Model of Metadata (Crosswalk) on the Iberoamerican and Caribbean Virtual Library

Javier Solorio Lagunas
CGSTI, Proyectos Especiales
Universidad de Colima, Mexico
Tel: 01 (52) 312 3161039
Fax: 01 (52) 312 3161039
solorioj@ucol.mx

Lourdes Feria Basurto
CGSTI
Universidad de Colima, Mexico
Tel: 01 (52) 3123161039
Fax: 01 (52) 312 3161039
lferia@ucol.mx

Abstract:
The Iberoamerican and Caribbean Virtual Library (in spanish, BVIC)) is a coordinated project between the University of Colima and UNESCO, it was born in 1999 and has the goal of create a great digital library for the Iberoamerican and Caribbean region. It is a consortium open to national, university, public and institutional libraries, and is currently available the 3.0 version called “El Dorado”. It is an initiative of free software y free contents for the users. An important element of this virtual library model is the implementation of a methodology, structured in three parts: 1. information digitization, 2. client-server system Z39.50 and 3. the register guide of information resources, through the utilization of Dublin Core, GILS (Government Information Locator Service) and MARC21 metadata standards, for the objectives and thematic of this conference we’ll focus on the description of this model of metadata.

Keywords: Metadata, Dublin Core, GILS, MARC21, Virtual libraries, normalization and information retrieval.

Introduction:
The Iberoamerican and Caribbean Digital Library project was born in 1988, in the framework of the INFOLAC program, and along with other four projects: 1. ANTARES, 2. UNESCO Regional Chair on New Information Technologies for the MERCOSUR, 4. System of cooperative cataloguing on the Internet for Latin America and the Caribbean were introduced to the Bureau of the Intergovernmental Informatics Program Council of UNESCO.

Of these five projects, the first three were selected by the said intergovernmental organism to be financed in a shared manner, in charge of a fiduciary fund granted by the Spanish Government. The authorization of these funds is a responsibility of the Superior Informatics Council of Spain, annex to the Public Administrations Ministry (in spanish, MAP). Within these three projects was born the creation of the Iberoamerican and Caribbean Digital Library, as well as in 2001 the UNESCO Regional Chair on Information Technologies.

In Latin America and the Caribbean exist thousands of web pages and, on the other side, abundant and valuable rich information of diverse significance (historic, cultural, technological and scientific) that must be organized conveniently on the Internet, in order to be able to compete in a global scale and contribute to the regional development over its on cultural roots. Aware of that spoken truth, a number of national institutions have been united to the UNESCO’s Information Society Division for Latin America and the Caribbean in the Office of the INF/LAC/UNESCO-Caracas Adviser (currently in UNESCO-Quito) to execute dozens of automation projects, a good deal of them focused to libraries and documentation spaces, within the INFOLAC program.

UNESCO called a group of experts from the Brazilian Information in Science and Tech Institute (IBICT), the Information in Medical Sciences...
Regional Centre (BIREME, Brazil), the National Medical Sciences Centre of Cuba (INFOMED), the Monterrey’s Superior Institute of Superior Studies and the University of Colima, in Mexico, for the creation of the Library model as well as a methodology that will allow its implementation in any kind of library.

Methodology

For the development of the general Digital Library Methodology, the group previously mentioned analyzed the standards referring to the cataloguing of registrations with metadata, like Dublin Core, MARC21 and GILS (Government Information Locator Services). Standards for the digitization of information (texts, images, audio, video, etc.), and for the interchange of information protocols between client-server systems. The methodology as well as the development of software progressed successfully and, after a year of work, ended in November 2002 with the 1.0 version.

The methodology includes three sections, the first one directed to the management of metadata, through a guide for the register of information resources, with MARC21 and cross references to Dublin Core and GILS, the second one dedicated to the digitization of information and the third one to the exchange of information between servers with the implementation of the Client-Server System Z39.50.

Metadata structure

The 1.0 version of this document was created according to the bases of the Users Guide for the Register of Resources on Information of Health Locator (LIS), produced by BIREME and adapted with the authorization of the intellectual author and the Health Virtual Library of Latin America and the Caribbean (BVS). Such metadata are an essential component for the information retrieval of the Iberoamerican and Caribbean Virtual Library (BVIC). The methodology as well as the development of software progressed successfully and, after a year of work, ended in November 2002 with the 1.0 version.

The methodology includes three sections, the first one directed to the management of metadata, through a guide for the register of information resources, with MARC21 and cross references to Dublin Core and GILS, the second one dedicated to the digitization of information and the third one to the exchange of information between servers with the implementation of the Client-Server System Z39.50.

1. - Institutions board

Where the fields related to the organizations whose data may appear in multiple register are consigned only once, the 2006 to 2015 and 2023 to 2033 of GILS.

2. Users board

Describe the elements that identify the person(s) responsible of the administration, assignation of metadata and approval of registers from the Information Locator. Also, it was considered a primary factor that is the practical aspect of index of those data by the participating institutions, as well as their use databases management system It’s a Client-Server implementation that has comprised different clients (programs/bookstores). It permits to add, and process engraved data in a database. At present the database agent plays a central role for data processing, as a tool, or as part of another application.

The metadata model is comprised of a data base in MySQL of 41 fields, structured in three boards, for the cataloguing, index, administration and retrieval of themselves. We will describe the boards according to the order in which they are used in the Virtual Library (BVIC).

MySQL database server

It is an SQL (Structured Query Language)
and maintenance. In addition, they maintain a relation of different fields like the Control Identifying with the Institution Code in the institution and users boards, and the originator and distributor with the name of the institution.

3. Main board

The main board consists of 20 fields where appear the data that has been brought up to date by each resource or digital object indexed in the data base.

We have built the relations of metadata elements or cross references between GILS, MARC21 and Dublin Core in the previous boards, that correspond to the interface of retrieval for the mid user of the virtual library (BVIC) system; where cataloguers, testers, administrative, who perform the activities of: cataloguing, modification, correction, elimination of metadata records, indexation and administration of the data base that is generated in the system, each one depending on the access level that he has, in this case, by the library administrator, who has all the rights, among them create, modifying, eliminate users (mid) and participating institutions.

Next, we will describe, first in a separate manner, each one of the used metadata standards, its elements or fields, and the ones which were implemented in the methodology of the library and subsequently the relations (crosswalk) among themselves.

Metadata description

GILS

The base of the metadata standard in the BVIC is the GILS system, which, as has been mentioned, is part of the information locator of the BVS, from it we took 28 fields or elements for its cross relation and correspondence with MARC21 and those which correspond with Dublin Core.

For the GILS Standard they are used 28 elements of this metadata system.

MARC21

In the case of MARC21 we used 27 elements or fields and subfields for the relations primarily with the GILS standard. In the next board it is shown the name of the field with its correspondent tag.
In the Dublin Core case and considering the library model that was desirable to implement, the expert group identified that DC, due to the amount of elements that contained, it should be used but with a crosswalk with other metadata systems. Our model utilizes 14 elements of DC:

1. Title
2. Subject
3. Description
4. Type
5. Coverage
6. Creator
7. Publisher
8. Contributor
9. Date
10. Format
11. Identifier
12. Language
13. Source
14. Relation

The element “Rights” it is not used in this model of virtual library, because its objective is to offer, in a free mode, digital objects to the users of the diverse servers and sites that participate with the BVIC. In some elements we have used the qualifications of DC, for example: thesaurus, abstract, creation date, resources type, temporal cover, etc.

**Crosswalk relations of metadata elements between the three standards**

In the next chart there are illustrated the relations of the elements (crosswalk) of the used standards.

1. In the Dublin Core case, in some elements we have used the DCQ (Dublin Core Qualifiers), as an example: in Description we used the relations with other standards in the field Abstract and Summary of GILS and MARC respectively. In the element Date we made the relations as well as in the date of publication as on the date of last modification. For the field Originator of GILS and Name of MARC we related the elements Creator, Publisher and Contributor of DC.

2. Considering an important element for the information retrieval, we have implemented in the interface of capture or metadata indexation in the field Thesaurus, the consult of different languages controlled of free access on the Internet; Macrothesaurus of OECD, UNESCO’s Thesaurus, DecS Thesaurus (Descriptors in Health Sciences), Informatics Thesaurus. As well, if a participating library doesn’t wish to use them, it will be able to include the name of the used thesaurus in the correspondent field or the list of subject headlines that the library utilizes to
assign themes or subjects to its documents.  

3. In our methodology it is descript in the Locator of Information Resources Guide the relations, cross references, between the elements or fields from the three standards, as well as a description of every field, its tags, etc.

**Information retrieval**

**Z39.50 Protocol**

Z39.50 is a protocol for information retrieval based on a client-server architecture, it facilitates the interconnection of information systems.

Its official name is “Information Retrieval (Z39-50): Application Services Definition and Protocol Specification. ANSI / NISO Z3950-1995” ó Z3950. The name is formed by two components: first corresponds to ANSI Committee number 39 and second to the standard 50 of the NISO.

The protocol’s architecture utilizes a “client-server” model, that allow the users to find information independently of syntax and semantics, because the client application converts the consultations to the formats indicated by the Z39.50.

The server transforms the data to his structure to consult the database. The answer is communicated in the same form inversely: the server converts the information to the form required by the protocol and the client transforms it to another form easier to be understood for the users.

For the information retrieval, metadata and other access points the Virtual Library uses the Z39.50 protocol. Through this protocol, the access is offered to the end users connected to servers of the participating libraries; it offers two options:

1. **Local searches in Z server.** They are executed in the institution server that has installed the BVIC. In this kind of searches the user can perform searches a) Basics, b) By metadata; title, author, abstract, resource type, themes, etc., as wells as in words of the full text c) By themes catalogue, d) By title catalogue, e) By author catalogue, the last three showing the letters of the alphabet to select in which of them you wish to consult a theme, title or author.

2. **External searches** The Z39.50 protocol allow us the exchange of information between servers, in this model the user can execute searches of information in the diverse servers installed in the participating libraries, in this option we offer the following kinds of searches: a) by libraries, b) By libraries of a country or specific library in that nation, or well, search in every server of the existing libraries.

It is important to mention than the cataloguing of metadata as well as the full text of the digital objects are located in each of the participating servers, the same as the ones that join the BVIC from new institutions. Therefore, it is a distributed system, for the cataloguing, the disposition of digital objects in the net and thru the Z39.50 the metadata is shared.
In the annex I. It is a graphic representation as is agreed to the web way databases (server Apache or IIS), consultation in the database MySQL and the answer of the server.


We can conclude that a system or model of Digital Library it is not finished as long as there are changes to do, modifications, improvements, etc., the main object is to utilize the international standards to the information registration, as well as for the exchange of it. The goal of our Virtual Library project is to offer to every kind libraries a free access software, with a methodology that facilitate them the creation of its own digital or virtual library able to offer its users the contents that have been developed by the institution.

Though is certain that the metadata standards (whoever used them, whichever is used) allow the normalization of the information, is final goal is the information retrieval by the users.

At present the Virtual Library server is operating in Spanish version; next August an English version will be offered thanks to the collaboration of the National Library of Trinidad and Tobago. In a near future it will also be available a Portuguese version, with the participation of the Federal University of Pernambuco, Brazil.

References

5. Departamento de Ciencias de la Computación, Universidad de Chile. Calificadores del estandar de metadatos Dublin Core (Dublín Core Qualifiers)[En línea]. Disponible en Internet en: http://www.ciw.cl/proyectos/calificadoresDC.html [con acceso el 1-4-2005]