

A DIGITAL LIBRARY FOR ELPUB PROCEEDINGS: THE USE OF A WEB-BASED PROTOTYPE

BOB MARTENS¹; PETER LINDE²; ZIGA TURK³

¹ Vienna University of Technology, Faculty of Architecture and Regional Planning
Karlsplatz 13, A-1040 Vienna, Austria
e-mail: b.martens@tuwien.ac.at

² Blekinge Institute of Technology, The Library
S-37179 Karlskrona, Sweden
e-mail: peter.linde@bth.se

³ University of Ljubljana, Department of Civil and Geodetic Engineering
Jamova 2, SI-1000 Ljubljana, Slovenia
e-mail: ziga.turk@itc.fgg.uni-lj.si

This paper presents the initial steps in the creation of a digital library of previous ELPUB proceedings. The series of ELPUB conferences started in 1997, but due to different reasons the proceedings appeared solely in a paper-based format. Six conferences on electronic publishing resulted in an interesting output of scientific papers and it makes sense to have this wider and easier available to the audience. The currently used interface of the ELPUB Digital Library was created in the framework of an EU-IST project, called SciX (Open Scientific Exchange of Information). First results on the gathering of initial content will be presented as well as related experiences. Furthermore other DL-prototypes - which were also developed in the context of the SciX-project - will be described as examples of interest. Finally, thoughts on future developments of the ELPUB Digital Library will be displayed.

Keywords: Digital Preservation; Scientific Publishing Model; Re-Engineering; Metadata Use

INTRODUCTION

The "history" of ELPUB conferences goes back to 1997 venue at Kent University in Canterbury, UK. Since then about 200 papers have been published in conference proceedings [1,2,3,4,5,6]. ELPUB stands for Electronic Publishing, yet the tangible output of ELPUB conferences - the proceedings - was not published electronically, but solely on paper. One of the reasons for this is definitively the copyright issue, another quite certainly, a lack of simple, available tools that would allow for the creation and maintenance of such publications. Tools that simplify the creation of such infrastructures are the goal of the SciX project (<http://www.scix.net/>). SciX is demonstrating that the Internet enables new business models for the scientific publishing process, which are much more cost and time efficient to the scientific community than the current practice. In addition, existing publishing practices are analyzed systematically and business models for re-engineering the scientific publishing process will be developed, taking into account also the legal, social and psychological barriers to change. When the SciX project was presented at ELPUB 2002 [7], links between the ELPUB community and the SciX project were established, that are resulting in the ELPUB Digital Library, freely available on the Web.

Nearly all data from the early days of ELPUB has been archived in a digital format, with the exception of the 1997 proceedings. E-prints were produced by means of scanning and then applying OCR to the scanned images for the creation of metadata (authors, title and summary) and this is quite labour intensive. The "reuse" of archived digital data sources, which may be in general only available to more recent publications, is more efficient as additional human effort is minimal. Added benefits are high quality images and possible animations that are impossible to reproduce in a black-and-white print. In order to allow for as much re-use of existing digital material, SciX is not only aiming at tools that support digital libraries, but will support large parts of the publication process including the support for a conference organisation, reviewing, etc. The result of SciX will be a set of ready to run applications (digital library, electronic journal, conference) and a set of web services (repository, review, discussion, annotation, etc.) that can be regarded as building blocks, out of which the mentioned or new web based applications could be created.

SciX is currently at a stage where prototypes of these applications are being developed and tested against selected user groups. The ELPUB digital library is one such prototype. These prototypes will evolve into open source applications and services towards the end of the project.

OPEN REPOSITORIES AND SELF-ORGANISATION

SciX stands for "Open, self organizing repository for Scientific information eXchange". The term "open" in this context has several meanings:

- **To users, SciX repositories are open access.** SciX is paying little or no attention to providing features that would close the repository to certain groups of users. SciX does allow for features through which known users have more functionality available to them, than anonymous users, but generally, the items in the SciX archives should be open for all.
- **To programmers SciX software is open source.** As a matter of principle, SciX is placing all its development into an open source license. The SciX code can be reused in the following ways:
 - open source applications can be implemented with or without modifications to provide customized applications.
 - open source services can be implemented (using XML calls via HTTP) by third party applications. Services provide a rather complete set of functionality related to an important business object (such as a paper, a review, a person, etc.). Furthermore, because services are open source, they can be augmented and customized.
- **To (web) librarians, SciX archives appear as open archives.** Access to SciX repositories is possible through the OAI-PMH protocol. SciX repositories can be harvested by the OAI-PMH compliant harvesters. Also, a module related to the SciX repository service allows for the harvesting of foreign OAI-PMH compliant archives and their inclusion in SciX archives.

The concept of self-organisation refers to the key idea that repositories would not be maintained by professional publishers, but by organisations that do not have a commercial interest in managing the repository (universities, professional associations, individuals, etc.) The repositories should be created in such a way that they could be managed with a minimum of human effort. A major effort in managing these repositories is the organisation of the content - how to group papers into categories, how to create hierarchical directories, like Yahoo, across the papers in the repository, etc. Not unimportant is the management of the users and maintenance of the consistency of the data. An important aspect of SciX is applying text-learning techniques to automatically categorize the content. However, self-organisation means that decisions towards data entry and extensions of available information are not taken by the SciX-consortium, but remain in the hands of the association in charge. Furthermore, a splitting of workload can take place among a number of collaborating activists, so that an accumulation of smaller tasks is feasible. The criterion of "ease to use" has to be mentioned as a prerequisite, both for the maintainers as the endusers. Proprietary software packages normally do not meet this criterion; data entry and viewing have to preferably be performed from any internet browser and furthermore, a minimal level of computer literacy should be required. In many cases the editing of the data is being handled by non-librarians, which means that data will not necessarily be imported into the Digital Library as "clean" and "proper" as a library would prefer. Maybe from a librarian's point of view a "quick and dirty" approach is not preferred, but in this context for the enduser, hardly less effective. Furthermore, as the full text version is attached to the record the justification of the need for an ultimate precise citation is narrowed.

The content providers behind a repository are responsible for what kind of information they input and how detailed or secured the pieces of information are in the projected repository. This does, for example, not cause serious problems in an academic context as long as reviewed publications (i.e. conference proceedings) are recorded. However, individual submissions with a variety of publication sources may lead to the installation of an advisory board who supports decisions concerning the relevance of entries, insurance of quality control, topical relevance, etc. For the moment, this is not planned by ELPUB, as the inclusion of previous proceedings has the highest priority. For the ELPUB association, a future model of self-organisation could well be that the person responsible for the editing of the proceedings also would be responsible for entering the bibliographical information, preferably including the full texts of each accepted conference paper. This should be done before the paper version of the conference proceedings has been printed.

The practical reason for producing paper-based proceedings is mainly twofold: It is handy for the delegates to have during the conference and it is a means of provide libraries and database vendors with information for further dissemination. However, all materials are available in a digital format and allow for a "produce once and publish twice" procedure. The reason why ELPUB papers have not been made available as eprints so far is simply that, without any other impetus than pure idealism, activities have been focused on organising and realizing future conferences. There has been no energy left to create a digital database for dissemination of the papers produced. And this is of course the reason why the ready to use SciX infrastructure was so attractive to the organisers of ELPUB. But at this point in time a rather stable network of people are connected to ELPUB. They will be the base not only for organising future conferences but also for dissemination

of the intellectual results of the organising efforts in a appropriate digital way. Afterall, we should try to live as we preach!

Talking about self-organisation, the "problem" of copyright has also to be regarded. The two latest ELPUB conference proceedings have been published by commercial publishers: IOS Press (Amsterdam) and Verlag für Wissenschaft und Forschung (Berlin). This means that ELPUB now must seek the publishers permission before eprints of the conference papers are republished. So, before these deals are successfully negotiated, only the bibliographical data can be entered for the records of ELPUB 2001 and 2002 in the database.

Having access to a digital publishing tool as presented by the SciX consortium means that the organisers of future conferences in the scale of ELPUB must consider if there is any need for a commercial publisher at all. The two reasons for printing proceedings in paperformat - providing the delegates with a working tool as well as providing libraries and database vendors with interesting information for further distribution, means that you only need to print and distribute around 100-200 copies of the proceedings. This cost could quite easily fit into a conference budget and publishing could be done, as with earlier conferences, by the organisation that is sponsoring the conference (in the ELPUB case of the ICCG).

So for proper self-organisation you need to have full control of your data as well as a organisational agreement on the administrative routines for entering data. This is in fact a very simple thing to administer since the web-based interface to the ELPUB database is intuitive to learn. This is vital for an organisation such as ELPUB which is geographically distributed over the whole world and consists of people from different cultures and language areas.

SETUP OF A PROTOTYPE FOR ELPUB

The creation of a Digital Library in the sense of the SciX-project did not cause an immense workload. However, in the process of setting up a repository, the demand for feedback and definition of a corresponding wishlist is given as a couple of decisions have to be taken. For example the visual representation of the database can be based on an existing corporate design scheme (layout styles etc.) of an association. In the case of the current version of the ELPUB-repository, not all decisions have been taken, such as the visual representation, but this can easily be implemented at a later stage.

It does not make sense to offer solely bibliographical data. Preferably abstracts (summaries) and keywords are to be entered in the records as well as email addresses of the author(s). The field "type" proves useful to identify the paper more clearly. Also most ideally, the corresponding full text (in pdf-format) is attached to an individual record. In case a possible interference with the revenue of selling conference proceedings is expected, a plain textversion could serve as a solution to make the difference with the printed counterpart clear. The argument of "commercial success" concerning sellings of proceedings may be on one hand important for a local organizers to earn a revenue for the investments. On the other hand, in terms of impact and visibility, a wide (if not the widest) dissemination seems to be of high interest to an association and leads to potential increased recognition of the authors. Search machines are able to scan the content of unrestricted pdf-files and can therefore be included in queries. A well filled repository incorporates a chance for newcomers to inform themselves in a ubiquitous way. The literal "reinvention of the wheel" is by intention not forbidden, but could be avoided in many cases. Also for the review of submissions, reviewers can consult the Digital Library in order to discover duplicating efforts. The same applies to future organizers who can derive on a concise overview on previous themes and topics.

Questions regarding "open access" initiate extended discussions as this is the point where many formally established associations come up with the issue of "added value" for their members. Indeed from the point of view of policy development of an association this is understandable as the membership is interested to know, "what do I get back for my membership fee". Therefore the accessibility of, for example, full papers can be limited to members to make the advantage of membership clear. However, this is not in line with the vision of "open source/access-movement". A reasonable solution can be seen in the following treatment: If a repository offers full access to all available information, no costs will be charged for setting up and hosting. In case of built-in limitations, additional work will be created and as the associations receive revenues, a sharing of finances with the SciX-host will take place.

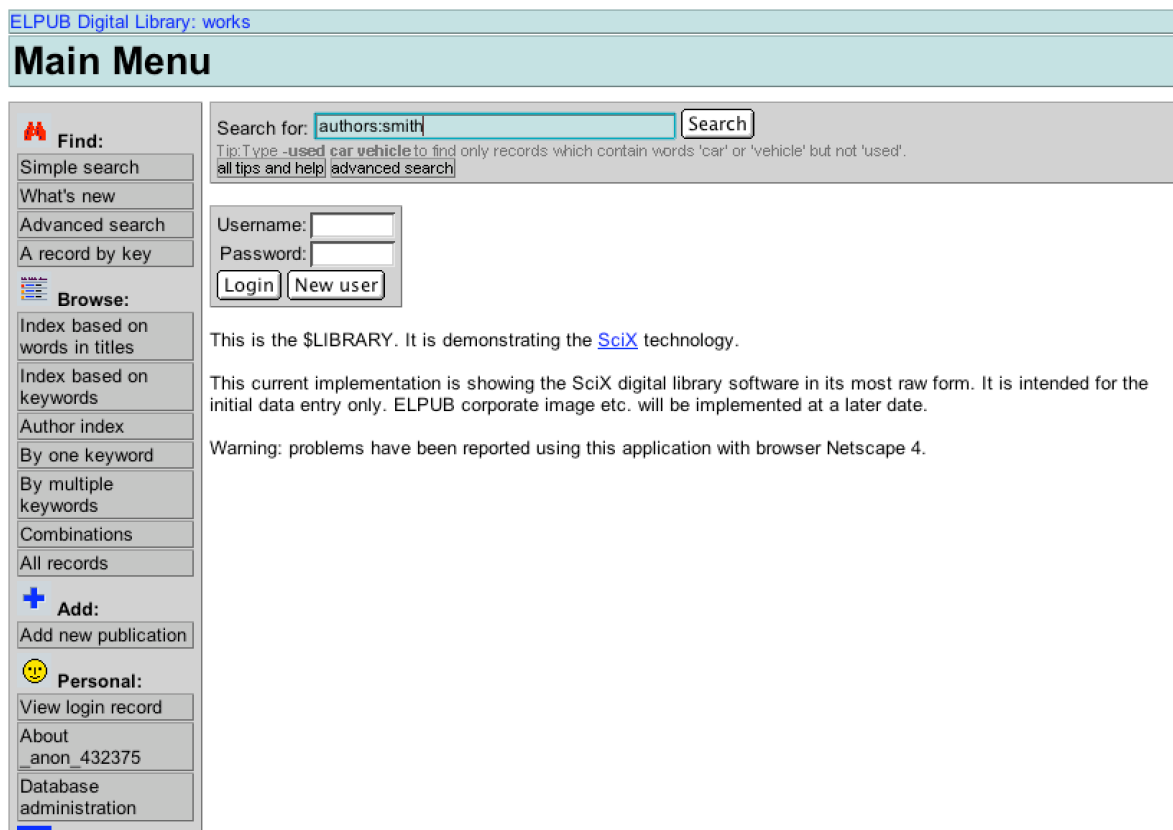



FIGURE 1 – SCREENDUMP OF ELPUB.SCIX.NET INDEX-PAGE. THE SEARCH IN THIS EXAMPLE IS LIMITED TO THE FIELD <AUTHORS>. MORE COMPLEX QUERIES ARE SUPPORTED IN ADVANCED SEARCH (SEE FIG. 6).











FIGURE 2 – THE RESULT OF THE QUERY FROM FIG. 1 LEAD TO ONE CITATION. A FULL PAPER VERSION IS AVAILABLE AS THE PDF-ICON INDICATES AND FURTHER DETAILS CAN BE CONSULTED (SEE FIG. 3).

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 ICCO/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.: ICCO 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 

FIGURE 3 – THIS SCREENSHOT SHOWS THE FULL RECORD (WITH SUMMARY) AND AN OPTION FOR DISCUSSION AS WELL AS RATINGS CAN BE ENTERED BY A VISITOR.

ELPUB Digital Library: works

Browse by [series](#) : ELPUB









Subcategories ([more](#)) re-order: [by count](#), [by reverse count](#), a-z, [predefined](#).

1. [ELPUB:1997 \(0\)](#)
2. [ELPUB:1998 \(0\)](#)
3. [ELPUB:1999 \(36\)](#)
4. [ELPUB:2000 \(25\)](#)
5. [ELPUB:2001 \(0\)](#)
6. [ELPUB:2002 \(0\)](#)
7. [ELPUB:2003 \(0\)](#)
8. [ELPUB:2004 \(0\)](#)
9. [ELPUB:2005 \(0\)](#)

Found items in ELPUB ([more](#)):

details full text citation check to select

-

ELPUB Digital Library: works


powered by 

FIGURE 4 – A CLICK ON THE HYPERLINK <SERIES> FROM FIG. 3 DELIVERS THIS SUMMARISATION.

ELPUB Digital Library: [works](#)

New publication

Show less help Display manual

id: Unique ID of the paper (Must be non-empty)

authors: All authors of the paper (Letters, dashes, spaces, semicolons. No dots!)

year: Year when this was published (4 digits.)

title: Full title of the publication or paper. Sentence case is suggested. (Starts with a capital and followed by up to 100 characters.)

source: Where was it published

summary: Brief summary

keywords:

series: If this was published in a series, select one:

type: Is this document of a special type:

FIGURE 5 – THE MASK FOR ENTRY OF INDIVIDUAL RECORDS ALLOWS FOR A "COPY-AND-PASTE-INPUT". FOR THE IMPORT OF MULTIPLE ENTRIES ALSO STRUCTURED CSV-FILES CAN BE UTILIZED.

ELPUB Digital Library: [works](#)

Advanced Search

Search for:

Paint found words: ☒

Additional search criteria

They are applied on records satisfying the list of words.:

satisfy all of the above

Max age of data in days:

Sorting: by number of matches

Formatting: default short format

Frames: no frames

Results per page: 20

Where applicable print the following fields:

id ☒ summary ☒ content ☒

id ☒ keywords ☒ discussion ☒

authors ☒ series ☒ ratings ☒

year ☒ type ☒ from ☐

title ☒ email ☒ at ☐

source ☒ more ☐

Search! Save form as

FIGURE 6 – THE FEATURE OF ADVANCED SEARCH ALLOWS FOR A MORE PRECISE FILTERING OF DATA.

RELATED PROTOTYPES

The previous situation of ELPUB regarding availability of e-papers situation is not unique and applies in principal to many other academic associations, which function mainly on the basis of volunteering work. For example the *International Association for People-Environment-Systems* (IAPS - <http://www.iaps-association.org/>) has a much longer history than ELPUB as activities started more than three decades ago. IAPS

is aware of the fact that their presence in especially Eastern-Europe could be improved, whereas a Digital Library could serve as strong accompanying measure. A start with a self-organizing repository has been made in the meanwhile (iaps.scix.net) and issues concerning full (back-) digitisation are currently being investigated. In a first round, scientific contributions from 1996 on have been recorded, as data in an electronic format could be gathered. Instead of searching individual proceedings one by one, an overview search can more conveniently be performed by the end user.

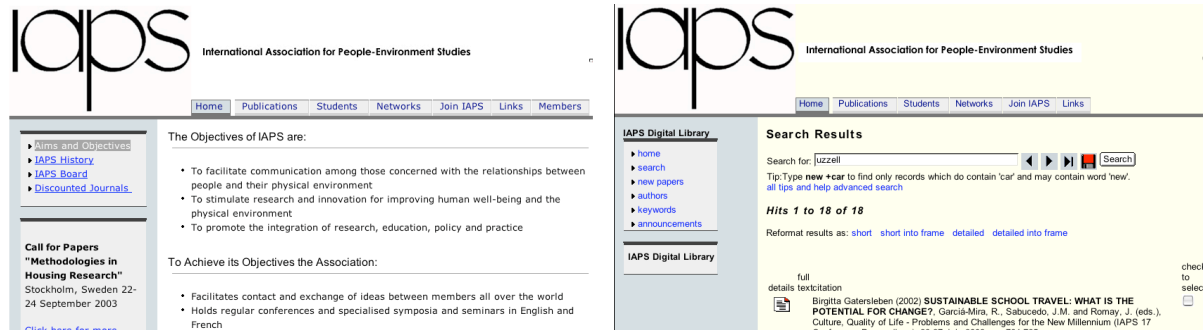


FIGURE 7A-B IAPS HOMEPAGE (LEFT) AND IAPS DIGITAL LIBRARY (RIGHT) - ADAPTATION OF VISUAL REPRESENTATION.

Another well-established database stemming from the SciX-project is *CUMINCAD* (CUMulative INdex on CAD - cumincad.scix.net). The first stage of development concentrated on a complete recording of CAAD-related Conference Proceedings. At present CUMINCAD features 5.394 entries and an increasing part of the entries also issue full-text versions (2.493). More than 90% of the entries also provide an abstract. The majority of entries, however, concentrates on papers from CAAD-related Conference Proceedings.

If a CUMINCAD-user enters anonymously, only bibliographic information is disclosed, this would be an incentive for registration as "friend" free of charge, as this status entitles the user to view summaries and access advanced research features. 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.

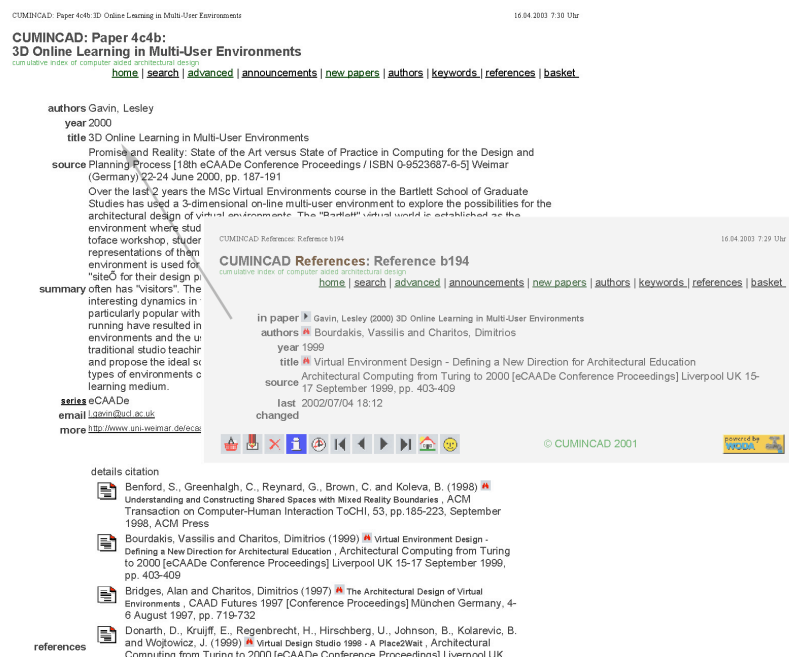


FIGURE 8 – CUMINCAD: RELATED DATABASE OF REFERENCES (ITC.FGG.UNI-LJ.SI/DB/USE/CUMINCADREF/)

CUMINCAD was recently extended with a related database of references (approx. 20,000 entries) recreated from the gathered content. Authors can now see who is citing his/her paper and for example a "TOP 25" of most influential publications can easily be created. Dissertations and theses in the field have also been extracted from this source.

What was clear from the very beginning was that the success of such a repository like CUMINCAD would strongly depend on its contents, i.e. the availability of a critical mass. This prerequisite was realized pretty soon as far as CUMINCAD was concerned and publications were provided by the associations.

FUTURE DEVELOPMENTS

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 is attributed to a number of factors, such as knowledge that the thesis exists, the contents of the thesis or dissertation availability. The relatively restricted access to print theses is the predominant reason for their under-utilisation. Making the full-text available from any computer desktop across the web would greatly increase knowledge, access and availability of such a significant resource. An extension with theses and dissertations is feasible within the elpub.scix.net-repository. In this respect the support of the ELPUB-community - both authors and supervisors - can be envisioned. It can furthermore be assumed that researcher are willing to (re-) create an electronic version of their previous work.

Furthermore, the procedure for setting up a keywording system for automated classification of entries should be studied. A traditional library handles classification manually, but this is rather cost-intensive as well as time-consuming. Accounting for the fact that thousands of printed pages will soon be archived in elpub.scix.net, one can hardly expect that the documented scientific knowledge can be handled autonomously. Using advanced search possibilities will make for a pretty sound result, as various fields of the database entries can be investigated simultaneously. The range of tasks can be limited to two major measures: compilation of a list of no more than 30 keywords with hierarchic ranking so required and definition of 5-10 subtopics each per keyword and selection of up to 10 relevant entries from the database. Naturally a specialized field like Electronic Publication does not consist of cemented topics and thus the systematics introduced above is to be adjusted in line with current developments. On completion of "Betaversion 1.0" tests with clustering processes can be initiated.

CONCLUSIONS

In this paper the conditions for setting up a digital library for ELPUB were elaborated and first results presented. The infrastructure used is currently provided by the SciX-consortium, and allows for the creation of OAI-compatible data. It can be stated that entering metadata did not cause major problems at all. Furthermore the initial work in terms of preparation, etc. was limited to a couple of emails with the previous conference hosts. The collection of all previous conference papers can be regarded as a first step towards content analysis. Soon ELPUB will celebrate the tenth anniversary and questions like "Where do we go from here?" will be posed. The number of records in the ELPUB Digital Library will hopefully soon provide a critical mass of recorded publications, which will enhance the authority of the repository and thus attract scientist to record further relevant publications in the field of electronic publishing.

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 Swedish Business School of Finland, Icelandic Building Research Institute, an eBusiness company Indra/Atlante (Spain), Vienna University of Technology (Austria), FGG Institute (Slovenia) and the University of Salford (UK). The home page of the SciX project is at <http://www.scix.net/>.

REFERENCES AND LINKS

1 Electronic Publishing '97. New Models and Opportunities. Proceedings of the 1st International ICC/IFIP Conference on Electronic Publishing held at the University of Kent, Canterbury, UK, April 1997. Edited by J.W.T Smith. ICC Press 1997. ISBN 1-891365-00-2.

2 Electronic Publishing '98. Towards the Information-Rich Society. Proceedings of the 2nd. International ICC/IFIP Conference on Electronic Publishing held at the Central European University in Budapest, Hungary, April 1998. Edited by Fytton Rowland and John.W.T Smith. ICC Press 1998: 286 p. ISBN 1-891365-02-9.

3 Electronic Publishing '99. Redefining the Information Chain - New Ways and Voices : Proceedings of the 3rd ICC/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.: ICC Press, 1999. 326 p. ISBN: 1-891365-04-5

4 Electronic Publishing 2000. Electronic Publishing in the Third Millennium: Proceedings of the 4th ICC/IFIP conference held at Kaliningrad/Svetlogorsk, Russia August 17-19 2000/Edited by Peter Linde, John W.T.Smith, Elena Emilianova. Washington D.C.: ICC Press, 2000. 239 p. ISBN: 1-891365-07-X

5 Electronic Publishing '01. 2001 in the Digital Publishing Odyssey : Proceedings of the 5th International ICC/IFIP Conference on Electronic Publishing held at the University of Kent, Canterbury, UK, July 2001. Edited by A.C.Hübner, P.Linde and J.W.T Smith. Amsterdam: IOS Press 2001: 330 p. ISBN: 1-58603-191-0.

6 Electronic Publishing '02. Technology Interactions. Proceedings of the 6th International ICC/IFIP Conference on Electronic Publishing held in Karlovy Vary, Czech Republic, November 2002. Edited by João Á Carvalho, Arved Hübner, Anna A. Baptista. Verlag für Wissenschaft und Forschung, Berlin 2002: 395 p. ISBN 3-89700-357-0.

7 Bjoerk, Bo-Christer; Turk, Ziga and Martens, Bob. "Re-Engineering the Scientific Publishing Process for the "Internetworked" Global Academic Community", in: Elpub 2002 Conference Proceedings, Karlovy Vary (Czech Republic), pp. 152-159

<http://elpub.scix.net> [ELPUB Digital Library]
<http://iaps.scix.net> [IAPS Digital Library]
<http://cumincad.scix.net> [Cumulative Index on CAD]
<http://www5.hk-r.se/elpub99.nsf> [ELPUB 1999]
<http://www.albertina.ru/elpub2000> [ELPUB 2000]
<http://library.ukc.ac.uk/iccc/2001> [ELPUB 2001]
<http://www.tu-chemnitz.de/elpub02/> [ELPUB 2002]