DESIGNING AND IMPLEMENTING A WEB-BASED FULL-TEXT DATABASE FOR PERSIAN ELECTRONIC JOURNALS[1]

J. MEHRAD, Ph.D.
President of RLST
Shiraz, I. R. of Iran
email: dean@srlst.com

S. KOLEINI, M.S.
Head, Dept. of Information Technology, RLST
Shiraz, I. R. of Iran
email: koleini@srlst.com

Abstract – By winter 2003 the Regional Library of Science and Technology (RLST) successfully completed the design and implementation of a Full-Text Database on Persian Electronic Journals. The main objective was to facilitate easy access to articles embodied in Persian journals without requiring the physical presence of the users. Small and medium-size libraries in Iran as well as those of other countries encounter ample problems and difficulties in their endeavor to add all the issues of a given journal to their holdings. This point clearly asserts that the existence of such a database can play a really important role in society. Due to the existence of the whole set of issues and volumes of different journals, this database will act as an invaluable source of Persian journals in the near future. In this article, foremost the term *electronic journal* will be defined. Steps will then be taken to design and implement the database.

Keywords - Persian Full-Text Electronic Journals, Electronic Journals, Regional Library of Science and Technology, RLST.

INTRODUCTION

Almost 200 years after the invention of printing press, the first generation of scientific journals appeared rather simultaneously in London (*Philosophical Transactions*) and in Paris (*Le Journal des savants*). This happened in 1665 A. D. In recent years the number of scientific journals have been increasing very rapidly. It is estimated that in the years 1800, 1850 and 1900, the total number of journals published were 100, 1000 and 10,000 respectively. Maintaining the same rate of increase would give way to the publication of 1,000,000 scientific journals by the year 2000. [3] Fortunately, this did not happen and the rate of increase diminished for one reason or another. It is estimated that there were 133,000 journal titles by the year 1991. [9] It is not clear how many of them are still published.

Today, journals on paper are the most important medium through which scientific progress and findings are reported. A printed journal is the center of a cycle in which the

1. The authors wish to acknowledge the RLST Scientifie Counеil for the research grant allocated to this project.

following items play their parts:

1. Scientists and researchers as producers and users of information.
2. Business and non-business publishers that provide editing, evaluating, abstracting and indexing services.
3. Libraries that hold the responsibility of selecting the needed journal titles to meet the needs of their users, to keep the records and to provide their affiliated institutions as well as others with their information needs based on the requests they receive and using the inter-library loan service for non-affiliated institutions and users.

The most important functions of journals are as follows: communications and dissemination of information; quality control of journals through the evaluation process and finally archival function. [1] Of course, there is a fourth function, which is directly related to the control of quality and that is determining priorities. Nowadays, these basic functions are carried out by printed journals, and are vital to the natural flow of scientific communications. These functions must apply to any new medium of communication that wants to bridge communication and record scientific endeavors. There are indications that printed information cycle may vanish soon. Library budgets can no longer support the crisis of an ever increasing number and cost of journals. What Karen Hunter has called the publication's "echosystem" is now in a gradual process of erosion. [8]

The increasing cost of journals as well as the budgets that libraries can allocate to the purchase of periodicals are two important factors that must be taken into account. Libraries are well informed that at least during the past two decades, inflation has been a serious problem. During the past few years a 10% to 15% increase has been observed in the average price of journals.

In each country or market, the rate of inflation varies from year to year, based on the fluctuations in the value of the local currency. For instance in Iran, during the first years of 1990s, the value of each U.S. dollar was lower than 100 rials for purchase of journals. During the last years of the same decade, this balance changed to 1750 rials for each US dollar. Based on the policies adopted for stabalising the value of the Iranian currency, the exchange rate is presently set at 8400 rials per US dollar.

Moreover, it is apparent that the printed journal will lose its main function as the dissemination and provider of information. Even before *the electronic age*, scientists had come up with faster, and of course more informal, ways of exchanging and sharing their ideas. But, after the emergence of the Internet – with numerous conferences being held through its list servers, news groups and bulletin boards – the world has actually transformed into a global village. In this new environment, one could hope for the emergence of an electronic version of the printed journals, which eventually occured late in the 1980s, 300 years after the printed journals.

ELECTRONIC JOURNALS: HISTORY, DEFINITION, FEATURES

The first endeavors made in the domain of electronic journals were projects that were trying to produce the electronic counterparts of the already existing printed journals. ADONIS, as one of the oldest projects, stored the articles published in printed journals on CD-ROMs. [11] After its other projects were carried out: TULIP related to Elsevier in the United States of America and Rightpages TM Service carried out in AT & T Bell laboratories. [6] Another example that is very similar to Elsevier is a project that was carried out at the library of Tilburg University. This project provided the users with an access to 100 electronic journals. [4] Nevertheless, more ancient examples are related to full-text information databases that are monitored by host organizations. All these projects embodied journals in electronic format; nonetheless, one cannot call them electronic since they were just electronic equivalents of printed journals.

The term Electronic Journal has been defined in a number of ways. Nevertheless, it seems that the definition provided below better illustrate the term. In general, electronic journals or periodicals could be defined as any type of journal, magazine, newsletter, E-zine, Web-zine[1], or any other type of electronic material that is accessible through the Internet. Based on this definition, it is possible to access the electronic titles using different technologies like, the Web, Gofer, File Transmission Protocol, TelNet, Email and Listservers. [2]

The Oxford Advanced Learners Dictionary [9] has defined the term *e-zine,* as a completely new form of electronic periodical, but has left the term *e-journal* untouched. In this dictionary, *zine* refers to any kind of periodical that includes some piece of news or current issues. Similarly, the prefix e – refers to the electronic exchange of information illustrated in the form of web or pages. Therefore, in addition to the previous items, concepts such as personal web pages, Blogs and Digests could be added to the definition of this term.

In fact, the concept of electronic journals was actually introduced in 1980. [10] Such journals, unlike the electronic journals already mentioned, are factual examples of electronic journals. The previous examples, including RLST's project of design and implementation of web-based full-text Persian electronic journals, illustrated only photos of the already existing printed journals. The journal entitled *Postmodern Culture,* which was put to print in 1990 and since then has been published by the Johns Hopkins University Press is a good example of an electronic journal. The publication of this journal coincided with the emergence of the Internet.

Based on a study [5], there were 115 electronic journals by the year 1995. This number rose to 1300 cases in the following three years. [6]

The personal experience of the authors of this article with respect to electronic journals go back to 1999 when RLST, together with Tehran and Tarbiat Modarres Universities,

accessed 1000 electronic journal titles through *the Document Delivery Plan.* Until the year 2004 a large number of Iranian universities and research institutes signed contracts with RoseNet and Ebsco, which enabled them to have access to journals related to different subject areas. Among these electronic journals there were cases that had no printed versions at all. Such an access embodied varying issues like training, organization, etc. The total number of e-journals was 10,000 in the year 1999, and rose to about 30,000 by the ycar 2004.

With respect to electronic journals three important endeavors have been made in Iran which will be explained below:

1. Namaye CD, including the full-text of articles, criticisms, texts of scientific and cultural newspapers, etc. This program enables the users to search their information based on items like, subject area, title of the article, title of the newspaper, ... and the name of the author. This CD is published by the Board of Trustees of Iran's Public Libraries affiliated to the Ministry of Culture and Islamic Guidance in 1998. The result of digital recording of periodicals has been 300,000 pages of information per year.
2. Aftab Software Company established Iran's Journal databank in May 2001. Its main objective was to create an efficient reference source for the country's journals on the Internet, as it intended to play an important role in the dissemination of information in Iran. Despite this objective, in the final output of the project there are only bibliographic information related to 750 journals, which are accessible at: www.magiran.com. To enrich the databank, this company requests publishers to deliver the full-texts or abstracts of the articles of their journals to the company.
3. The project entitled, *The design and implementation of web-based full-text Persian electronic journals*. This project is being carried out by RLST.

PERSIAN E-JOURNALS AT RLST

RLST has entered Persian journals into its homepage. They have been added from cover to cover with their whole contents and have been arranged alphabetically according to their volumes and issues. This enable the users to retrieve and print the article(s) they need from the system.

All in all, the Persian e-journal database was designed to achieve the following goals:

1. To access the full-text of articles of Persian journals through the Internet.
2. To resolve the problem of access to the previous issues. Often libraries do not have all the back issues.
3. To eliminate the problem of purchase and binding of journals. This might be a difficult task for domestic and foreign small and medium size libraries.

The Development of the Persian E-journal Database

Acquiring Entire Issues of a Journal

The first step was to collect all the issues of a given journal each year. The possible problems related to this stage was that not all the issues of certain journals were available.

Producing HTML[2] Pages for Each Issue of Each Volume

HTML pages were produced for each issue based on the table of contents available in each journal. If the Persian OCR[3] Program was at hand, a large number of problems would be resolved. To save time, use was made of RLST's main database to extract the titles and index related to Persian journals. A large number of such articles had been indexed before in this database. So they were taken and placed on HTML pages. Problems encountered at this stage were as follows:

1. There were orthographic problems in the titles of the articles taken from Library's Database.
2. The journals' indexing process was incomplete, some articles had not been indexed, thus there was a need to add the title of such articles.
3. Often titles are repeated on more than one occasion in a single issue of a journal.

Example

In certain journals the part related to introducing different books has just one title, 'Introducing Books', and the title of the book is not mentioned. To solve this problem, one title, 'Introducing books' was used – exactly the same format as in the table of contents of journals.

Scan of Articles

In this stage, articles are scanned, and each article receives a code based on the number assigned to that document in the indexing process. Those pages of the article that include color pictures or photos are scanned with the format of *grey scale*[4] and, thus, such pages occupy more space compared to pages lacking such photos.

Making Links

The title of each article is linked to its full-text version. This is achieved when the HTML pages are made.

Preparing HTML Pages Related to the Issues and Volumes of Each Journal

During this stage, the title of the journal as well as the cover photo are placed on the first page. When all the issues of a given journal are made available for a certain year, they are written down together with the information concerning the number of issues. Later by selecting a given year, all the issues related to that year are displayed. Finally a link is established between each issue and the table of contents of that particular issue.

Producing Alphabetically Ordered Pages

Thirty-two HTML pages were produced, corresponding to the 32 letters of the Persian alphabet. These pages include titles of journals all the issues of which have been processed. These titles have been arranged alphabetically. A link is also established between each journal title and the HTML page, and the final analysis is carried out.

The Use of the Persian E-Journal Database via the Internet

A standard proposed and used by Elsevier, a valid corporation, was employed to illustrate journals in the Persian E-journals Database.

First go to RLST's homepage, www.srlst.com, and click on 'E-journal' in the section entitled, 'Farsi Databases' (Figure 1). After entering the ID PASSWORD, a page will open containing all letters of the Persian alphabet. Journals have been arranged alphabetically. Selecting a certain letter will provide us with a list of all journals starting with that letter. In case the name of a given journal is clicked on, all the years related to that journal will be listed. Clicking on each year will give way to all issues and numbers related to that year. Then, the user may select the specific issue required. Clicking on a given article, from the table of contents, will take the user to the full-text of that specific article, in pdf[5] format. At this stage, the content of the article could be stored or printed. Of course, in order to see the full-text article, the computer system used by the user must have the Adobe Acrobat Software. Let's give an example:

Selecting the Persian letter 'K' all the names of journals starting with this letter will be listed (Figure 2). If the name of a given journal is selected, the relevant information related to that journal will be shown. For example, selecting the title, *Ketābdār va Ettelā' Rasān* "Librarianship and Information" all the years related to this journal are given as a list – the availability of each year means that all issues related to that year have already been processed (Figure 3). Clicking on a certain year, i.e. 1380/2001 will provide the user with the issues related to that year. Similarly, clicking on a given issue, i.e. volume 4, no. 2, 2001/1380, will take us to the table of contents of that specific issue (Figure 4). At this stage, if the title of a certain article is clicked on, i.e. A'zāye Hei'at-e Elmi-e Dāneshgāhe Shiraz va Tovlid-e Ettelā'āt-e Elmi: Barresiye Kammi Salhaye 1369-1378 (Faculty Members of Shiraz University and Production of Scientific Information: A Quantitative Review 1995-1999) the full-text of the article will be provided.

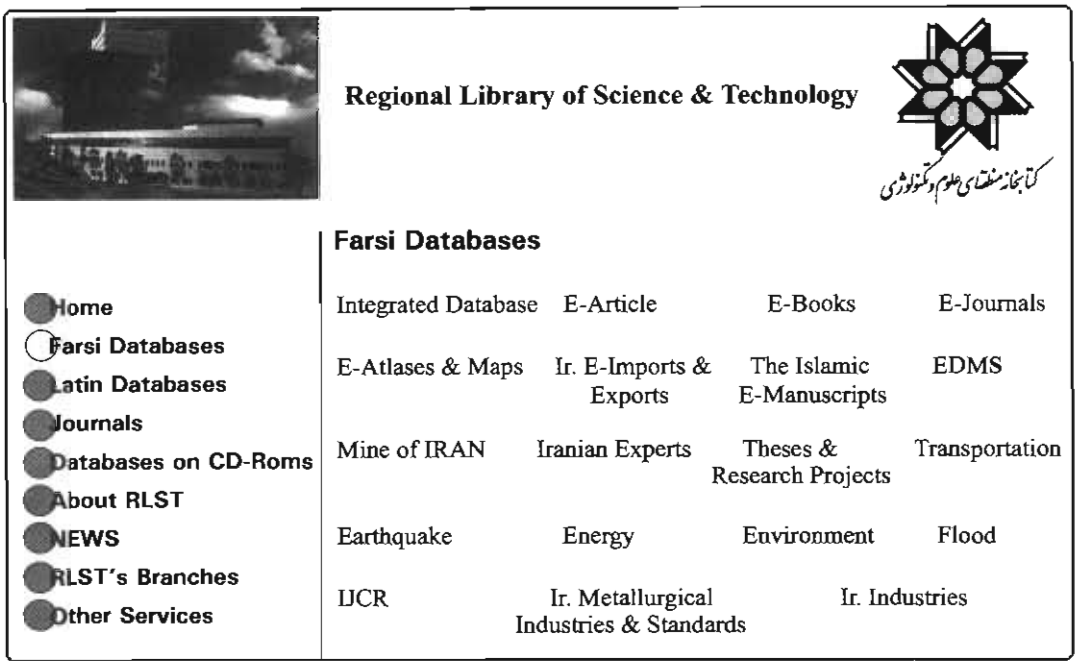

Figure 1: RLST's Homepage

- *Kārshenās*
- *Kāmposit*
- *Ketābe Māh-Olūm Ejtemā'i*
- *Ketābdār*
- *Ketābdār va Ettelā' Rasān*
- *Kelk*

Figure 2: The List of Journal Names Starting with the Persian Letter 'K'.

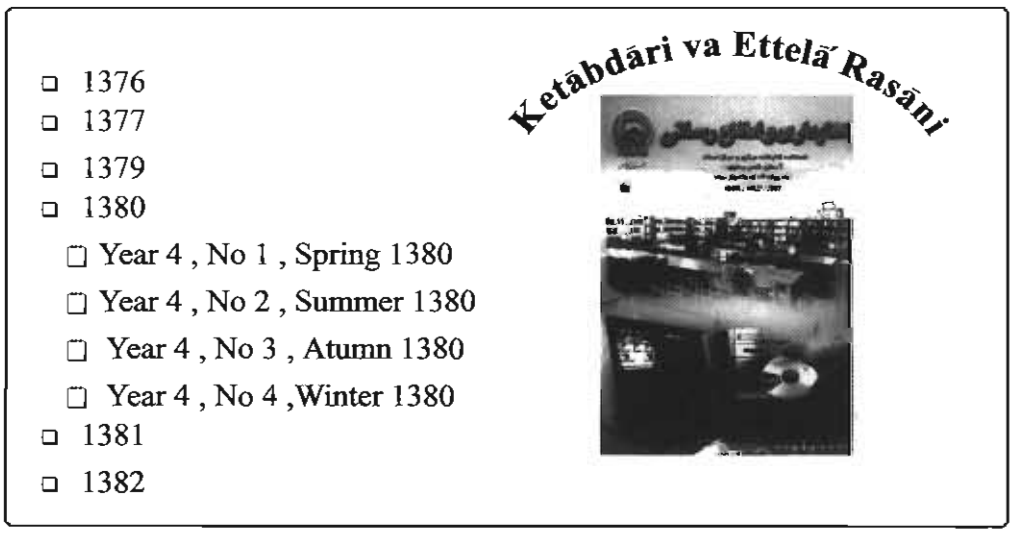

Figure 3: Different Volumes, Related to Different Years, of the Journal of *Ketābdār va Ettelā' Rasān.*

Ketābdār va Ettelā' Rasān
Year 4, No 2, Summer 1380

- Anjoman-e Ketābdār va Ettelā' Rasān.
- Neeyaz be Tahavvol dar Barnāmehāye Darsi-e Kārshenās Ketābdār va Ettelā' Rasān
- A'zāye Hei'at-e Elmi-e Dāneshgāhe Shiraz va Towlid-e Ettelā'āt-e Elmi: Barresiye Kammi Sālhaye 1369-1378.
- Barresi Avāmel Mo'asser Bar Bahrevari-e Ketābdārān-e Ketābkhānehāye Dāneshgāh-e Olūm Pezeshki-e Mashhad
- A'n-e Ketābat dar Noskhehāye Khatti-e Gharn-e 7 tā 9 dar Ketābkhāne-ye Āstān-e Qods-e Razavi
- Tavānmandsāzi, Tavānmandsāzān: Ta'n Jaygāh-e Herfe-mandān-e Ketābdār va Ettelā' Rasān dar Āyande Tose'e-ye Reshte va Herfeye Olūm Ketābdār va Ettelā' Rasān.

Figure 4: The Table of Contents of the Journal of *Ketābdār va Ettelā' Rasān*, No. 2, Summer 2001/1380.

Problems

A large number of problems were encountered during the full-text production of Persian journals, some of which are mentioned below:

1. Sometimes a title is introduced in the table of contents, the full-text of which is absent in that issue.

Example: In the table of contents of the journal entitled, *Tāzehāye Terāfik* vol. 3, no. 9, spring 2001/1380, there is the article, " System Āshkār-sāzi-ye Terāfik-e Āzād-rāh" but no article with this title is really found in the table of contents of the journal. In fact, another article, "Tarrahi-e Docharkhe-ro-hāye Shahri" is on that page.

Solution: In such cases the title of the article that appears on the article must replace the article provided in the table of contents.

2. Some titles that exist in the content of a journal are non-existent in the table of contents.

Example: In the journal called, "*Peyvand,*" No. 271, Ordibehesht 2002/1381, the title, "Āshenā' bā Ketāb" could be observed of which there is no mention in the table of contents.

In the journal, "*Nāfeh,*" vol. 1, no. 11 & 12, Farvardin & Ordibehesht 2001/1380, there is no mention of the article, "Negareshe Khātami be Moshārekate Siāsiye Zanān" in the table of contents. However, the article appears on page 63 of the journal.

In the journal, "*Hesābras,*" vol. 4, no. 18, Bahman & Esfand 2002/1381, there is no mention of the article, "Ā'n Nāmeye Arzyābi-ye Amalkarde Dastgāhhāye Ejrā'-e Keshvar" in the table of contents.

Solution: The titles of such articles were added to the table of contents and a link was made to the full-text articles.

3. Some discrepancies are found between the titles given in the table of contents and those given in the actual content of the journal.

Example: In the journal, *Sanāye,* vol. 7, no. 3 & 4, summer and fall 2001/1380, the title, "Sākhtāre Kalān Projeh" is found, which appears in the content as, "Tarrāhi-e Sākhtāre Kalān Projeh."

4. There are cases where an article cannot be found on the pages given in the table of contents.
5. With respect to certain journals, varying numbers of issues are published each year.

Example: The journal entitled, *Majaleye Hoghūghi* has been published with varying numbers of issues in different years.

6. Information related to volumes and numbers are not standard in certain cases.

Example: With respect to the journal *Nashriye-e Daneshkade-e Adabiyyāt va Olūm-e Ensāni-e Dāneshgāhe Tabriz* the following information is found:

Year 45, no. 1, spring 2002/1381.
Year 45, nos. 2 & 3, summer and fall 2002/1381.
Year 45, winter 2002/1381.

7. In journals that have an English section, there are problems in numbering the pages and the table of contents.

Example: In, "Maghāleye Elmi Pezeshki-e Ghānūni," vol. 7, no. 22, spring 2001/1380, the English abstracts start from page 76, rather than 74, as mentioned in the table of contents.

In, "*Ta'lm va Tarbiyat,*" vol. 16, no. 1, spring 2000/1379, the page numbers related to the English articles are written using Persian numericals in the table of contents, but with English numericals in the text.

8. Incompleteness of some articles.

Example: In the journal, *Pajūheshhā va Siāsathāye Eghtesādi*, vol. 2, no. 2, fall 1994/1373, the article, "Nazāme Novin-e Jahāni va Moghe'iyyat-e Iran" is expected to appear on pages 61-89, of which only four pages are given on pages 61-64.

In the journal, *Pajooheshhā va Siāsathāye Eghtesādi*, vol. 2, no. 2, fall 1994/1373, the article, "Tajziye va Tahlil-e Jaryān-e Vojohe Naghd dar Bakhshhāye Eghtesādi Iran" is expected to appear on pages 90-112, of which only pages 97-112 appear in the content.

In the journal, "*Esteghlāl,*" vol. 17, no. 1, Shahrivar 1998/1377, the article entitled, "Estefāde az Yek Ravesh-e Jadid Mote'āref Sākhtan-e Antegrālhāye Monfared Se Bo'di," which appears on pages 121-132, pages 121, 124 and 125 are blank.

9. Many Persian periodicals are published using low quality paper, which makes the scanning problem more difficult.
10. Problems related to the publication of journals will cause subsequent problems in the production of e-journals. For example one or more pages of many articles are blank or have not been printed appropriately.

CONCLUSION

Developing information databases has been one of the most important responsibilities of RLST. Therefore, the creation of the Persian Full-Text Electronic Journal Database should be considered as one of the most important achievements of RLST since its establishment in 1991.

With the establishment of this database, RLST has taken a significant step forward in developing digital libraries in the country. Despite the large number of problems existing concerning Persian journals and periodicals, RLST is trying hard to provide the users with the content of these journals through this database.

The final objective has been to include the largest number of scientific research and applied research journals and periodicals, which are mostly published on a monthly, bi-monthly or tri-monthly basis.

The process of producing e-journals will be made easier and faster with a Persian OCR software that has the capability of converting visual files into text files correctly.

Finally, if the set standards, i.e. paper quality, printing quality, assigning correct volume and numbers to each issue, etc., for the publication of journals are observed by different publishers, we will be empowered to implement the Persian E-Journals database on a qualitative basis that can compete with rival databases throughout the world.

ENDNOTES

1. Zines refer to publications with limited circulation, between 15-2500 copies. They usually cover controversial topics, i.e. the life of a publisher, the latest secret pieces of music, etc. Zines are better known with their privatization rhythm, rather than the subject areas they cover. To define the term, it will be much more easier to identify what it is not. They are not professional publications, although some of them delve into details. Generally speaking, they do not make much money. The area that comes between zines and magazines is called 'Zine.' Such publications have a circulation more than 2500, and even 10,000 copies have an attractive and colored cover, and

in some cases are profitable. They cover subjects that are often not dealt with in magazines and ordinary publications. They are published independently and are often more attractive than zines, though they are not as attractive as colored journals often found at the newspaper booth.

2. HTML is a coding system used to produce hypertext documents in such a way that they can be used on the Web. It includes a number of codes, attached to a file, by the use of which the user's browser can arrange the texts and pictures in a way that matches the main page designed.
3. OCR is a type of process that converts a printed or written text into a file that is based on electronic letters. This process requires that each letter of the text be scanned and analyzed, and that the visual letters be rendered into digital codes – usually Aski codes are used. During the scanning process, a visual copy of a page is scanned, and in order to identify each letter of the alphabet, the gray and bright scales are analyzed.
4. Grey scale refers to a grey area in a picture. In scanners, it could be identified through the number of grey spectra or the amount of black and white colors on a visual page.
5. Portable Document Format (PDF) is a file format produced by Adobe. It provides a standard format for storing or editing of printed materials. Documents produced in this format can be accessed and retrieved through different systems. This format is quite common on the web .

REFERENCES

[1] Bennion, Bruce C., "Why the Science Journal Crisis?" *Bulletin of the American Society for Information Science*, 25-26, March 1994 .

[2] Colorado Alliance of Research Library e-journal definition; accessed January 2004 <http://www.stcc.cc.tx.us/main/e-journalsearch.html)>

[3] De Solla Price, Derek J. *Little Science, Big Science*, Columbia University Press, New York, 1965.

[4] Dijkstra, Joost. *A Digital Library in the Mid-Nineties: Ahead of Schedule*, Information Services and Use, 1993.

[5] Hitchcock, S., et. al. "A Survey of STM Online Journals 1990-1995: the Calm before the Storm," Open Journals Framework Project.

[6] Hitchcock, S. et. al. "Web Journals Publishing: a UK Perspective," Serials 10 (3): 285-299.

[7] Hoffman, Melia M. et. al., "The RightPages TM Service: An Image Based Electronic Library," *Journal of the American Society for Information Science*, 44(8), 446-452, 1993.

[8] Hunter, Karen, "Three views of the Internet" in Ann Marie Cunningham and Wendy

Wicks, eds., NFAIS Report Series no. 3, NFAIS, Philadelphia, PA, 1993.

[9] Metz, Paul and Gherman, Paul M., "Serials Pricing and the Role of the Electronic Journal," *College & Research Libraries*, 52 (4), 315-327, 1991.

[10] Some examples of the first electronic journals are as follows: Public-Access Computer System Review (1990-1998) <http://info.lib.uh.edu/pr/pacsrev.html>; Postmodern Culture (1990 to date). Text-only archive at <http:www.iath.virginia.edu/ pmc /text-only/back-contents.html>; New Horizons in Adult Education (1987 to date) <http;//www.nova.edu/~aed//newhorizons.html>; Newsletter on Serials Pricing Issues (1989 to date) <http:www.lib.unc.edu/prices/>;and Psycoloquy (started 1989) <http://www.cogsci.ecs.soton.ac.uk/psycoloquy/)>

[11] Stern, B.T., and Campbell, R.M. "ADONIS, Publishing Journal Articles on CD-ROM," Advances in Serials Management, 3, 1-60, 1989.

[12] *The Oxford English Dictionary*, 2nd ed., Oxford University Press. Definitions taken from online the version <http://dictionary.oed.com/>

[13] The rise in the number of electronic journals between 1990-1997 could be studied in the following source: "ARL Directory of Scholarly Electronic Journals and Academic Discussion Lists," at <http://db.arl.org/dsej/2000/mogge.html>