

Theme:

Title: **The new role of professional organisations in the management of scientific knowledge**

Author(s): Žiga Turk
Bo-Christer Björk
Bob Martnes
Etiel Petrinja
University of Ljubljana, Slovenia
Swedish School of Economics and Business Administration, Finland
Technical University of Vienna, Austria
University of Ljubljana, Slovenia

E-mail(s): Ziga.Turk@itc.fgg.uni-lj.si
Bo-Christer.Bjork@shh.fi
b.martens@tuwien.ac.at
etial.petrinja@hotmail.com

Abstract: *The Internet is changing not only the way the architects and engineers work, but also how the scientists collaborate. In the paper based world, professional organisations, such as the CIB and scientific publishers provided for getting the right people together and for making sure their work gets distributed to their peers. Electronic networks, such as the Internet, are providing the scientists the means to take those activities into their own hands. In the paper we present the goals of an EU project called SciX. The goal of SciX is to analyse the business processes of the scientific publishing, invent new publication models and through a series of pilots demonstrate how this should work. In the envisioned scenarios, professional organisations such as the CIB play an important role. Their members are the potential users of SciX's platforms, authors and readers of the papers. Organisations could also become the publishers and archivers of the knowledge created within their respective community. The goal of the paper is to engage the members of the CIB-W78 community in the developments in SciX, to help fine-shaping the goals, defining the requirements and monitoring the usability of the pilots.*

Keywords: *scientific knowledge management*

Introduction

The history of the scientific publishing starts in the 17th century when the Royal Society of London created the Philosophical Transactions of the Royal Society of London (Gudeon 2001). The intention was to create a public registry of ideas - a logbook or journal - of the "present undertakings, studies and labours of the ingenious" who thought of what first - to protect intellectual property and ensure the rapid evolution of scientific knowledge (Fig. 1). For a long time, scientific publishing remained largely in the hands of learned societies and similar, scientist-driven institutions. Publishers 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 Universities around the developed world grew quickly. "What librarians (of these Universities) viewed as crucial core journals, publishers translated as the constitutive elements of an "inelastic market", i.e., a market where demand was little affected by pricing (and vice versa)" (ibid.).

The business model of the publishers is a rather fascinating one. The scientists do the research, they write the papers, they review their peers' work and they edit the scientific journals. They give away the copyright to their work, for free, to a party that has not been taking part in the value chain before. They then subscribe to usually rather expensive journals, so that they can learn about the work of their peers. In

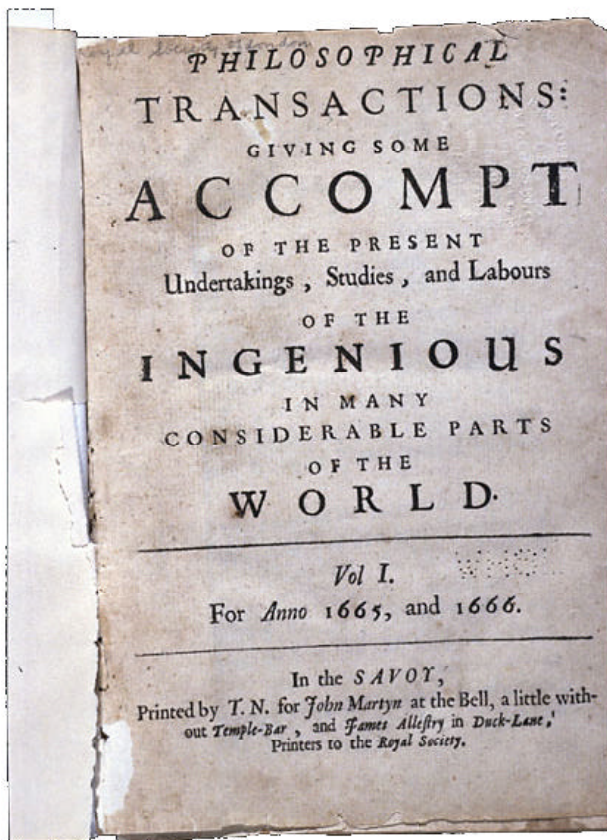


Figure 1: Cover page of the Philosophical Transactions

and the publication has been below 6 months. Each published paper had an average of about 1000 readers that viewed the abstract and about 1400 that downloaded the full text.

Since 1999, Bob Martnes and Ziga Turk have been managing CUMINCAD - the Cumulative index of CAD - the largest freely available database of papers related to computer aided architectural design, particularly related to the education in this area. At conferences organised by regional organisations of CAAD teachers (ECAADE in Europe, ACADIA in North America, Sigradi in South America and CAADRIA in Australasia) thousands of papers have been published. Rarely were the proceedings published by a professional publisher, therefore, the texts were not entered into commercial indexes and neither were they sold commercially. The full texts were not broadly available; only conference attendees had copies. On the other hand, the professional organisations retained the copyright to this work and could therefore allow its publication/archiving in the CUMINCAD. In this way this work is available on the net and rescued from oblivion. At the time of writing, CUMINCAD includes 3831 papers with abstracts. 883 of papers are available in full text as well.

A similar effort (with a couple of hundred papers) was created with the EGSEAAI community.

The CIB-W78 in 1996 in Slovenia was the first to use the Internet as the only medium to support the workflow of conference organisation. The W78 in 2000 in Iceland made it almost 100% web based.

Goals of this paper

The goal of this paper is to engage the CIB-W78 community in the SciX project. Since 1988, about one thousand papers have been published in the CIB-W78 proceedings. Most of these proceedings are the so-called grey literature - published by the workshop organisers - not generally available to a broader audience. And yet in this community valuable contributions have been made, particularly in relation to computer integrated construction and product modelling.

According to a study (Umich, 2001), about 50% of the costs related to making some literature electronically available is related to scanning and further 20% to the digitalisation of the material. By

the SciX project we believe that giving away the right to copy (copyright) and distribute results of scientific work to commercial publishers hinders the efficient exchange of this information and make scientific results harder and more expensive to get.

The above is the baseline of the SciX project. SciX (Open, self managed platform for scientific information exchange - IST - 2001 - 33127) 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), Technical University of Vienna (Austria), FGG Institute (Slovenia) and the University of Salford (UK). Project's home page is at <http://www.SciX.net/>.

Previous work

The partners have been active in the field of electronic publishing since the mid 1990s. Bo-Christer Bjoerk and Ziga Turk have been the editor and one of the co-editors of the Electronic Journal of Information Technology in Construction (ITcon). The average time from submission of a paper

working closely with the scientific community and with the scientists who authored the material, these costs can be saved.

In the SciX project we envisioned a creation of a target user group with the representatives of the main professional organisations where the scientific publications are being created. The role of this group is to comment on the work so that the results are relevant to the community.

On the other hand, within SciX, services and tools will be created and placed on the open source license, which could be useful for a community like CIB-W78, providing it accepts some guidelines on the publication of the proceedings.

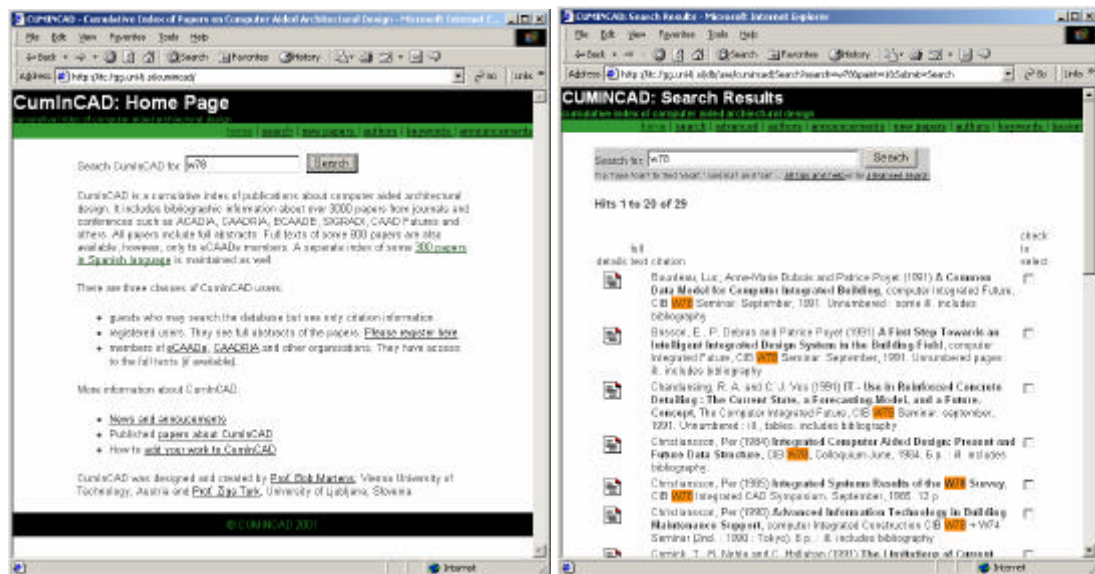


Figure 2. User interface of the CUMINCAD database.

Related work

Commercial

Both professional organisations, groups of publishers as well as specialised companies are providing added value services related to scientific publishing. An example is the CIB's database ICONDA, as well as several others. Several bibliographical databases are providing sophisticated search engines on bibliographic information about publications (such as titles and abstracts). Full texts are, as a rule, not available.

Table 1: Commercial indexes and bibliographic databases.

| | Ei Compendex | ICONDA | RSWB | CumInCAD | CiteSeer |
|-------------------|--------------|---------|---------|----------|-----------|
| Number of records | 6.000.000 | 500.000 | 575.000 | 3.000 | 2.500.000 |
| Availability | \$ | \$ | \$ | Free | Free |

The Internet represents a threat to traditional publishers. While some years ago, the Internet was a first resource for getting scientific information (Bjoerk and Turk, 2000) it is today becoming the only resource, particularly with the young researchers. Traditional publishers are responding with services such as ScienceDirect that allows pay-on-demand access to the full texts of published papers.

Another strategy of the publishers is to avoid dealing directly with the readers of the journals and attempting to close direct, long term deals with either whole universities (Landesman, Van Reenen 2000) or whole countries (<http://www.lib.helsinki.fi/finelib/>). Although discounts are offered if an institution subscribes to a full spectrum of journals the economies of such deals for the funding bodies and the researchers are not necessarily positive.

Related work in free publishing

The idea to use the Internet for scientific publication 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 such archive is the Los Alamos or arXiv preprints archive (<http://www.arxiv.org/>).
- Electronic journals (eJournals) and magazines (eZines). Similar to ITcon they provide similar quality control mechanisms as paper based publications. 400 such journals have been believed to have existed in 1999, including a Journal on Electronic Publishing. Today this number is estimated at over 1000.
- On-line bibliographies are collections of papers (usually without full text) from a certain discipline. An example from a related domain to W78's is the ARCOM collection of abstracts from construction management and economics (<http://www2.auckland.ac.nz/lbr/prop/propres.htm#ARCOM>). After having been published as a booklet for a number of years the abstracts are currently freely available through a database on the web. Another well known example is the CiteSeer service offering full texts of some 2.5million 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, index and cross-referenced.

The problems of all kinds of services include:

- Sustainability. Although the funds required to run such services are rather small, after the initial work done by the enthusiasts, a stable funding is required. The mortality rate of the electronic journals was 25% over two years (Wells 1999).
- Copyright. Many services include material that has been previously published in a way that required the transfer of the copyright.
- Prestige. An important factor in deciding where to publish is the prestige of a journal (Bjoerk and Turk, 2000), as perceived by the universities' or national research review processes. It is not uncommon, that a publication in a fully reviewed electronic journal is less valuable than publication at a conference where the author actually paid a fee to get their work published in an impressively hard-bound proceedings.

Exemplars in the field of software

The policy of the ARPA and the NSF in the United States was that all research that was funded through public funding should make the results available for free. This has not been entirely true for published papers, but has worked excellently with software. Programs written in the context of research projects were made available - for free, usually including source code - on the Internet. In fact, the software to run the Internet in the first place was available for free. This created the critical mass for the so-called open-source initiative (<http://www.opensource.org/>). An increasing number of operating systems, application programs and tools are available for free. Market share of those systems is growing and they are being used as a platform for vertical applications by companies such as the IBM.

On the other hand, the European funded research projects (such as the 4th and 5th Framework projects) never made a requirement for making the results publicly available. The excuse used was that commercial companies are co-funding this work and that they are not interested in making available what could be their competitive advantage. We are not aware of the scientific community challenging this system. Labelling most of the reports "restricted" actually restricted the readership to the project officers and the reviewers.

Exemplars in the field of standards

Standards organisations, similarly to journal publishers, do not fund the writing of new standards, yet they are given the copyright of a standard. They support their organisational activities by the selling of the paper copies of the standards. Several research efforts that deal with the computerisation of the building codes stopped at a prototype level, because of the problems with the copyright to the text of the standard.

Contrary examples are the standards that govern the Internet and the Web. The well known "request for comments" documents (RFCs) are the results of the work of groups of individuals and made available, for

free, on the Internet, to comment as well as to write compatible software. In the early 1990s there has been a competition between the ISO and Internet based networking, best demonstrated through the use of two different email addressing schemes. The Internet solution, based on open freely accessible RFC822 standard, prevailed.

The development of product modelling standards too started with the restricted publication model. Only recently the IAI is correcting the mistake and is making the entire IFC standard available on the Internet for free.

The open archives initiative

In the paper based publishing, a few dozens of publishers control most of the scientific publications and making a rather complete index involves including the publications of the few biggest ones. If, however, thousands will be creating digital archives on the Internet, indexing that information could be quite challenging. Web search engines, such as Google or Altavista are a most appropriate tool to search for scientific information, because they quite indiscretion index everything that they crawl into.

The open archives initiative (<http://www.openarchives.org/>) is standardising the metadata structure and the API of an archive, so that the archive can (1) be indexed so that (2) several archives can be search by the users at once. Moreover, Open Archives Initiative is developing standards that aim to facilitate the efficient dissemination of content.

Goals of SciX

The objective of this project is to demonstrate

- (1) that the scientific knowledge management can be made much more efficient,
- (2) show alternative business models to the ones practised in the past, and
- (3) analyse social, economic, legal and other related issues.

2-4% of the European GDP that is spent on research and development - on creating new knowledge. While several projects deal with the management of knowledge that is created within the industry, little has changed in the past hundred years in the ways that knowledge, created by scientific research and published in scientific journals, is handled. The current mainstream scientific publication process has so far been only marginally affected by the possibilities offered by the Internet, despite some pioneering endeavours. This is not so much because of lack of enthusiasm, but because there is a lack of sound business models and pilots to demonstrate the benefits of totally free scientific publication archives to the organisations, which ultimately should fund the development and maintenance of these.

The objectives of this project are:

- to enable scientists time- and cost-efficient access to their peers' work by creating a repository of electronic publications;
- to make the scientific materials in the repository also available to non-scientists - engineers, architects from the industry and explore new business scenarios.
- to support building a virtual on-line community of authors and readers.

To do so in SciX we will:

- create the necessary services infrastructure and populate it with at least 5000 papers from the domain of architecture and engineering;
- strengthen the already started transition to new modes of scientific publishing process so that the cheap dissemination channels of the Internet are put to efficient use; we will do so by setting up infrastructure to set up an electronic journal and make it available under open source licensing;
- perform a social-economic analysis of new business;
- investigate the legal, social and psychological obstacles to using eWork approaches in this area as well; this will include a survey among about 300 of our colleagues on their views on e-publishing.
- develop a method to benchmark scientific journals based on user requirements in the Internet era.

Most technologies and software to implement that is either freely available or has even been developed by the partners in this project in the past. See section "Previous work".

Enable efficient access to scientific results

The current methods for accessing scientific results are highly inefficient in view of the technical potential offered by the Internet and the fact that scientific research results, from the viewpoint of the public sector financing research, are aimed for reuse in other research and application in industry, not as a commodity to be sold per se for a profit. It would seem to be in the interest of the public R&D funding bodies (e.g. European Commission) and of the academic community as a whole to have a completely free cyberspace of scientific information, in order to speed up the scientific research process and save costs. The objectives of the project are to explore business models and techniques which: speed up the process from submission to final publication, allow a more rich content (multi-media), provide readers more efficient mechanisms for retrieving publications of interest and increase readership through the abolition of barriers such as subscriptions.

Provide access to non-scientists

Making the information available for free and with efficient search mechanisms could also dramatically increase the number of readers from industry and from smaller universities (the SMEs of the scientific community), who seldom can justify paying expensive journal subscriptions or have the time for extensive searches for information. For industry readers, a crucial factor is also the time spent in searching and retrieving relevant information, they want relevant information just-in-time, and can't afford the luxury of reading hundreds of papers just-in-case there is something of use for them.

Re-engineer the publishing process and perform a social-economic analysis

In this project a process reengineering view of the whole life-cycle process of scientific papers will be taken, with the intent to find savings of 80-90% in the distribution – retrieval costs. Compared to the 10-20 % approaches often taken in development projects initiated by commercial publishers and libraries, these savings are very promising. The key issue is the paradigm shift to see scientific publications not as a commodity to be sold or archived but as an essential part in a larger scientific communication process, and to look for solutions based on the premise of globally free information on the World Wide Web, thus side-stepping some of the traditional intermediaries altogether.

Automate repository management through self organisation

The amount of digitally stored technical data, both general and corporate, is growing rapidly - more rapidly than the ability of the humans to appropriately structure, classify or index it, so that it could be found and (re) used. Typically, this information is available through different search techniques. Searching, however, implies that the user knows what she is looking for. Another approach to access the data is through browsing, which requires a certain structure imposed over the data items. The main function of the structure is to provide user navigation through the data. The structure should tell the user what items are similar, which are different, and how. The simplest structures of this kind are clusters or groups of similar data items. By using data mining it is possible to create an algorithm that would create clusters of data automatically so that the clusters would be similar to the human interpretation of that data. For example, given one or a few papers related to certain topic, the machine should come up with a cluster of similar papers, which should be of interest to the reader as well. Such clustering becomes very interesting when applied to large repositories of publications, such as the one planned in this project.

The freely available clustering and automatic classification software will be put to use in the project.

Simplify use through intelligent personalised agents

Another important part of the project is a user profiling system that would add value in combination with the automation described above. Automatic notification on new papers matching the profiles' interest and selective searches will be provided without the need to build a very sophisticated profile. The user will be able to semi-automatically modify the query with assistance of the system and update her user profile.

Investigate legal, social and psychological issues

The main problem to a new vision of information exchange in science is the copyright that researchers currently give away to the commercial publishers for free, and which results in severe obstacles for potential readers to retrieve the information they need. There are also other barriers for a shift to free repositories dealing with perceived risks of Internet publishing, sluggishness of academic department to change their "rating" systems, etc which need to be studied. A survey we conducted in the year 2000 in the field of construction IT and management showed interesting results in relation to what scientists think

about where to publish and what to read. We intend to continue this survey over the next years so that the trends could be monitored as well as the impact of the proposed repository.

Develop benchmarking methods for scientific journals

Typically scientific journals have been rated by prestige, often based on subjective evaluations or to some extent on the use of citation indexes. Ratings have been done implicitly through university departments, for instance in shortlists of accepted publications for promotion etc. Little attention has been paid to questions of how quickly and efficiently the information passes to experts for whom the information could be useful. Thus it would be very useful to develop methods, which would allow the benchmarking of journals for also other factors than the scientific quality of the papers (turnover time from submission to publication, availability, readership etc.). Such a benchmarking tool will be developed in the project and tested with a number of journals of different categories. The main value of such a tool would be as a tool increasing the awareness within scientific communities of the deficiencies of their current communication process, which hopefully will trigger activities to change the process.

The demonstrator

The main components of the demonstrator include:

- a digital library of works published elsewhere.
- an electronic journal.
- support for a workshop of a conference.

The audience is invited to address the requirements for such services to the authors of this paper. Open source solutions and/or rentable Web infrastructure will be created and made available to potential users.

Conclusions

In early 1990s the W78 was the only forum where issues related to computer integrated construction could be presented. The role of CIB was decisive, because it had membership lists, newsletters to keep the membership informed, publishing and organisational experience. At that time, organisations too, had to be from the real world and paper based. Since then, several new tracks of conferences have emerged, fighting for similar audiences, including rather regular events such as the ECPPM, the CE and CEC concurrent engineering conferences, as well as a number of unique events. Creating groups of people has never been easier. Email addresses can be collected by spidering the Web, the added value of a clout such as the CIB is diminishing. Organisations are getting virtual, they do have a cyberspace address but not an office space address.

One of the advantages of the longer rooted communities, such as the CIB W78, is its track record and prestige; in the hundreds of papers published by people, who may now be the authorities in the field. This track record, however, is remembered by a few dozens who have been attending the W78 regularly. It is documented in the proceedings which are mostly locked in the cabinets of those attending. All others could appreciate the achievements of the W78 if they were electronically and freely available. In the SciX project we will make an infrastructure available that will allow the W78 to put their history on-line, in a very similar fashion, as the ECAADE already did. It will also provide some content. For more we will ask for the support of the previous conference organisers as well as the W78 board in setting up some guidelines for the future organisers.

Acknowledgement and disclaimer

The presented work has been conducted in the context of the SciX project, funded by the European commission under the contract IST - 2001 - 33127. The home page of the SciX project is at <http://www.scix.net/>. The contribution of the funding agency as well as that of the industrial partners in the project is gratefully acknowledged.

The opinions expressed in this paper are that of the authors and do not necessarily represent the opinions of their employers, of the SciX consortium or of the European commission.

References

- Guedon, J.C. (2001). **In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing**, proceedings, Association of Research Libraries, Proceedings of the 138th Annual Meeting, Toronto, Ontario, May 2001 - <http://www.arl.org/arl/proceedings/138/guedon.html>
- Björk, Bo-Christer, Turk, Ziga (2000). **How Scientists Retrieve Publications: An Empirical Study of How the Internet Is Overtaking Paper Media**, Journal of Electronic Publishing, Michigan University Press, Vol. 6/2. 2000, <http://www.press.umich.edu/jep/06-02/bjork.html>
- Landesman, M., Van Reenen, J. (2000). **Consortia vs. Reform - Creating Congruence**, Journal of Electronic Publishing, Vol 6/2, <http://www.press.umich.edu/jep/06-02/landesman.html>
- Wells, A. (1999). **Exploring the development of the independent, electronic, scholarly journal**, M.Sc. in Information Management, Electronic Dissertations Library, Department of Information Studies, University of Sheffield, UK, <http://panizzi.shef.ac.uk/elecdis/edl0001/index.html>
- Univ. of Michigan Library Services (2001). Assessing the Cost of Conversion, http://www.umd.umich.edu/pubs/moa4_costs.pdf

All URLs mentioned in this paper worked in April 2002.