

BIREME / PAHO / WHO

Latin American and Caribbean Center on Health Sciences Information

SciELO Methodology

Converter

Version 3.1

São Paulo - 2005

Copyright © 2005 - BIREME / PAHO / WHO

Converter

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

Card catalog

BIREME / PAHO / WHO (Brasil)

Converter. / BIREME (org.). São Paulo : BIREME / PAHO / WHO, 2005.

29 p.

1. User manual. 2. Information access. 3. Information systems. 4. Information management. 5. Public health. 6. Public Health services. I. BIREME II. Title

Warning - Any mention in this document to companies, institutions, persons or products are not an endorsement or recommendation given by BIREME / PAHO / WHO, thus it does not mean a preference to a similar one, cited or not.

BIREME / PAHO / WHO

Latin American and Caribbean Center on Health Sciences Information

Rua Botucatu, 862 - V. Clementino

This document was produced with the Documents Conformation Methodology (NorDoc) developed by BIREME.

Methodology document set

The complete set consists of 9 documents:

1. SciELO Model Guide
2. Secondary Pages Creation and Update
3. File Preparation Procedures
4. Installing PC Programs
5. Code Manager and Title Manager
6. Markup and Parser
7. **Converter**
8. Local Site
9. SciELO Processing Procedures

Table of contents

Methodology document set	I
Abbreviations used	III
How to use this manual.....	VI
1 Preface.....	1
1.1 About BIREME.....	1
1.2 The Virtual Health Library (VHL)	3
1.3 About the SciELO Methodology.....	4
2 Introduction.....	7
3 Directory Structure.....	8
4 Requirements to initiate conversion.....	9
5 Using Converter program.....	10
6 Main problems.....	14
7 Bibliographic references	15
8 Glossary.....	16

Abbreviations used

- AACR2. Anglo-American Cataloguing Rules - 2nd Edition.
- ABNT. Associação Brasileira de Normas Técnicas [Brazilian Association of Technical Standards]
- ASCII. American Standard Code for Information Interchange.
- BIREME. Latin American and Caribbean Center on Health Sciences Information.
- BVS. Biblioteca Virtual em Saúde (*see* VHL).
- CGI. Common Gateway Interface.
- CNPq. Conselho Nacional de Desenvolvimento Científico e Tecnológico [National Council for Scientific and Technological Development].
- CNS. Conselho Nacional de Saúde [National Health Council (Brazil)].

- CSS. Cascading Style Sheet.
- DeCS. Health Sciences Descriptors.
- DTD. Document Type Definition.
- FAPESP. Fundação de Amparo à Pesquisa do Estado de São Paulo [The State of São Paulo Research Foundation].
- FAQ. Frequently Asked Questions.
- HTML. HyperText Markup Language.
- HTTP. HyperText Transfer Protocol.
- ISBN. International Standard Book Number.
- ISI. Institute for Scientific Information.
- ISO. International Organization for Standardization.
- ISSN. International Standard Serial Number.
- LILACS. Latin American and Caribbean Health Sciences Literature.
- MEDLINE. Medical Literature Analysis and Retrieval System Online.
- NLM. National Library of Medicine.
- PAHO. Pan American Health Organization.
- PDF. Portable Document Format.
- SciELO. Scientific Electronic Library Online.
- SeCS. Serials in Health Sciences.
- SGML. Standard Generalized Markup Language.

- URL. Universal Resource Locator.
- WHO. World Health Organization.
- XML. eXtensible Markup Language.

How to use this manual

This manual aims to helping the user of Converter, one of the SciELO methodology programs.

This manual is structured in topics as follows:

- **Introduction:** presents a brief explanation on the converter program, and its functionalities.
- **Directory structure:** - describes the structure in which the file and directory should be arranged so the program converter can operate successfully.
- **Requirements to start conversion** – demonstrates the procedures executed before using the converter
- **Using the converter program:** - explains in detail with illustrations the procedures to use the converter.
- **Main error causes:** reports the most frequent errors that disable the journal's files conversion in the program.

1 Preface

1.1 About BIREME

Year after year, BIREME has been following its mission of being a center dedicated to scientific and technical health information for the region of Latin America and the Caribbean. Founded in Brazil in 1967, under the name of Regional Medicine Library (which the acronym BIREME comes from), it has always met the growing demand for up-to-date scientific literature from the Brazilian health systems and the communities of healthcare researchers, professionals and students. Then, in 1982, its name changed to Latin-American and Caribbean Center on Health Sciences Information so as to better express its dedication to the strengthening and expansion of the flow of scientific and technical health information across the region, but kept the acronym.

Networking, based on decentralization, on the development of local capacities, on sharing information resources, on developing cooperative products and services, on designing common methodologies, has always been the foundation of BIREME's technical cooperation work. It has been like this that the center established itself as an international model that fosters professional education

with managerial and technical information with the adoption of information and communication paradigms that best meet local needs.

The main foundations that gave origin and which support the existence of BIREME are following:

- ✓ access to scientific and technical health information is essential for the development of health;
- ✓ the need to develop the capacity of Latin American and Caribbean countries to operate their sources of scientific-technical health information in a cooperative and efficient manner;
- ✓ the need to foster the use and to respond to the demands for scientific-technical health information from governments, health systems, educational and research institutions.

BIREME, as a specialized center of the Pan-American Health Organization (PAHO)/ World Health Organization (WHO), coordinates and conducts technical cooperation activities on the management of scientific information and knowledge with the aim of strengthening and expanding the flow of scientific health information in Brazil and in other Latin American and Caribbean countries as a key condition for the development of health, including its planning, management, promotion, research, education, and care.

The agreement that supports BIREME is renewed every five years by the members of the National Advisory Committee of the institution (PAHO, Brazilian Ministry of Health, Brazilian Ministry of Education and Culture, Secretary of Health of the State of São Paulo, and Federal University of São Paulo – Unifesp). The latter provides the physical infrastructure necessary for the establishment of the institution.

In 2004 the institution took on the responsibility of becoming a knowledge-based institution.

1.2 The Virtual Health Library (VHL)

With the rise and consolidation of the internet as the prevailing means of access to information and communication, BIREME's technical cooperation model evolved, as of 1998, to build and develop the Virtual Health Library (VHL) as a common space for the convergence of the cooperative work of producers, intermediaries, and users of information. The VHL promotes the development of a network of sources of scientific and technical information with universal access on the internet. For the first time there has been a real possibility of equal access to health information.

To BIREME, the Virtual Health Library is a model for the management of information and knowledge, which includes the cooperation and convergence between institutions, systems, networks, and initiatives of producers, intermediaries, and users in the operation of networks of local, national, regional and international information sources favoring open and universal access.

Today, every country in Latin America and the Caribbean (Region) participates either directly or indirectly in the cooperative products and services offered by the VHL, which includes over 1,000 institutions in more than 30 countries.

The VHL is simulated in a virtual space of the internet formed by a collection or network of health information sources in the Region. Users of different levels and locations can interact and navigate in the space of one or many information sources, regardless of where they are. Information sources are generated, updated, stored and operated on the internet by producers, integrators, and intermediaries, in a decentralized manner, following common methodologies for their integration in the VHL.

The VHL organizes information in a structure that integrates and interconnects reference databases, specialist directories, events and institutions, a catalogue of the information resources available on the internet, collections of full texts with a highlight for the SciELO (*Scientific Electronic Library Online*) collection of scientific journals, selective information dissemination services, information sources to support education and decision-making, news, discussion lists, and support to virtual communities. The space of the VHL is, therefore, a dynamic and

decentralized network of information sources based on which it is possible to retrieve and extract information and knowledge to support health decision-making processes.

The Virtual Health Library can be visualized as a distributed base of scientific and technical health knowledge that is saved, organized and stored in electronic format in the countries of the Region, universally accessible on the internet and compatible with international databases.

1.3 About the SciELO Methodology

The access to adequate and up-to-date scientific and technical information is essential for the economic and social development, specially to support decision making process in planning, formulation and implementation of public policies and to support professional development and practice. The results of scientific research are mainly communicated and validated through publication in scientific journals. This is valid for developed and developing countries. However, scientific journals from developing countries face several distribution and dissemination barriers, which limits the access and usage of locally generated scientific information.

SciELO - Scientific Electronic Library Online is a model for cooperative electronic publishing of scientific journals on the Internet. Especially conceived to meet the scientific communication needs of developing countries, particularly Latin America and the Caribbean countries, it provides an efficient way to assure universal visibility and accessibility to their scientific literature, contributing to overcome the phenomena known as "lost science". In addition, the SciELO model comprises integrated procedures for the measurement of usage and impact of scientific journals.

SciELO Model is product of a partnership among FAPESP (the State of São Paulo Science Foundation) <<http://www.fapesp.br>>, BIREME (the Latin America and Caribbean Center on Health Sciences Information) <<http://www.bireme.br>>, as well as national and international institutions related to scientific communication

and editors. A pilot project, involving 10 Brazilian journals from different subject areas, was successfully carried out from March 1997 to May 1998, aimed at the development and evaluation of an adequate methodology for electronic publishing on the Internet. From June 1998, the project begins to operate regularly, incorporating progressively new journal titles and expanding its operation to other countries. Since 2002, the Project is also supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) <<http://www.cnpq.br>>.

The SciELO Model comprises three components:

The model's first component is the SciELO Methodology, which enables the electronic publication of complete editions of scientific journals, the organization of searchable bibliographical and full text databases, the preservation of electronic archives and the production of statistical indicators of the scientific literature usage and impact. The methodology includes also journal evaluation criteria based on international scientific communication standards. SciELO full texts are enriched with dynamic hypertext links with national and international data bases, as for example, LILACS and MEDLINE.

The SciELO Model's second component is the application of the SciELO Methodology to operate web sites of collections of electronic journals. The SciELO Model envisages the operation of national sites as well as thematic sites. The pioneer application is the SciELO Brazil site <<http://www.scielo.br>>. Nowadays, Chile <<http://www.scielo.cl>> and Cuba <<http://www.scielo.sld.cu>> are also operating applications regularly. Several other countries are evaluating and/or being trained on the SciELO Methodology. SciELO Public Health <<http://www.scielosp.org>>, a regional thematic library covering Public Health scientific journals from Latin America and Spain, was launched in December 1999. A portal to integrate and provide access to the network of SciELO sites operates at <<http://www.scielo.org>>.

The Model's third component is the actual development of partnerships among national and international scientific communication players — authors, editors, scientific and technological institutions, funding agencies, universities, libraries, scientific and technological information centers etc, aiming at the dissemination, improvement and sustainability of the SciELO Model. The operation of the SciELO

network is highly based on national infrastructures, which contributes to guarantee its future sustainability.

The successful development of the proposed SciELO network of Latin America and Caribbean scientific journals in the next years will contribute to make locally generated scientific information readily available, which will ultimately contribute to increase the usage of scientific and technical information on decision making process at different levels.

2 Introduction

The Module Converter is responsible for loading the database using complete text documents previously marked according to DTD SciELO 3.1. Therefore, it should be used after Markup and Parser modules.

Its functions are as follows:

- To read HTML files, already marked and processed by Markup Module;
- To identify the markups and associate them with SciELO database fields;
- To load the database so that any article text has the register structure to be available in retrieval interface.

This module uses as input files:

- SGML declaration;
- DTD SciELO - Version 3.1;
- the DTD element conversion table for database fields;
- the files that contain the articles marked according to DTD SciELO 3.1.

Apart from the article database, the program generates an inverted file of the base and a log file, that enables configuring the tracks of the files used and created.

3 Directory Structure

In order to load the database, it is necessary to have a uniform structure of directories to organize the documents.

The structure was defined by the Methodology and it is not possible to change it. There are rules to name the directories.

```
\scielo\serial\acronym\v*n*\  body
                               img
                               markup
                               pdf
                               source
```

The directories in this structure may contain the following files:

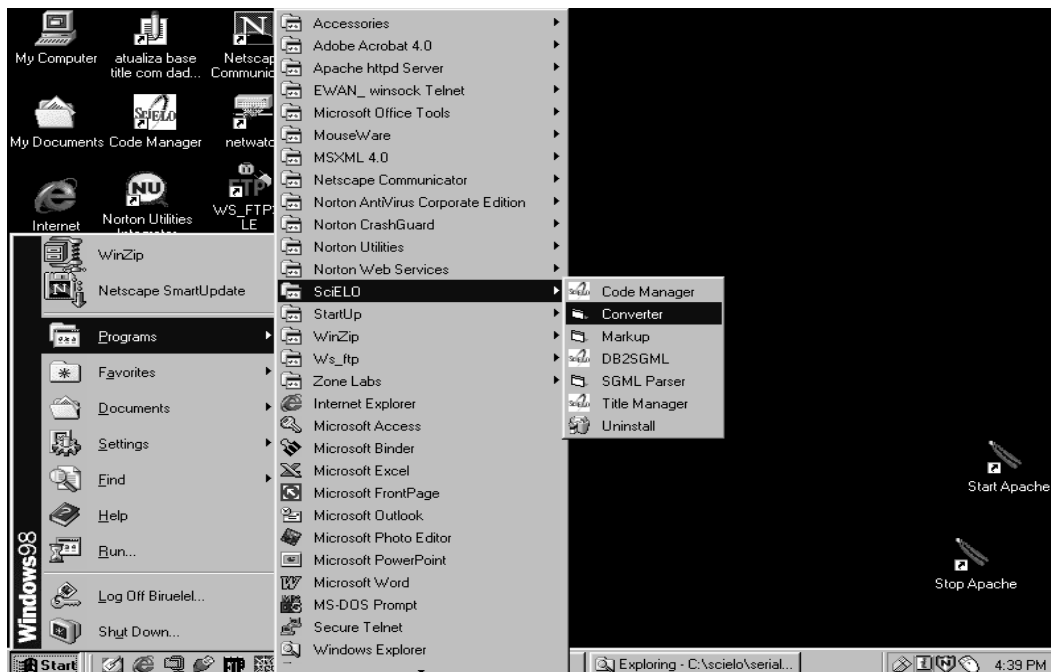
- body – HTML files without markup
- markup – HTML files with markup
- img – image files
- pdf – files in PDF format (optional)
- source – original files of the articles

4 Requirements to initiate conversion

- Check if all the files are in the correct directories. It is necessary to have HTML files in Body and Markup directories.
- The HTML files in Markup directory should be correctly marked.
- The number of files should be the same in the same in the directories (body and markup) and in Title Manager.

5 Using Converter program

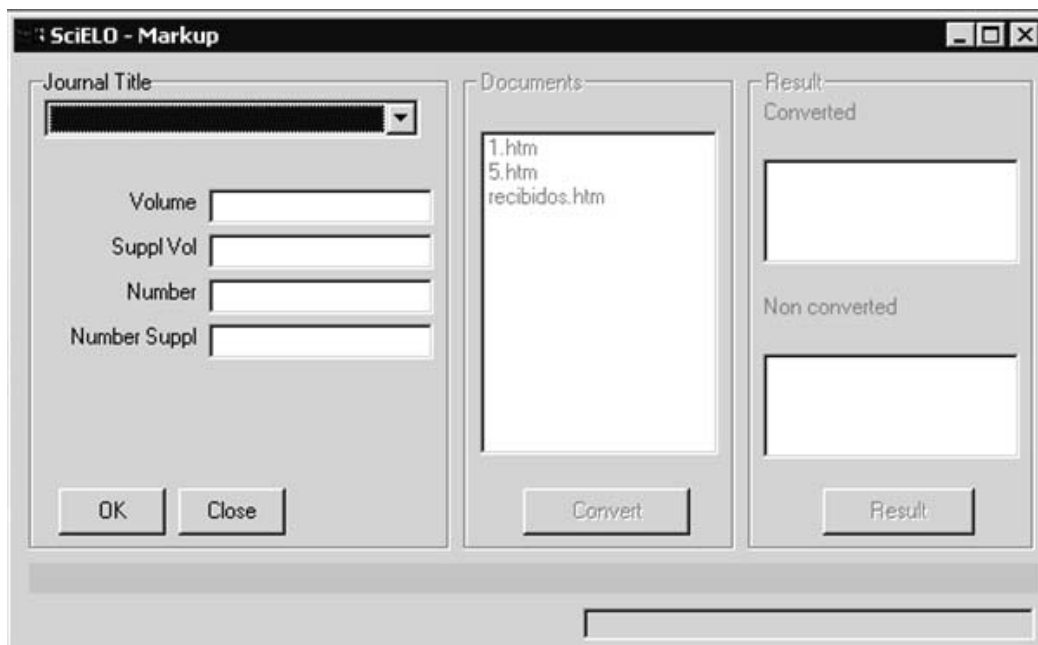
In Windows, use menu Start, Programs, SciELO, and open the executable program converter. After conclusion, the following screen will be displayed:



Select option Open and a screen will be shown to fill out the form.



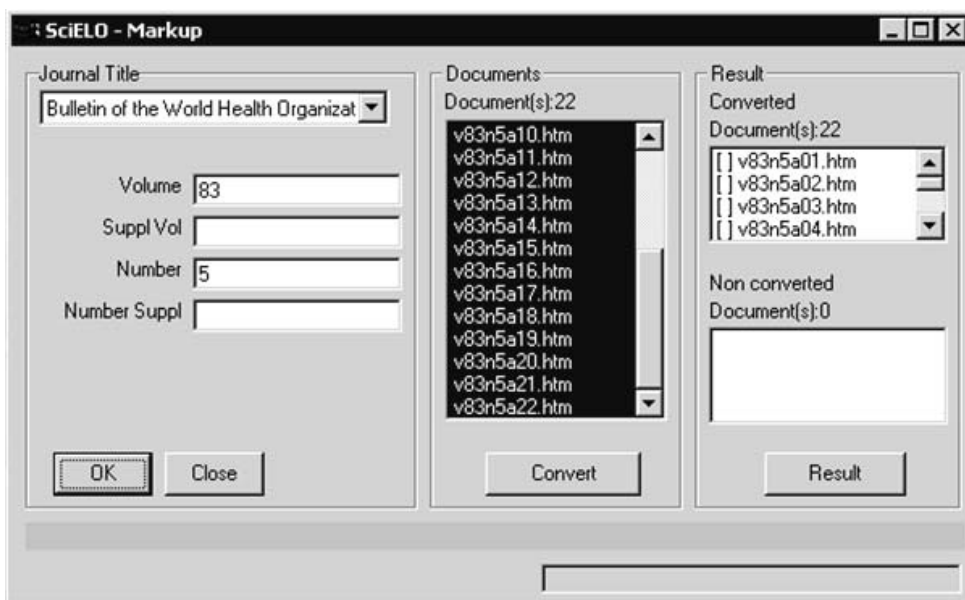
Fill out the form that indicates the journal title, volume, volume supplement, number and number supplement.



After completing the form with data identifying the issue, click on Accept. All files already selected for conversion will be displayed in a box on the side. If you need to convert only some files, select them.

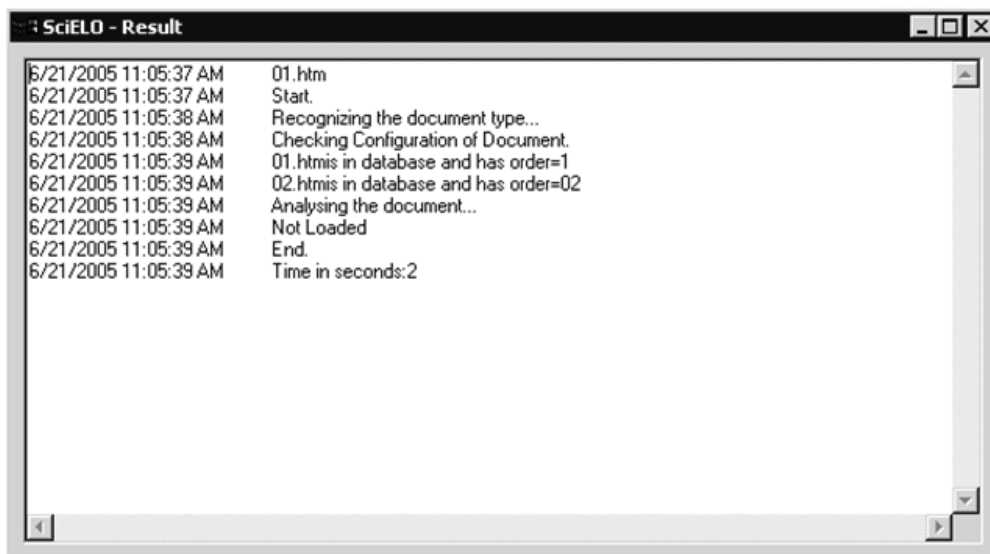
After choosing the files, click on Convert and the program will start converting the files to the database. During file conversion, some processing messages are shown on the screen.

Conversion result is shown in box Results, and the articles with no severe markup error are displayed in the box Converted. The files with some markup or configuration problems will not be converted and are displayed in box Not converted.



Some articles are presented in the Converted box with an [X] before its name. It means they may have some markup errors, but not severe enough to avoid conversion.

It is possible to follow the loading process of each file in two lists (converted and not converted), selecting one file at a turn and clicking on Result.



```
SciELO - Result
6/21/2005 11:05:37 AM 01.htm
6/21/2005 11:05:37 AM Start.
6/21/2005 11:05:38 AM Recognizing the document type...
6/21/2005 11:05:38 AM Checking Configuration of Document.
6/21/2005 11:05:39 AM 01.htm is in database and has order=1
6/21/2005 11:05:39 AM 02.htm is in database and has order=02
6/21/2005 11:05:39 AM Analysing the document...
6/21/2005 11:05:39 AM Not Loaded
6/21/2005 11:05:39 AM End.
6/21/2005 11:05:39 AM Time in seconds:2
```

The results of this process is stored in the directory Base, automatically created by the program.

It is strongly recommended to use this program only in the computer where the local site is installed (<http://localhost>).

6 Main problems

- Files not marked;
- Different files with the same order number;
- Total number of files is not the same specified in Title Manager;
- Information specified in Title Manager, such as ISSN, shortened title, volume and number do not comply with information of marked file.
- HTML code inserted in the DTD tag: [`titlegrp`]

In order to solve these problems, one should make the necessary corrections in the marked files and/or Title Manager and convert the files again.

7 Bibliographic references

1. PACKER, Abel Laerte. SciELO: metodología para la preparación, almacenamiento, diseminación y evaluación de revistas científicas electrónicas. In: *Congreso Regional de Información en Ciencias de la Salud, 4* [online]. San José, 1998. Available from internet: <<http://www.bireme.br/cgi-bin/crics4w/text0?id=crics4-mr3-co3>>.
2. PACKER, Abel Laerte; et al. SciELO: uma metodologia para publicação eletrônica. *Ciência da Informação* [online]. 1998, v. 27, n. 2 [cited 2005 Set 21]. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000200002&tlng=en&lmg=en&nrm=iso. ISSN 0100-1965.

8 Glossary

- **Affiliation.** Institution to which the author belongs or to which he/she is subordinated.
- **Analytical.** Part of a document, such as the article of a periodical or the chapter of a book.
- **Application.** Program used to execute tasks in connection with an application, such as the creation or edition of texts, drawings, animations, layout, etc. E.g.: text processor, database manager, Internet browser, etc.
- **Backup.** Procedure used to duplicate one or more files and/or directories in another storing device (tape or disc), thus producing a backup copy that may be restored in the event of accidental deletion or physical damage to the original data.
- **Bibliographic Database.** Electronic version of a catalog or bibliographic index.

- **Bibliographic Description.** Description of a bibliographic item by using attributes such as author, title, edition, size, etc.
- **Browser.** Internet page navigator, such as Internet Explorer and Netscape Navigator.
- **CDS/ISIS - MicroISIS.** Software programs developed and maintained by UNESCO to treat bibliographic data.
- **CGI.** The Common Gateway Interface is a standard for interfacing external applications with information servers, such as HTTP or Web servers.
- **Controlled or structured vocabulary.** Collection of related terms, organized according to a methodology, in order to facilitate the access to the information previously indexed with those terms.
- **Cooperating Center.** Institution that participates in the VHL and/or contributes bibliographic records to Bireme.
- **Database.** Collection of data that are structured to be easily accessed and handled. It is formed by units called records whose attributes are represented by fields. For example, in a file called "customer base", each customer is a record, with several fields such as "NAME", "CUSTOMER CODE", "TELEPHONE" etc.
- **DeCS Server.** Application developed by Bireme using the IsisScript language to manage the database of health descriptors (DeCS).
- **Descriptor.** Embodies a concept accepted in a controlled vocabulary (like a thesaurus.)
- **DTD SciELO.** Describes the article structure and other periodic texts of scientific, identifying and defining of necessary form its structure and the bibliographical elements constituent, the context

where they appear, its obligatoriness and its attributes. The DTD is used for the description and computerized treatment of texts.

- **Editorial Committee.** Group of professionals and specialists of the publication area of a periodical whose objective is to establish the rules and editorial conventions and to evaluate the contributions received by the publication to guarantee a certain quality standard.
- **Electronic Format.** Any form of storage, retrieval or presentation of information that may be transmitted on-line or recorded in magnetic or optical media.
- **Field.** *See* Database.
- **File.** In computing, a set of data that may be saved into some type of storing device. The data files are created by applications, such as a text processor for example.
- **Glossary.** Vocabulary for specific or controlled use, used in publications to clarify the meaning of technical or restricted terms which are not widely used.
- **Guide.** Defines the processes needed for the production of a source of information or phases of a methodology.
- **Indexing.** Procedure to identify and describe the content of a document with terms that reflect the corresponding subject matters to allow the document to be retrieved later.
- **ISO Format (of files).** Standard established by the ISO to allow the exchange of data between institutions, networks and users.

- **LILACS Format.** A bibliographic description format established by BIREME, based on the UNISIST Reference Manual for Machine-readable Bibliographic Descriptions.
- **Manual.** Set of steps and operations, whether automatic or manual, required to provide users with instructions on a certain application, program or methodology.
- **Methodology.** Set of rules and conventions used to standardize a process or the production of a source of information.
- **National Coordinating Center.** Institution that cooperates with the VHL and whose primary role is the coordination of a region's cooperating centers.
- **PDF.** File format developed by Adobe whose objective is to maintain the presentation format of a document designed for printing when this document is stored in digital media.
- **PubMed.** Service of the National Library of Medicine that includes over 15 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.
- **Quotation.** Excerpt originally written by a third person. Quotations appear between inverted commas in a publication, with a mention to the author.
- **Scientific production.** Collation (gathering and analysis) of all the literature on a certain theme or literature produced by a specific author for the purpose of analysis, especially of a quantitative nature.
- **SGML.** Metalanguage standard of the ISO (International Organization for Standardization) used for the definition of languages of marking

of electronic texts, making possible the interchange and the distribution of documents in the most varied formats, from one same source of data.

- **Style.** Element which defines the form of a character, a set of characters or a paragraph for viewing or printing purposes. *See* template.
- **Stylesheet.** File which contains the definition of the styles of a publication. *See also* template.
- **Technical Cooperation.** Exchange between developing countries or between developing countries and developed countries to enable cooperation in certain areas, such as the exchange of specialists and faculty members, development or transfer of technology, exchange of information, exchange of information and experiences to improve sanitary conditions.
- **Template.** File which contains the basic definition of the type of document that will be used, with style, predefined text, etc.
- **Thematic area.** Specific set of information on the subject matter of a VHL which allows user topic-based navigation.
- **Treatment Level.** Codified definition of the degree of depth applied to the document upon its bibliographic description.
- **URL.** Standard defined for the addressing of data contents via the TCP/IP protocol. Internet browsers use the URL to access Web pages.
- **Vancouver Group.** Was created in 1978 for the elaboration of articles, including the norms for the bibliographical references and was supported of the National Library.

- XML. Language created to allow the arrangement of data in a structured and hierarchical manner, thus facilitating data communication between different systems and platforms.