SciELO – An electronic publishing model for developing countries

Abel Laerte Packer
scielo@bireme.br
Director, BIREME/PAHO/WHO

1. Introduction

The SciELO – Scientific Electronic Library Online Project (www.scielo.br) is becoming a common model for the cooperative publishing of scientific and technical electronic journals in Latin American and Caribbean countries.

Since its launch in March 1997, promoted by a partnership pilot project led by FAPESP[1] (www.fapesp.br), BIREME[2] (www.bireme.br) and a group of Brazilian scientific editors, the SciELO Model was already adopted by Chile, and there are several ongoing initiatives that lead us to expect that five more countries will adopt SciELO during 1999. By the end of the year 2000, it is expected that at least 100 journals will be published under the SciELO Model throughout Latin America and the Caribbean.

The SciELO Model aims at the operation of electronic libraries of scientific journals on the Internet. It comprises two major components: the methodology for electronic publishing, and the Internet site that actually operates the journal collections.

2. SciELO Model's brief history

The SciELO Model was developed by the project conceived in 1997 by BIREME and FAPESP for the development of a methodology for the preparation, storage, evaluation and dissemination of electronic publications. The methodology was intended to be applied in the operation of electronic journal collections (Perez[3], Meneghini[4], Packer et al[5]).

The project was launched in March of 1998 with the engagement of 10 Brazilian scientific editors as partners in the development of the methodology, as well as on the pilot operation of an online library of electronic journals to test the methodology.

The pilot project was finished in May 1998. A second phase started aiming at the development of a national scientific electronic library, which is planned to operate from 75 to 100 Brazilian scientific and technical journals. In March 1999, the SciELO Brazil already operates 27 journal titles, totalling more than 2,400 articles.

In September of 1998, the Departamento de Información of the Comisión Nacional de Investigación Científica y Tecnológica - CONICYT of Chile, with the collaboration of the Chilean National System on Health Sciences, started the implementation of the SciELO Chile as a cooperative project with SciELO Brazil. SciELO Chile is expected to be released by April 1999, with a collection of 10 journal titles (www.scielo.cl).

Colombia, Cuba and Venezuela have also decided to adopt the SciELO Model, and the implementation of the preliminary activities has already started. National collections are expected to be publicly operated in those countries by August 1999.

BIREME has adopted the SciELO Model for the publishing of the scientific literature in the context of the Virtual Health Library (VHL). The VHL will operate a network of decentralized health sciences information sources throughout Latin American and Caribbean countries. In principle, such a decision implies that the SciELO Model will be adopted by most of the countries of the region in the next years, at least for the operation of national health sciences literature.

3. The SciELO Methodology

The SciELO Project was conceived with the general objective of bringing a contribution to the advancement of Brazilian scientific research through the improvement of mechanisms and procedures of scientific communication by using information technology, specially the Internet. Three main specific contributions were also envisaged. First, to facilitate and make feasible the transition of scientific communication from printed publication to electronic publishing. Second, to increase the accessibility and visibility of journals through their publication on the Internet. And third, to establish procedures for the evaluation of the usage and impact of scientific journals.

From its very beginning, SciELO was planned to be a common methodology. Actually, one of the justifications for its development was to strengthen the challenging process facing Brazilian editors and publishers with reference to the transition to electronic media. The usage of a common methodology would decisively contribute to overcome those challenges. It would also contribute to increase accessibility and visibility by providing universal access via a common interface.

The SciELO Methodology is a set of tools for the preparation, storage, publication and evaluation of scientific and technical journals. It comprises guides, standards, manuals and software (Packer et al, Packer[6]).

Although the SciELO Methodology can be used to operate individual titles on the Internet, it is oriented for operating collections of journals. The SciELO Methodology presents the following basic characteristics:

- Its application does not affect the process of the print publication, because the methodology makes possible the use of the same electronic files already employed in the printing process. This characteristic is specially useful in the promotion of the transition to electronic publishing, once editors and publishers are not asked to take radical decisions regarding the media, process or procedures they are using for paper printing.
- Journals editorial policy and individual characteristics are preserved. The methodology allows full operation of individual journals in the library collections.
- It stresses the obedience and compatibility to international standards, as well as to the most recognized practices in electronic publishing carried out among the international community of scientific communication. This characteristic is of high importance to assure that SciELO journals may be interoperable with international electronic journals.
- The information technology based tools follow open architectures on both hardware and software in order to be operable in an affordable way in the most common infrastructure found in Latin America and Caribbean.

The current version of the SciELO Methodology implements the electronic publication of a journal issue as a technical process, including the following modules:

- DTD Module – comprises a collection of SGML Document Type Definitions (DTD) that describes the structure of SciELO texts. As a general principle, the SciELO DTDs focus on the actual structure and content of the texts, specially the bibliographic data elements, while the presentation of the texts uses the HTML DTD.
- Markup Module – implemented as an add-on over MS Word Processor, it manages the markup process of the elements of articles and other texts according to the correspondent DTD. It is operated in desktop computers under MS Windows.
- Conversor Module – parses and normalizes the marked documents, and loads them into a full text database to be later exported to the Internet server which operates the SciELO site database. The Conversor Module is also responsible for exporting data to external systems. It is operated in desktop computers under MS Windows.
- Interface Module – manages the entire SciELO Internet site, including the full text database and the http client-server based interface. It is operated both under MS Windows and different Unix operating systems.
- Bibliometric Module – produces reports of usage of the journals, their issues and articles, as well as indicators of citation and impact. This module permits the generation of bibliometric indicators compatible with those of the ISI (Institute for Scientific Information).

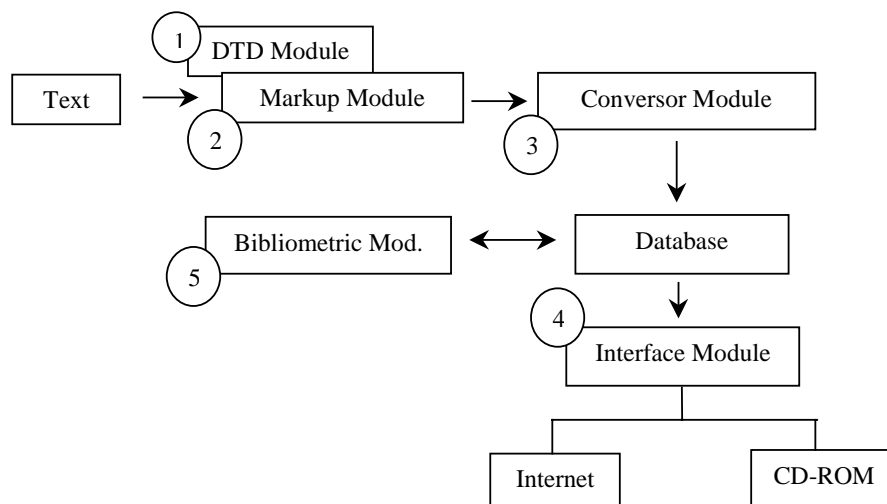The following figure depicts the flow of data through the different modules:

Figure 1 - SciELO Methodology data flow diagram

4. The Internet SciELO site

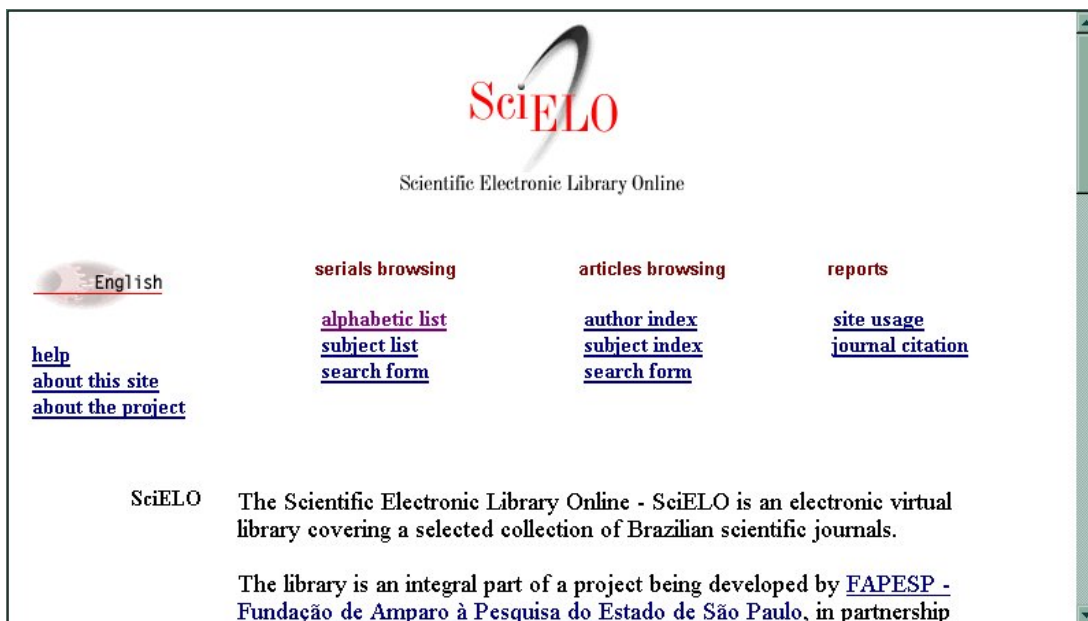The principal product of the SciELO Model is the Internet site that operates an electronic journal collection.

Figure 2 - Entry page of the SciELO site
The site interface offers access to individual titles through alphabetic and subject lists.

Figure 3 - Example of a SciELO alphabetic list of journal titles

Once a title is selected, its home page is displayed, where links provide access to information about the journal, to the holdings, to individual issues, as well as to abstracts and full texts of articles. Full texts are available online (in HTML), and also in PDF. The figures below are examples of SciELO interface pages related to individual titles.
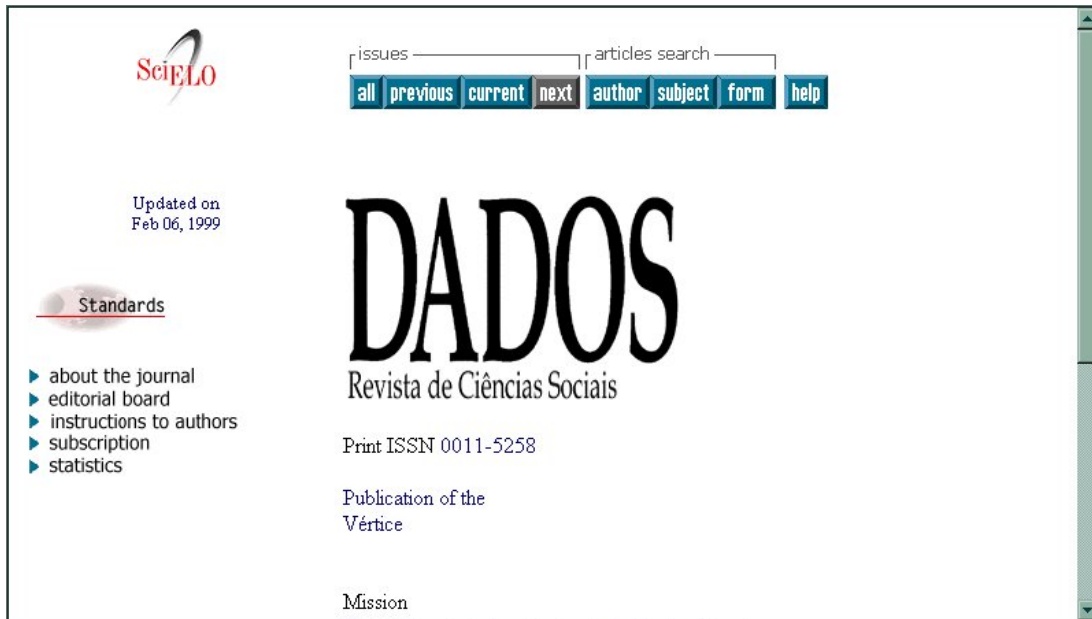
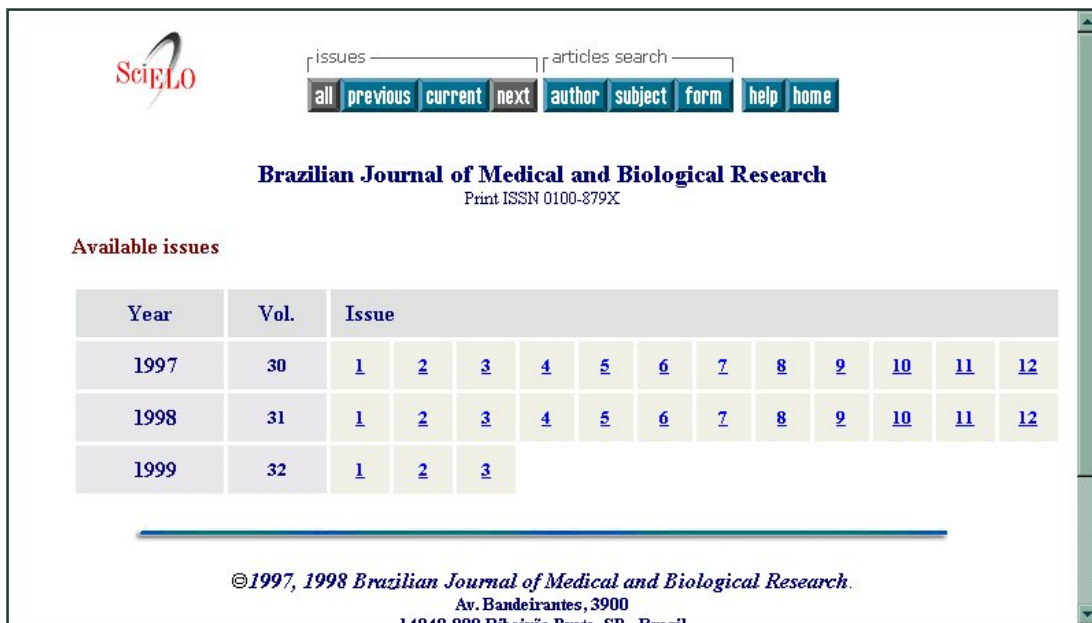*Figure 4 - Example of a title main page*
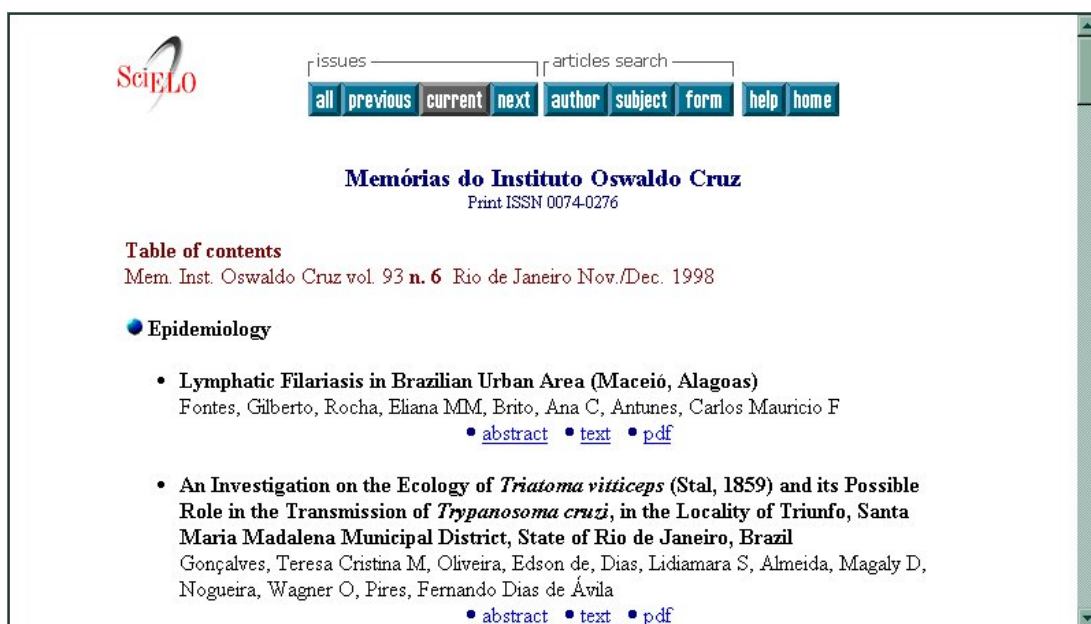

Figure 5 - Example of a holdings page

Table of contents
Mem. Inst. Oswaldo Cruz vol. 93 **n. 6** Rio de Janeiro Nov./Dec. 1998

● **Epidemiology**

Figure 6 - Example of an issue table of contents page

Abstract

Chang, Y.K., Martínez-Bustos, f. and Lara, h. Effect of Some Extrusion Variables on Rheological Properties and Physicochemical Changes of Cornmeal Extruded by Twin Screw Extruder. *Braz. J. Chem. Eng.*, December 1998, vol.15 no.4. ISSN 0104-6632

The effect of extrusion variables, such as barrel temperature (100 to 170°C), feed rate (100 to 500 g/min), feed moisture (20 to 40 g/100 g wet basis), screw speed rate (from 100 to 500 rpm), and slit die rheometer configuration (0.15 and 0.30 cm height) were studied using a co-rotating intermeshing twin-screw extruder coupled to a slit die rheometer on the rheological properties of yellow cornmeal. An increase in feed rate decreased WAI and WSI, but increased the viscosity values. The temperature interacts strongly with screw speed in affecting the WSI. The most important factor in starch degradation was the screw speed. Increasing the screw speed completely modifies the organised structure of starch (crystalline region).
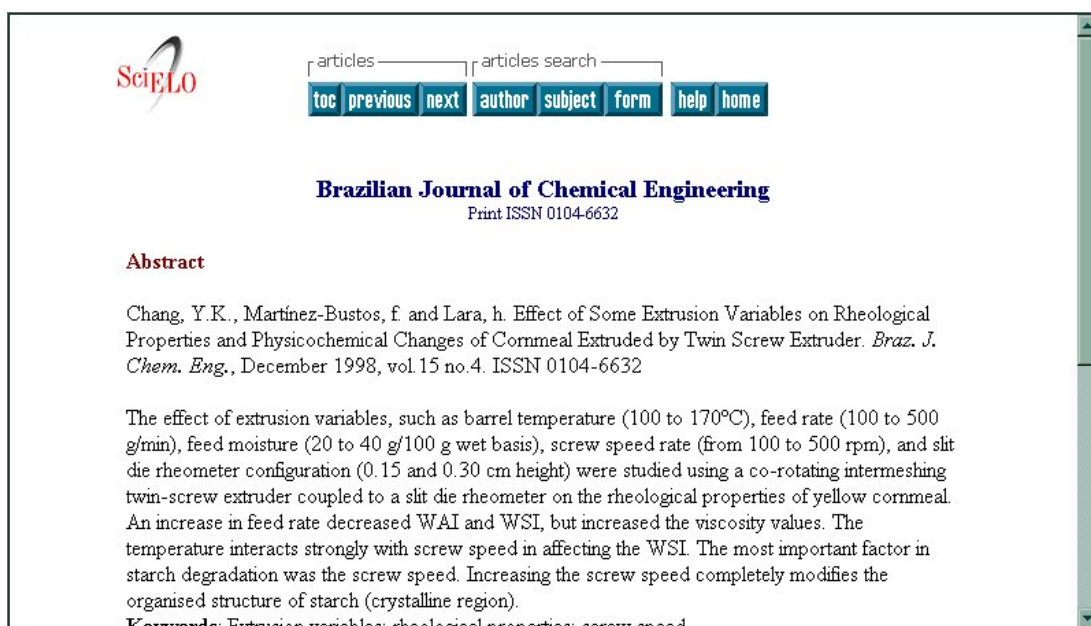**Keywords**: Extrusion variables; rheological properties; screw speed

Figure 7 - Abstract page

Providing Context to Web Searches:
The Use of Ontologies to Enhance Search Engine's Accuracy

Flávia A. Barros, Pedro F. Gonçalves and Thiago L.V.L. Santos
Departamento de Informática
Universidade Federal de Pernambuco - Caixa Postal 7851
50.732-970 Recife (PE) – BRAZIL
fab@di.ufpe.br , pfg@di.ufpe.br , tlvls@di.ufpe.br

6 Evaluation of Precision and Recall

This section presents an statistical evaluation of the performance of OMF using the Brazilian BRight! IB.

Recall and precision are two inherently conflicting metrics. That is, when an IR system is enhanced to increase precision, recall generally degrades, and vice-versa. A more convenient metrics to evaluate the overall performance of an IR system is the *F-Measure* [28, 10], given by:

$$F\text{-Measure} = \frac{(2*Recall*Precision)}{(Recall + Precision)}$$

Our goal is to improve the F-measure of the system, bringing it as close as possible to 1. Note that F-measure only goes up when both precision and recall go to 1 simultaneously. When only one of these indicators goes to 1 (and the other goes to 0), the F-measure goes down.

The Web size is currently estimated in about 320 million pages in 1.5 Terabytes [17]. With these figures, the Web exceeds by orders of magnitude the document collections used for research in Information Retrieval, "which have recently reached 7.5 million documents in 20 Gb for the Very Large Corpus track." [29].

Due to the Web's size and dynamics, it is infeasible to assess retrieval quality by the standard recall measures since, for each query, every Web document would have to be inspected by a human and marked relevant or irrelevant. Instead, *relative recall* can be used [25, 10]. The trick with relative recall is that a set of queries about a given topic is considered as a whole, and all the relevant documents retrieved for this set of queries is then regarded as "the set of

[3] R. Dolin, D. Agawal, A. El Abbadi, L. Dillon. Pharos: A Scalable Distributed Architecture for Locating Heterogeneous Information Sources. *In Proceedings of the Sixth International Conference on Information and Knowledge Management (CIKM'97)*. Nevada, pages 165-169. 1997. http://pharos.alexandria.ucsb.edu/publications/

[4] R. Dolin, D. Agawal, A. ElAbbadi, J. Pearlman. Using Automated Classification for Summarizing and Selecting Heterogeneous Information Sources. *In D-Lib Magazine*. 1998. http://www.dlib.org/dlib/january98/dolin/01dolin.html

[5] S. Dumais. Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments & Computers*. 23(2): 229-236, 1991.

[6] E. A.Fox. Rethinking Libraries in the Information Age: Lessons Learned with Five Digital Library Projects. School of Information & Library Science, UNC Chapel Hill. 1996. http://fox.cs.vt.edu/talks/UNC96/

[7] W.B. Frakes. Introduction to Information Storage and Retrieval Systems. In Frakes, W.B. and Baeza-Yates, R. (eds.), *Introduction to Information Retrieval - Data Structures and Algorithms*. Prentice-Hall, New Jersey, pages 1-12, 1992.

[8] P.F. Gonçalves, A.C. Salgado, S.L. Meira. Digital Neighbourhoods: Partitioning the Web for Information Indexing and Searching. In Olivè, A., Pastor,J.A. (Eds.) *Advanced Information Systems Engineering. 9th International Conference (CAiSE'97)*. Barcelona, Spain, June 1997. Springer Verlag, Lecture Notes in Computer Science 1250: 289-302. http://www.bright.org.br/ .

Figure 8 - Three images illustrate an article page: front, body of text and bibliography

The site interface also permits access to individual articles through author and subject indexes. It is also possible to retrieve a set of articles through boolean search on authors, title words, subject etc. From the indexes or searches it is possible to display abstracts and full texts of individual articles. The following figures show pages related to the access to individual articles through indexes and search form.

SciELO

serials browsing ─── articles browsing ───
alpha | subject | form | author | subject | form | help

Library collection

Author index - selection

722 names found!
Select one or more names and press the Search button to display articles references.
Displaying from 51 to 100

CAMACHO, LUIZ A. B.
CAMACHO, ROBERTO GOMES
CAMADRO, ELSA
CAMARA, G.M.S.
CAMARA, G.N.N.L.
CAMARA, MILENA
CAMARA-LOPES, L.H.

Display references in the format defined by the ISO 690 standard.
Display up to 10 references per page.

Search

Browse index

50 Previous | Next 50

275

Figure 9 - Author index

Figure 10 - Form search

Figure 11 - Result of a search

On the site interface and on the home page of each journal title, there are links to quantitative reports, which present the distribution of access, the distribution of the content and citation based reports.

Library collection

Journal citation reports

Source data listing as of Mar 16, 1999

▼ - click to select column order
▲ - indicates current order

Source data statistics on SciELO articles published in the years indicated									
journal title / year ▼	no. of issues ▼	no. of articles ▲	no. of citation ▼	average articles per issue ▼	average citations per issue ▼	average citations per article ▼	% of total issues ▼	% of total articles ▼	% of total citations ▼
Braz J Med Biol Res	27	459	14117	17.00	522.85	30.76	17.88	18.46	27.11
1997	12	206	5850	17.17	487.50	28.40	7.95	8.28	11.23
1998	12	203	6906	16.92	575.50	34.02	7.95	8.16	13.26
1999	3	50	1361	16.67	453.67	27.22	1.99	2.01	2.61
Mem. Inst. Oswaldo Cruz	13	350	5994	26.92	461.08	17.13	8.61	14.07	11.51
1997	6	160	2824	26.67	470.67	17.65	3.97	6.43	5.42

Figure 12 - Source table sorted by number of articles

5. The establishment and development of a SciELO site

The SciELO Model implements national libraries of scientific electronic journals, covering different subject areas or a specific area such as health sciences.

The successful development of a SciELO national collection requires, in principle, the participation of one or more national institutions in the field of scientific communication, such as science councils, scientific information systems, scientific editors associations, as well as the participation of individual editors and publishers. It is expected that different consortia configurations will be set up in the countries. Two main national entities – the national council of science and the national system on health sciences information linked to BIREME – are expected to play key roles in most of the national SciELO sites.

A SciELO national collection requires a central unit to be responsible for its overall operation, which includes planning, management and execution of activities and tasks related to funding, staff administration, technology infrastructure, relationship with editors, marketing, reception of data from editors, managing the preparation of the electronic journal issues through the SciELO methodology modules, relationship with external systems, and the operation of the Internet SciELO site server.

It is possible and even recommended that the preparation of electronic journals, specially the markup process, be carried out through decentralized units, including individual publishers. In those cases, the marked files are sent to the central unit for inclusion in the SciELO site server.

SciELO sites mirrors may be implemented in order to increase their availability and accessibility, as well as to assure backup and preservation.

From the previous description, it is not difficult to conclude that the enterprise of establishing and developing national SciELO collections is highly complex in several aspects.

First of all, a national decision needs to be taken by scientific related authorities to sponsor and promote the establisment of a national SciELO.

Second, it is necessary to obtain the effective participation of scientific editors and publishers as active partners in the process.

Third, an institution with the required managerial, technical expertise, and infrastructure in information technology is required to manage the SciELO central unit.

Fourth, it is necessary the establishment and continued improvement of criteria and procedures for the evaluation of scientific journals regarding the selection process to enter and remain in the SciELO collections.

Fifth, it is necessary to develop economic models based on local conditions that combine the objective of increasing the accessibility and visibility of scientific production with the demand for sustainability.

In order to face such a complex enterprise in an efficient and effective way, the SciELO Model calls for the technical cooperation among national and international related organizations. This strategy calls for a rationalization in the usage of the scarce national resources in order to maximize outputs relating the creation and development of electronic collections, as well as in relation to further developments of the SciELO Methodology.

6. Conclusion

The SciELO Model comprises an ample aproach to the electronic publication for developing countries. On one side, it provides a powerful and affordable methodology for managing collections of electronic journals on the Internet, including a secure transition process from paper to electronic media. On the other side, it increases the accessibility and visibility of scientific journals. In addition, the SciELO Model creates an environment that induces the improvement of the quality of the scientific communication.

The adoption of the SciELO Model by several Latin American and Caribbean countries will contribute to the development of a regional cooperative program toward the establishment and operation of national and regional indexes for the control, dissemination and evaluation of scientific communication.

[1] FAPESP – Foundation for the Promotion of Scientific Research of the State of São Paulo.

[2] BIREME – Latin American and the Caribbean Centre on Health Sciences Information / Pan American Health Organization / World Health Organization.

[3] Perez, José Fernando. Seminário sobre Avaliação da Produção Científica: sessão de abertura. *Ciência da Informação*, Brasília, vol. 27, no. 2, p. 217-8, 1998. Available at WWW: [http://www.scielo.br/cgi-bin/fbpe/fbtext?got=last&pid=S0100-1965(98)02700217&usr=fbpe&lng=en&seq=0100-1965-005&nrm=iso&sss=1&aut=71981947]

[4] Meneghini, Rogério. Avaliação da produção científica e o Projeto SciELO. *Ciência da Informação*, Brasília, vol. 27, no. 2, p. 219-20, 1998. Available at WWW:

[http://www.scielo.br/cgi-bin/fbpe/fbtext?got=next&pid=S0100-1965(98)02700218&usr=fbpe&lng=en&seq=0100-1965-005&nrm=iso&sss=1&aut=71981947]

[5] Packer, Abel Laerte; Biojone, Mariana Rocha; Antonio, Irati et al. SciELO: uma metodologia para publicação eletrônica. *Ciência da Informação*, Brasília, vol. 27, no. 2, p. 109-21, 1998. Available at WWW: [http://www.scielo.br/cgi-bin/fbpe/fbtext?got=last&pid=S0100-1965(98)02700202&usr=fbpe&lng=en&seq=0100-1965-005&nrm=iso&sss=1&aut=71981947]

[6] Packer, Abel Laerte. SciELO, Scientific Electronic Library Online. Paper presented to the ICSU Press Workshop, Keble College, Oxford, UK, 31 March to 2 April 1998. Available at WWW: [http://www.bodley.ox.ac.uk/icsu/packerppr.htm]