalog" where this kind of academic work can be searched. In fact the user has to visit individual university libraries one by one and to get acquainted with a variety of interfaces. This is time-consuming and makes sense only if a specific piece of work has to be retrieved (i.e. university and author are known). In this case particularly "younger" dissertations and theses have a good chance to be already available in pdf-format (as digital data may be archived), and can be retrieved for free.

Concerning the "production" of CAAD-related dissertations, no exact figures as those issued for the conference proceedings exist. A closer look at the full collection of the Faculty of Architecture at Eindhoven University of Technology – which provides of a complete set of doctoral e-dissertations (90 since 1974) – shows that a smaller number (4) focus on topics related to CAAD. The number of architectural education sites in Europe (approximately 250) could be used as an indicator. For various reasons, not every location has the same research output in terms of defended doctoral dissertations. The total number of architectural education sites in a worldwide perspective probably does not exceed the number of 1.000. The same number probably accounts for those individual researchers explicitly dedicating their work to CAAD.

Especially in Asia, a remarkable number of Ph.D.-students are currently working on dissertations (as observed by the submission entries for review in annual conferences). An estimation for the Nordic countries in Europe concludes that approximately 15 Ph.D. theses in Construction IT and CAAD were defended since 1995. A worldwide figure of 25 dissertations per year and a total of 250 since approximately 1990 seems to be reasonable.

Further analysis of both the dissertation indexes as well as references from conference papers assisted in the creation of a new type of series within CUMINCAD. Nearly 200 entries have been added so far in this new series and approximately a third includes full-text versions. In this respect, the support of the CAAD-community – both authors and supervisors – worked out well. Some scientists were even willing to recreate an electronic version of their previous work.

5. Extension of Cumincad with Automated Classification

CUMINCAD consists of a couple of thousand records and still growth can be envisioned. This means that a framework for the organization of stored knowledge is called for. A traditional library handles classification manually, but this is rather cost-intensive as well as time-consuming. As we are dealing with a specific field of knowledge, we may assume that those looking for information have previous knowledge and thus will achieve a satisfactory search result. Accounting for the fact that ten thousand printed pages are archived in CUMINCAD, one can hardly expect that the documented scientific knowledge can be handled autonomously. Using advanced search possibilities will make for a rather sound result, as various fields of the database entries can be investigated

simultaneously. Not all contributions issue keywords submitted by the author and some author-submitted keywords might not issue the proper clue.

Due to the specified reasons, we decided to take a closer look at the possibilities of a hierarchical system of topics, defined by keywords and prototype papers. Putting to use the co-author's positive experience with such a system for the field Construction Informatics [6], a similar procedure for CAAD was developed. The range of tasks was limited to two major measures: compilation of a keyword category list and selection of up to 10 relevant entries from the database. Hierarchical browsing of no more than 30 keyword categories was desired, which required the definition of 5-10 subtopics each per keyword.

First of all a provisional list was established based on the frequency of entries in the CUMINCAD-database field "keywords":

- 2D Representation
- 3D City Modeling
- 3D Modeling
- Animation
- Artificial Intelligence
- Case Based Reasoning
- Collaborative Design
- Communication
- Computer Integrated Construction
- Constraint Based Design
- Database Systems
- Design Methodology
- Design Process
- Digital Design Education
- Digital Media
- Environmental Simulation
- Generative Design
- Human-Computer Interaction
- Image Processing
- Interactive Design
- Knowledge Modeling (KM)
- Learning Environment
- Object Oriented Modeling
- Performance Simulation
- Shape Grammars
- Virtual Design Studio
- Virtual Environments
- Virtual Reality
- Visualization
- Web Design

Regarding the definition of related subtopics and typical publications, expert-help within the CAAD-community was requested. In order to avoid additional workload, the experts did not see the whole set of keywords, but only one category. As this is work in progress, the inquiry is not yet completely finished. To give an impression of the result, one example is displayed here:

- Collaborative Design
    - Collaborative Teamwork
    - Collective Authorship
    - Computer Supported Collaborative Design
    - Computer Supported Collaborative Work
    - Distributed Modeling
    - Distributed Workgroups
    - Groupware
    - Groupwork
    - Information Sharing
    - Multi-User Workspace

The provisional list is covering a vast amount of information in order to feed the bibliographic database. Naturally a specialized field like CAAD does not consist of fixed topics and thus the system introduced above is to be adjusted in line with current developments. On completion of a "Beta version 1.0", tests with clustering processes can be initiated. The user community might be particularly interested in hints as to comparable publications.

6. Conclusion and Outlook

A strategy to create a complete and relevant topical digital library depends on good relationships with the associations, whose members are involved in research work and wrote the publications. It is of 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.

CUMINCAD supports the important task of information management in the area of Computer Aided Architectural Design, as no similar activities have taken place so far. An annual entry of about 500 "new" publications seems feasible and would be accompanied by full-text documents for the major part of the records. Further research on the stored content may follow and would principally allow for an innovative analysis of the information.

This article attempted to describe current extensions for CUMINCAD such as community building, dissertation additions, citation indexing and

hierarchical topic browsing. First steps were also made to address multilingual content in languages other than English. WODA, as the user interface language, is multilingual. Services such as CUMINCAD.ES, CUMINCAD.DE, CUMINCAD.FR etc. – including machine translation between major languages – will bring CAAD research even closer to people.

In terms of quality management, CUMINCAD will be steered by an advisory board that will include key persons from the fields concerned.

Acknowledgements

References

1. Y.E. Kalay, CADLINE, *A Bibliographic Database on CAAD-Related Publications*, 1991, State University of New York, Buffalo.

2. Martens, B. and Turk, Z., The Creation of a Cumulative Index on CAD: "CUMINCAD", *ACADIA Quarterly*, 2000, Vol. 3, No. 19, pp. 18-19.

3. Martens, B. and Turk, Z., Work in Progress on CUMINCAD, *ACADIA Quarterly*, 2000, Vol. 4, No. 19, pp. 25-26.

4. Martens, B., Björk, B.-Ch. and Turk, Z., The SciX Project: Re-Engineering from Paper-based to Free Electronic Publishing, Thresholds – Design, Research, Education and Practice, *Proceedings of the 2002 Annual Conference of the Association for Computer Aided Design In Architecture*, Pomona (California), pp. 24-27 October 2002, pp. 179-185.

5. Björk, B.-C., Turk, Z., How Scientists Retrieve Publications: An Empirical Study of How the Internet Is Overtaking Paper Media, *Journal of Electronic Publishing*, Michigan University Press, 2000, Vol. 2, No. 6, http://www.press.umich.edu/jep/06-02/bjork.html.

6. Turk, Z., Cerovsek, T., Mapping the W78 Papers onto the Construction Informatics Topic Map, *Proceedings of the 20th CIB W78 Conference on Information Technology in Construction Waiheke Island*, Auckland, New Zealand, 23-25 April 2003 (forthcoming).

Bob Martens[1] and Ziga Turk[2], [1]Vienna University of Technology, Karlsplatz 13, A-1040 Vienna, Austria, [2]University of Ljubljana, Jamova 2, SI-1000 Ljubljana, Slovenia

[1]b.martens@tuwien.ac.at
[2]ziga.turk@itc.fgg.uni-lj.si