_____

ELECTRONIC BIBLIOGRAPHY - ITS RELIABILITY AND ITS IMPACT ON THE CONCEPT OF BIBLIOGRAPHY IN GENERAL

Jadranka Stojanovski
Institut Ruder Boskovic
Zagreb

Aida Slavic
Department of Information Sciences
University of Zagreb

Summary

The authors will present a project of Croatian Bibliography of Research Papers in the web environment. The bibliography was supposed to provide prompt and accurate information about the papers published within the research projects funded by the Croatian Ministry of Science and Technology. The bibliographic database has been designed for the web environment and the authors were given the capability to enter bibliographic records, as well as search and browse the database. Public access to the bibliography and the way the database has been constantly generated on-line opens many possibilities and questions. This paper will discuss both the positive and negative aspects of this approach based on a one year study. The problems of accuracy and reliability of the records provided will be discussed as well as the competencies of the authors from the point of view of the record validity.

Introduction

New information and communication technologies open up enormous potentials and have changed the way we use bibliographic tools. They have influenced our expectations of what we know as bibliography. Traditional printed bibliography nowadays cannot compete with the scope and possibilities of an electronic form. Printed bibliographies of scientific papers, for most scientific fields, are today more documents of historical value than sources of information needed in current scientific research.

Electronic bibliography in the network environment embraces bibliographic information itself, information with so called "added value" (place of the document within the domain knowledge or ranking of the document' contribution), and even a full-text document (Arnold, 1995). It could inform us of the impact that a document has on the usage of bibliography, as well as its impact within the field to which it contributes. This bibliography will very often contain the records of the documents which have not been published in paper form yet, or which will never be. This kind of bibliography can hold information about patents, computer programs or other kinds of authors' contribution to the science which would never enter in a traditional printed bibliography together with published items. Being in electronic form and equipped with searching, sorting, filtering and ranking tools, it is an easily accessible and available source for scientometric analysis of all kinds on a national or global scale. Therefore electronic bibliography is not only the information tool within some information service but rather an information service itself.(Tyckoson,

_____

1991)

Besides having numerous advantages, the full-text document concept imposes problems connected with rights and permissions, and electronic publishing in general. Articles being published in paper or electronic form of some scientific journal are usually the subject of some copyright agreement and very often their display in the bibliography is restricted. Furthermore what works on paper does not work on-line, the electronic document is not static. Both full-text documents and bibliographic records could be constantly and rapidly changing in form and content and this fact very often raises the question of reliability and accuracy. Finally, there is another obstacle to applying old bibliographic control rules and concepts in an electronic document. This document (or bibliography) when available on the World Wide Web very often is in hypertext format. In this case it consists of text itself and intricate web of links within the document or links to some other, remote documents which can be in the form of text document or in the form of image, sound or animation. (Fillmore, 1993)

This paper will be presenting the project of such an electronic bibliography with the stress being put on one of the features which challenges information professionals the most. Namely, the electronic bibliography which is going to be presented here is generated by the authors themselves, whilst librarians are those who had established standards, bibliographic forms and who supervise and control the process. The entire bibliography is available on the World Wide Web and could be edited by the authors or accessed by public (http://bib.irb.hr).

This paper will discuss specific problems connected with this approach and will try to summarise achievements and problems that occur. The experience presented here will give an insight into the information and computer competencies of the researchers and the changing role of the librarians.

Background

Significant changes have been occurring recently in world-wide scientific information systems. Some surveys and analyses have shown that the scientific community is under diverse pressures. These pressures are connected with shortages of funding which result in an increased rate of publication by researchers in order to keep up with competition for research grants. Researchers are more and more exposed to the growing burden of teaching and administration to be carried out along with research. In addition, it would point out some additional pressures as the result of reorganisation of many research and scientific institutions. Furthermore, growth in availability of information technology for the last two decades contributes to the overall pressure (Meadow; Buckle, 1992). Although this is a well known problem world-wide, it is even more apparent in the less developed and economically weaker countries such as Croatia, in which research funding is much more limited.

It should be noted that research papers published in Croatia are covered by the Croatian National Bibliography (issued by National and University Library). This bibliography is organised in three sections (A-monographs, B-serial Sciences and C-serial Arts, Humanities and Social Sciences) and is available in paper edition only. Due to technical and organisational problems its publication has been delayed by several years. The fact that information in the bibliography is very much out of date and the fact the bibliography does not cover all papers published by the Croatian authors (but only those published in this country), has always been a problem for the Croatian scientific community. For the researchers as well as for the bodies who finance the research, everything published by

_____

Croatian authors is of the greatest interest, no matter in which country is published. Most of the researchers strive to publish their papers in journals that have an important impact on their field of research, however, most of those journals are published abroad. This is specially the case when it comes to the field of science and medicine, in which very few papers are published in Croatian journals, covered by National Bibliography. Data about these papers has been scattered in the individual bibliographies of researchers, national bibliographies of various countries and commercial databases. Although the availability of information about their publications is very important for the career appraisal of every researcher, it appeared to be hard for the Ministry of Science and Technology (MST) who finance most of the research in the country, to gather the data.

These are the reasons the Croatian Ministry of Science and Technology has tried to collect bibliographic information in an electronic form in order to make an accurate and reliable database. In 1994 MST started a project called SVIBOR in order to collect data about current projects and papers published by the researchers. For this purpose special a software package was designed and distributed on diskette. The authors were supposed to input data about their papers and return the diskette to the MST, which would prepare the database for Internet display. The programme package had various imperfections (which occasionally caused the loss of data) and was found to be rather unfriendly and hard to use. Therefore in 1996 a simple MS Word template was created and distributed to the authors. Lots of technical problems appeared in the process. Most of them occurred because of the incompatibility of existing hardware and software equipment and different code tables used for Croatian fonts. However, the survey of the researcher showed that, although finding the MS Word template to be cumbersome, they preferred digital form of any kind to paper one. It was concluded that SVIBOR, at that phase, was not as successful at it could have been for following reasons:

- software package problems
- authors were not able to add or to correct data once the forms were filled in and sent to MST
- database was available on the Internet with significant delay (few months)
- library professionals weren't involved in database maintenance or any kind of bibliographic control.

However, these experiences enabled future projects of electronic bibliography to be better defined and known pitfalls were avoided by technological and communication advantages.

The scope of the Croatian Electronic bibliography of Research Papers

During the past two years better equipment and the development of the Croatian Academic Network (CARnet) offered the opportunity to advance and improve the idea of SVIBOR.

The vision of the electronic bibliography which could be edited and accessed on the World Wide Web came as a natural result of technological and communicational facilities and was encouraged by information professionals at the Ruder Boskovic Institute. They have been striving for years to improve scientific communication and ease access to research data in Croatia. Therefore the project has been realised by a team of information professionals in this very institute.

SVIBOR has already established the rule that bibliographic data should be the responsibility of the

_____

authors who possess all the necessary information and who can provide the information the very moment the paper is ready for publishing. The bibliographic references were the part of the paper work every researcher had to keep in order to get his career appraisal or project funding. So one of the tasks of the electronic bibliography project was to ease this paper work as much as possible.

The electronic bibliography of research paper is designed to provide:

- complete and fast information about research papers (published and those accepted to be published)
- information about ongoing research projects (to be browsed and retrieved by project name, institution or subject field)
- reliable access to electronic documents
- provision of full-text document retrieval and download facilities from the bibliography archive (respecting copyright agreement)
- easy to use and available 24 hours hours a day as a source of reference (which scientists can use while working on a project, writing their reports or preparing paper for publishing)
- a source of reference for different research and academic institution to ease the maintainence of their local databases
- the facility to merge all existing records in the archive of the MST bibliographic records into one unified database at some point in the future
- the facility to include in the databases of other already existing bibliographies of research papers scattered in different institutions
- provide an exchange of information on an international level

It could be expected that an electronic bibliography of this scope would improve the scientific communication and enable research projects to be acknowledged and properly evaluated. An electronic bibliography is designed to hold information about projects, researchers and bibliographic references. Projects are indexed and can be retrieved by the subject field to which they belong. Bibliographic records provide the possibility to rank each reference, according to whether or not the journal in which it is published is cited by Current Contents or the Science Citation Index. Besides being a help to MST, this bibliography is designed to be a bibliographic tool to researchers, as well as evidence of their scientific achievements. Moreover it is a potential source of numerous scientometric analyses of paramount interest to the scientific community as a whole.

To fulfil its function this bibliography covers information about:

- scientific production of the individual authors
- scientific production within the research project/programme/theme
- scientific production of particular institution financed by MST

Technical details, organisation and management of the bibliography

The bibliographic form available on the World Wide Web (http://bib.irb.hr/) collects data into several database tables (sorted by the type of paper: book, article, paper presented at the conference etc.). Records are edited in several levels and at the moment fields are only partially controlled (it is not possible for the main fields to be left empty, for instance), whilst thorough validation and field

control will be established in the current phase of the project.

Database is designed in Mini SQL 2.x first because it was free of charge for the academic purposes, then because of its interface with Perl (Practical Extraction and Reporting Language) used in cgi-scripts on Web servers. Perl is found to be easier to use for data processing and analysis and it enables direct access to records in a database, which is not always the case with other languages. The advantages of Perl found to be important for this project are:

- usage of "lite", a script-language which enables dynamic generation of web pages in the retrieval process
- the possibility to administer access to individual parts of Web-tree while manipulating with data
- data can be easily export in any other database using standard SQL language

However Mini SQL has some bad sides as well: a few bugs, SQL commands are somewhat reduced, there is a limitation of types of data that can be stored. Its performance regarding retrieval speed and quantity of data was not noticed to be a problem. Script-language is used to enable access, retrieval, processing and display of the records in the database.

On the level of bibliographic description, document exchange and retrieval, the usual standards are in the process of implementation. UNIMARC has been chosen for the bibliographic standard, PDF for the document exchange and Z39.50 is planned on the level of network access.

The editing procedure should be rather simple. The very moment the paper is prepared for publishing and sent to the publisher the researcher will then enter his reference in the bibliography on the web with a comment about the paper status (application only, unpublished or published reference) and if he wants he can upload the full-text document into the archive. Once the paper is published, the status of the reference in the bibliographic database has to be changed from "unpublished " to "published".

Records in the bibliography being linked to the project and the official identification number of researcher thus become the part of his official annual report and author need not to report about his papers again.

The form they fill in on the web consists of:

- author form (name, researcher identification number, number of the project, name of the institution), which will be supported with the authority files, after which authors will only select their name and corresponding identification number
- bibliographic form (different for each type of publication, offered in menu) which consists of usual bibliographic fields which can be easily filled in using copy/past function from the source document
- field for the full-text document (file could be easily uploaded from the local machine only by specifying directory and document format, after file is uploaded hyperlink is generated automatically, and URL address of the file in the archive is reported)

Full-text documents, if the author requires, are stored in the database permanently and therefore available to the author only whenever or wherever he needs it. If author wishes his full-text paper to be available to the public he has to give his signed copyright agreement to the Ministry of Science

_____

and Technology. Bibliography is supervised and bibliographic records are controlled and corrected by the librarians working on the project. The WWW interface is designed to enable the document to be edited in a different application, avoiding re-typing errors. And this interface provides converting facilities to correctly display Croatian fonts coming from different code tables, a problem not solved in SVIBOR.

The authors are free to correct their records by accessing the database on the Web using a special correcting worksheet, however corrections are not updated until being controlled by professional librarians. Access to the editing module was not restricted to one person only, and is not protected by password in this initial phase of the project. The idea was to allow all the research collaborators to enter the data about the project as well, and thus avoid problem with IT literacy of senior staff. The main concern was to gain confidence and a good response from the researchers.

Database is available for browsing and searching. Searching could be simple (one field, or combination of two fields), advanced (Boolean logic) or filtered (adapted for special type of evaluation). Precise instructions and help was provided for each field and most of questions could be addressed and promptly replied through e-mail. (Stojanovski, 1997)

First experiences

From December 1997 till the September 1998 almost ten thousand bibliographic records were entered in the bibliography:

article in journal	3913
paper in proceedings	3483
book	280
paper in book	674
thesis (master or	881
other	724
TOTAL	9955

During this first period statistics for accessing, editing and correcting bibliographic records were carefully observed. At this point it was possible to have some reliable information about database performance as well as have some overall view of the whole concept of electronic bibliography. For the project team, however, feed-back from the researchers was of paramount importance.

The observation about electronic bibliography performance can be grouped into technical problems and human errors, as usual. They greatly influence the next phase of the bibliography project as it is going to be explained further.

Technical problems

The most problems researchers had during editing their references were connected with the compatibility of their Web browsers (Netscape, MS Internet Explorer) with any of used code table for Croatian fonts. Before editing, researchers were offered a display of Croatian fonts according to

_____

the four code tables most used in Croatia. They selected the right table clicking on the correct display of Croatian fonts as they saw it on their monitors. However, in spite of this facility and precise instructions given, Croatian fonts were not always properly edited. When noticing that, some researchers simply avoided special fonts, using instead of them simple letters (c, s, z). This caused the irrecoverable loss of some data. This problem will disappear the very moment one code table will be accepted as a standard for the whole country.

And, as an observation, it should be noticed that Netscape browser proved to offer a better environment for editing corrections. Namely, the corrections of bibliographic records which were already in the database were done in a special worksheet sent by e-mail, in order to be checked by librarians before entering into the database. When Internet Explorer is used MS Outlook should be defined as default e-mail programme.

Human errors

These observations are of the great importance for the whole concept of the bibliography which is generated by non-professionals, and should be taken into consideration when discussing the advantages and disadvantages of this approach.

The most errors that have been noticed and corrected in the records belong to one of the following categories:

- instruction to give the title in the "language of the source document" was misunderstood, so many titles were unnecessary translated into Croatian

- one paper is entered more than one time when the author did not take the time to check his references

- on the question is the paper cited by Current Contents or Science Citation Index, author would reply "NO", simply to avoid checking (although Current Content is purchased and available to the most of Croatian researchers on-line)

- part of record is missing (number of pages, summary etc.)

- papers published in the proceedings are lacking details (was the author invited, was it presentation with poster or was the conference international, was the paper reviewed by international committee or not etc.)

- usage of c, z, s instead of Croatian fonts

It should be noted that authors were given precise instructions with all of the above mentioned cases. It could be stated that these errors are the result of the layman's attitude and a lack of understanding of the importance of the validity and accuracy of data. However it should be underlined that these mistakes were not made "on purpose" since in most cases the researchers themselves were to suffer the consequences. It will be their loss if they neglect to point that their paper was cited in Current Content for instance, or if their paper or name is lost in the retrieval process because they did not use Croatian fonts.

_____

These reasons have been always the strongest arguments for the tedious work of cataloguing done by professionals and not by people who had "more important problems on their mind" when doing the research. Still, in spite of this well know attitude of librarians and researchers, this project with its overall success proves that reality could be different and announced the change to be accepted both by librarians and by researchers. Researchers will continue to have their "paper work", only this time it will be better managed and taken care of by librarians. Researchers will emphasise speed of information flow, and librarians will continue to worry about organisation, accuracy and validity of bibliographic records.

The behaviour of the researchers is easy to predict and most of problems that occurred because of this advocated approach had been expected. Information technology today provides the facility to supervise, control and correct processes, and information professionals have the competencies to extend their role to be advisors, educators and knowledge facilitators. As it is shown here they have competencies to design information systems to embrace a wide scale of functions, bibliography being only one of them.

Future development

In the next phase of the project alot of work is going to be concentrated on the control and correction of records and the following steps are planned to be finish in the near future:

- correction and concordance between bibliographic database and the database of projects (MST)
- improvement of search speed (it is 10-20 seconds at the moment)
- better retrieval facilities with sorting, printing and saving functions
- control and the correction of the records (weeding duplicates, correcting mistakes etc.)
- input of work done one the old, finished project
- input of old data collected by the MST in the last years (1991-1996)
- name authority files (authors, projects, institutions, journal names)
- electronic full text archive with standard network formats and open public access
- implementing bibliographic, document exchange and network retrieval standards
- improving database maintenance by library professionals

The pace of these improvements will, of course, depend on funding available but most of them have been already started.

Conclusion

An electronic bibliography project of research papers was presented here to show the scope and reach of current changes initiated by the development of information and communication technologies. Information overload in the scientific and academic community is a growing problem world-wide, and the concept presented hereof bibliography should be viewed in the context of global changes and trends. The aim of this paper was to show, on the example of one project, that a comparison between paper and electronic document is. Furthermore, the use and purpose of an electronic bibliography of research papers, for instance, differs a great deal from a printed

_____

bibliography of the same field.

The electronic bibliography as it is showed here has not much in common with traditional bibliographic tools and its advent changes both users' and information professionals' role. Its priorities are promptness, speed, and access. Being a rapidly changing organism it has a more tolerant approach towards the accuracy and validity of data and different mechanisms to handle these. Paper documents cannot tolerate mistakes, once printed they will always stay there. Electronic documents can be corrected but can not tolerate being presented late or not being accessible. An electronic bibliography has advantages and imposes problems common to the rest of its electronic family.

Managing bibliographic records of scientific papers in the electronic and network environment has many sub-themes. The main question of this paper, however, concerns the way in which information technology helps scientists handle information and the way scientists are included in the process of cataloguing bibliographic references. As an example, with the Croatian Bibliography of Research papers we tried to advocate the approach that those who generate information should be the first to record it. Information technology today provides facilities to help researchers handle their project paper work and the Croatian Ministry of Science Technology was the first to realise that bibliographic references once made by the authors of research papers could be better organised and professionally managed.

The question of reliability and the accuracy of data edited by authors was specially underlined here. It could be said that this problem is best to illustrate the scope of current changes in the field of information profession. The change in the role of librarians in order to become knowledge facilitators, organisers, managers, and supervisors of the information flow process is obviously inevitable in the new technological environment.

Lots of questions such as those connected with electronic publishing issues, copyright agreements and electronic journals as well as those connected with indexing and retrieval are yet to be discussed and analysed within this project. Future surveys, if undertaken, could contribute to the understanding of the digital library environment and new possibilities. The question is not whether or not we should take the advantage of them, but the reason we did not do so yesterday.

_____

References

1 Arnold, K. The Body in the virtual library : rethinking scholarly communication. // The Journal of Electronic Publishing, January, 1995 Volume 1, Issue 1
 http://www.press.umich.edu/jep/works/arnold.body.html

2 Fillmore, L. Internet Publishing in a Borderless Environment: Bookworms into Butterflies :Notes for the Presentation at the Electronic Media Conference at the Frankfurt Book Fair, Frankfurt, Germany October 7, 1994. http://www.press.umich.edu/jep/works/fillmore.butterflies.html

3 Fillmore, L. Internet Publishing: How We Must Think : Presented at Meckler's Internet World '93 Conference Internet Publishing Seminar New York, New York December 7, 1993

4 http://www.press.umich.edu/jep/works/fillmore.think.html

5 McCarty, W. The shape of things to come is continuous change : fundamental problems in electronic publishing. http://www.kcl.ac.uk/humanities/cch/ohc/shape.html

6 McKiernan. Parallel universe : the organisation of information elements and access in a World Wide Web (WWW). // Advances in Knowledge Organization, vol. 5 (1996), 254-260.

7 Meadows, A. J; P. Buckle. Changing communication activities in the British scientific community. // Journal of Documentation , vol. 48, no. 3 (1992), 276-290.

8 Stojanovski, J. SVIBOR : o projektima Ministarstva znanosti i tehnologije (About projects of Ministry of Science and Technology). // MOST, 22. prosinac 1997. http://www.mzt.hr/mzt/hrv/most/most-22/most-222.htm#SVIBOR

9 Stojanovski, J. Hrvatska znanstvena bibliografija na Web-u. (Croatian Bibliography of Scientific Papers on Web). // MOST, 24, lipanj 1998. http://www.mzt.hr/mzt/hrv/most/most-24/most-242.htm#HRVATSKA ZNANSTVENA BIBLIOGRAFIJA NA WEB-U

10 Tyckson, D. A. The twenty-first century limited : designing catalogs for the next century. // Cataloging and Classification Quarterly, 13, 3/4(1991), 3-27.