Cataloging conventional and digital objects: new tools with old names or old names to new tools?

Michalis Gerolimos
Department of Archival and Library Sciences, Ionian University, Greece
(mike@ionio.gr)

George Papadourakis
Department of Applied Informatics and Multimedia, Technological Educational Institute of Crete, Greece (papadour@cs.teicrete.gr)

Michail Nikitakis
Library, Technological Educational Institute of Crete, Greece (nikit@lib.teiher.gr)

Anestis Sitas
Library, Aristotle University of Thessaloniki, Greece (sitas@lit.auth.gr)

Abstract

Information concerning the creation of tools and bibliographic description standards of library material are referred. Detailed reference to their development and adjustment to the data created so as to reflect the new conditions and needs, the way these are formed in the contemporary environment, is made. We present the efforts made by the bibliographic agencies to develop new and to improve the traditional/conventional tools for the description of the conventional documents as well as the electronic resources available in the World Wide Web.

1. Towards a digital information environment

The current information environment, always in mutation, is in the process of change and readjustment of its constituent parts, mainly, due to the technological progress in the area of Communication Technologies. These technologies diversify it, by prompting it towards the incorporation of certain technological elements into the new work models and of the provision of services. In general terms, however, the present environment keeps the basic structure and the morphology of the conventional/traditional environment. Due to the established practices, as well as to the experience vested on fundamental matters of its administration, the current "hybrid environment" administers and develops models, practices and processes of two different but interrelated "worlds", the conventional and the digital world.

The main characteristic of the current environment of the libraries and the information agencies concern the form of the objects that compose their collections. Libraries, in addition to the conventional material, contain also new forms of documents, can use automated tools for metadata production – at least as far as the digital documents are concerned- aim at the fulfillment of the process of documents description as soon as

possible and finally, libraries offer their services outside the conventional work environment.

2. Bibliographic description

The processing of a library's documents through a series of procedures and the use of specific standards and tools so as for the information contained to become accessible to the users, is a relatively "recent" development in the area of libraries. The systematic processing of the material, aiming at the creation of access tools to the library material for the user, was introduced in the second half of the 19th century along with the introduction of the first rules concerning the description of the bibliographic material. The bibliographic description of a document – of a publication - aims at its description as a bibliographic unit, the recording that is to say, all of the elements that attribute a unique identity to each document and to determine the access points, which enable its location. Moreover, the documents are given the possibility of topical access, since it is attributed the right term during its subject description. Although being different for each document, these elements altogether make up the "bibliographic record" which is added to the library catalog.

3. Bibliographic catalogs

The main tool and the main service with which the library material becomes accessible to its users is its catalog, which is the most important searching tool that allows the user to search for and retrieve its documents. The general process of the creation of catalogs followed in the course of time, is the following: Book catalog – Card catalog – Online Public Access Catalog (OPAC).

The Book Catalog is the first attempt to create a tool for the recording of a collection's material. It prevailed up until the 19th century, when it was replaced by the Card catalog. The Card catalog constitutes the form of a library catalog that has dominated for almost a century, not only as a means of recording the documents, but also of the public's access to them. In the early 80's, due to the possibilities of computerization of the collections, the Card catalog began to be replaced by OPACs. The OPAC constitutes the tool to which the bibliographic data of a document is inserted in an electronic format in order for the user to be able to look it up in a computer, without restrictions of place or time.

4. Conventional/Traditional cataloging tools

4. 1. Cataloging Codes and ISBDs

The bibliographic description, in other words the input in a catalog of the identification elements of a document, is done according to standard codes and instructions, as they are determined by special tools and standards. For library cataloging, tools such as the AACR2, LCRI, MARCs, LCSH, LCC, DDC, etc., are used. The main tools of bibliographic description which are applied for the description of documents, or objects, are the cataloging codes/rules. Among the first cataloging codes, the most

important was the *Panizzi's British Museum Rules for the Compiling of the Catalogue* (1841) and *Cutter's Rules for a printed dictionary catalogue* (1876). These rules had the greatest effect on the later ones, such as the Anglo-American Cataloging Rules (ACCR).

The AACR are detailed standard rules established in 1967 in order to catalog documents of the American, Canadian and British libraries and have been based on Paris Principles (1961). Very quickly, they took the form of "universal" rules, since they were adopted by many other countries. In 1978, they were published in a volume, followed by their revisions of 1998, in order to reflect the changes in formats, to which the information is now available. Their current publication is *AACR2 Second ed., 2002 Revision* (with 2003, 2004, and 2005 updates) (AACR JCS, 2005). Rules concerning the Internet resources and other types of resources are included, without however constituting an entirely new code. The next edition, entitled "RDA - Resource Description and Access", is expected to be published in 2009. (AACR JCS, 2005)

The need to standardize the cataloging processing on an international scale has leaded the IFLA (International Federation of Library Associations and Institutions) to the creation of a series of standards for bibliographic descriptions, known as International Standard Bibliographic Descriptions (ISBDs). The ISBD is a group of standards whose creation started in 1971, having as an objective the preparation of the descriptive part of bibliographic records. The ISBD determines the structure that bibliographic records must have, the order of the elements of description and the punctuation system which offer the possibility to exchange bibliographic documents and to convert them into a format that is readable by a computer. (AACR2 Homepage, 2006) The incorporation of the ISBDs into the AACR2 structure has constituted one of the greatest developments in cataloging.

The library collections, therefore the tools for bibliographic description, are determined by the format of the material included. While, initially, the collections were made of paper material, in the course of time new types of documents appear which lead to the creation of new tools, in order to make their library processing possible. The IFLA, driven by the increasing need for a separate ISBD concerning especially the computer files, published in 1990 the ISBD (CF- **C**omputer **F**iles). The electronic resources, however, are the products of a technology that is developing quite quickly. As a result of this assessment, IFLA decided to reassess and revise the ISBD (CF), which leaded in 1997 to the publication of ISBD (ER- **E**lectronic **R**esources). (IFLANET, 2005)

Subject access tools

The assignment of topical terms to the documents constitutes an effort of encoding their contents in order to make the search easier to the user. For their assignment, various tools have been created, which aim at the compilation of controlled vocabularies of topical terms assignment. The subject headings as well as the thesauruses are such tools. There are various tools for subject description of the documents available; among the most frequent are the Library of Congress Subject Headings (LCSH), the MeSH (Medical Subject Headings) and the Eurovoc.

Although the creation and use of the LCSH started in 1897, their basic 9th printed edition was published in 1980; now the 29th edition is available. Apart from the printed format, the LCSH have been available in microform (microfilm and microfiche), from

1976 to 2005 and in CD-ROM from 1988 to2005 as well. Since 1995, they are available via Internet and via the *Classification Web*.[1]

The MeSH are the subject headings created to be used by the National Library of Medicine (USA). The first printed edition was published in 1954 while the access to the online version is already available for free from the *MeSH Website*.[2] Finally, the EUROVOC is a multilingual thesaurus that covers all the European Union activity areas. It was created in 1982 and was first published in 1984. Its current edition, 4.2, is available in the 16 official languages of the E. U. and the access to the *Eurovoc thesaurus* webpage[3] is free.

Several classification schemes add to the task of the subject description of the documents, such as the DDC, LCC, UDC etc. Among the most frequent used classification schemes is the Dewey Decimal Classification (DDC), whose development started in 1876 and the Library of Congress Classification (LCC), which was developed mainly between 1899-1940.

5. Description tools in automated environments: MARC

The software that automatizes the works in the libraries need to implement a standard for the automate data readout. The MARC (Machine Readable Cataloging) format gives this possibility. Its creation started from the Library of Congress (USA) in the 60's with the creation of the LC-MARC format. This format, together with the identical CAN-MARC, has harmonized their differences, resulting, in 1999, in the publication of a volume entitled MARC21. In the meanwhile, other bibliographic agencies have been developed on over 20 different MARC formats, like the UKMARC, etc. In 1977, the IFLA tried to solve the problems caused by the number of deferent formats, their exchange and their compatibility, with the publication of the UNIMARC. In 1994 a newer and more complete edition of the UNIMARC was published.

Since 1981 the *ISO 2709* has been adopted, which has been replaced by ISO 2709:1996 "Information and documentation - Format for Information Exchange". It concerns an international exchange scheme and has actually originated from the original LC-MARC format. Its aim is to provide an internationally accepted standard, to which all the existing formats can be adapted.

One of the most significant and indicative adaptation procedures of MARC to the new means of publication of the documents, started in 1993 and concerns the creation of a special field containing necessary information for the location and the access to electronic resources, with the use of the 856 field "Electronic Location and Access" in the USMARC formats. Right after, this field has been adapted by UNIMARC.

6. Means of tools publication

The conventional media of publications of all the tools is the printed edition. With the advance of the technology, these tools began to be available in other formats as well. Initially, they were transformed into microforms, but mainly into CD-ROMS, which gave

[1] LC, Classification Web, http://classificationweb.net/
[2] NLM, MeSH Browser, http://www.nlm.nih.gov/mesh/MBrowser.html
[3] Eurovoc thesaurus, http://europa.eu.int/celex/eurovoc/

the possibility of their setting up in local networks. The best known relative service is the **Cataloger's Desktop**, which, besides the cataloging rules, provides access to the most significant tools for its application. However, the format that gradually dominated is the access through the Web. The Library of Congress has stopped its publication in CD-ROM with issue 4 (2005) and the access is possible only through the web version[4], and also through the **Classification Web**, which provides access nearly to all services and tools needed (classification schemes, subject headings, authority files, etc). This kind of access offers more developed many more possibilities, for example the hypertext linking between the AACR2 and the Library of Congress Rule Interpretations (LCRIs), etc.

Nevertheless, in the new environment where the Internet is prevalent, there are many tools that, not only do they change the medium of availability, but can partially be replaced without cost, since the access to these tools is free and they can be used as alternative sources of information retrieval for the catalogers. Certain free access documents are identical with the commercially available ones. There are others that provide an acceptable substitute, like the *Amazon.com,* which, after its launching in 1995, can partially function as an alternative solution for the *Books in Print*.[5] (Arms, 2000). There are also other free sources offering help to catalogers without demanding subscription, such as the Online LC Catalog, the Library of Congress Authorities, the MARC21 Concise formats and UNIMARC formats.

7. Search Engines vs Catalogs

The evolution of the Web has had an influence on the libraries' nature. The constant increase of digital objects (files, images, sound, computer software, multimedia, etc) available through the Web has led the libraries to the incorporation of the digital objects to their collections. It is already being discussed that the libraries must move their interest from the strict keeping of the library rules to standards that can be supported by automated systems for organization of conventional or digital libraries. This fact often obliges the libraries to transpose their efforts from the strict keeping of the description and cataloging standards for the creation of their catalogs towards the implementation of a more flexible policy to describe all the types of documents.

There are significant differences between the Web search engines and the catalogs, whether we are referring to the catalogs of conventional or digital libraries. Catalog is controlled for its quality and its bibliographic information is selected and is created on the basis of standard and generally accepted description rules. The use of the authority files, as well as, the assignment of the topical terms, the classification numbers, etc. lends quality to the information provision and the user's search. Also, the procedure for the choice of the documents that precedes the cataloging is based on detailed selection and evaluation of resources.

On the contrary, the choice made for the documents that are indexed by a Web search engine, is often based on arbitrary assessments, the recalled records can be characterized as unelaborated, the control of the authority files does not exist and the reduction of duplicated records if far from the desired goal. On the other hand, however, the Web search agencies are proven to be powerful where the catalogs are weak. Although the

---

[4] Cataloger's Desktop, http://www.loc.gov/cds/desktop/
[5] Books in Print Professional, http://www.booksinprint.com/bip/

cataloging is expensive, the Web indexing is not. The most considerable Web search engines index many millions of websites, much more that the overall number of MARC records that could ever be created. But the Web's great advantage is that everything one can find in its indexes is directly accessible. (Arms, 2000)

Nevertheless, it would be wrong to claim that the conventional catalogs are superior due to their quality control, as it is equally wrong to claim that the Web search services are superior due to the low cost, the wide range of resources provided and the constant update. Their value depends on the degree of the user's satisfaction. (Arms, 2000)

8. Incorporating digital objects in library catalogs

A considerable number of libraries have already incorporated many Internet information resources in their websites or in specialized portals. Some libraries have begun cataloging, with conventional means, the Internet resources that they consider useful for their users, incorporating in this way the bibliographic descriptions of these digital objects in their catalogs. Some other libraries register their bibliographic records in metadata repositories, so as for the records of their natural collections to be available in the same way and at the same time with the digital ones. (Grant Campbell and Fast, 2004) This leads to the need to have data bases that contain complex bibliographic descriptions based on different standards, accordingly to the type and the kind of each described object.

9. Metadata

For organizing traditional format documents, it has been sufficient to a certain degree, the use of tools used by professionals for the processing of the libraries' material for decades. However, data as the author, the title and the topical terms are proven to be insufficient for many types of material, mainly as far as the digital objects is concerned. (Hunt and Ethington, 1997) In the conventional libraries, the creation of bibliographic records was, to a certain degree, limited and controlled. In the Internet, the lack of organizational control of the available search tools and search engines contribute to the ineffectiveness of the discovery of its resources. It seems that the most reliable approach for the discovery of the Internet resources is the application of procedures used by librarians for the description of conventional documents: the use of metadata. Metadata is the data for the data. More specifically, metadata is structured information that surrogates the real described object.

9.1. Dublin Core Metadata Element Set

The need to have a common description standard for all the concerned communities has led to the creation of standards that combine, not only the requirements of most of the communities, but also the need of a standard which can be at the same time simple and applicable, without demanding high-level specialization. The descriptive metadata ranges from the very enriched and high value, for example the MARC format and the TEI headers, to the relatively simple one with 15 elements Dublin Core (DC). The creation of *DC Metadata Element Set* started in 1995 by OCLC in collaboration with other

organizations. Its aim is to describe and locate the Web digital objects in order to constitute a simple descriptive tool which will make the object description and recovery easy. It can be encoded in HTML or XML, formats that are more applicable than the MARC formats, which are limited in the libraries' communities. (Lee-Smeltzer, 2000)

9.2. MARC, XML and RDF implementations

The MARC formats constitute the foundations of library automation. The availability however of the XML (eXtensible Markup Language) and RDF (Resource Description Framework) language, seems to form a dynamic development that gives programming new possibilities. The XML was created for the handling of electronic publication; in this way the information of the structured documents can be used in the Web. It is a new language describing the Web resources, that is a useful component for the description of its field and leads to what we call Semantic Web. The RDF is a language for representing information about resources in the Web, like title, author, modification date of a site, and information concerning copyright issues of the content of that site. Is a data integration mechanism between applications and the Web, is the language in which Semantic Web meta-data statements are expressed. It provides the ability to transfer data between various formats and allows late binding of application schemas to that data. (Manola and Miller, 2004) The XML and the RDF are not metadata formats, but general data formats that can be used in any application. (Coyle, Feb. 2005)

The efforts of creating software for the transformation of MARC records into XML, have started since the late 90's. The Library of Congress Network Development and the MARC Standards Office develop the MARCXML, which allows the MARC records to be encoded in XML. In addition, the LC, since 1995, has created the DTDs that allowed the conversion of MARC data into SGML. With the development of technology, this SGML DTD turned into XML DTD, which has been available for trial use since May 1993. Additionally, in 2004, the IFLA, in an effort to adjust the tools that it develops, in collaboration with the Library of Portugal and the BookMARC, have created a prototype of UNIMARC manual in XML.

9.3. MODS & METS

The MODS[6] (Metadata Object Description Schema) is compatible with XML and XML MARC and its goal is to carry selected data from the existing MARC 21 records and to allow the creation of original resource description records. For the electronic resources, it is particularly useful in the cases where the records must be transferred with the use of the XML schema language. In 2002 it became publicly available for trial use. The newest version of the scheme is version 3.1 which was released in July 2005.

The METS[7] (Metadata Encoding & Transmission Standard) provides a XML documents format for the encoding of the metadata needed for the digital objects management of the libraries in a repository and for their exchange between different

---

[6] The Library of Congress, *MODS, Metadata Object Description Schema official website*, http://www.loc.gov/standards/mods/
[7] The Library of Congress, *METS, Metadata Encoding & Transmission Standard official website*, http://www.loc.gov/standards/mets/

repositories (or between the repositories and their users). It doesn't include descriptive metadata, for the output of which various standards can be used, such as usually the DC and the MODS.

9.4. TEI (Text Encoding Initiative)

The existence of the Web and the possibilities it gives for direct and dynamic publication, have led to the creation of a complex repositories of full documents. The bibliographic description no matter how well it represents the document, it remains a simple description, which keeps on not providing the essential point for the user: the direct access to the document. For this reason, the scientific community was led to the research to encode the full document, providing other dimension and other dynamics to the access, its use and its processing. One of its aims is to use the information given by the full document at the library catalogs. TEI[8] is an effort started in 1987 and has become a powerful means of research, of indexing and of information storage. It is compatible, initially with the SGML language and then with the XML. The TEI header and the MARC record can exist in a parallel way, since none of them replace the other and it is allowed to create a connection between the TEI document and the corresponding MARC record. (McCallum, 2004)

10. Automatic metadata creation

The majority of the libraries support warmly the idea that the libraries can create the metadata for the total amount of their material and to make them available in the Web, without having to sacrifice the necessary descriptive standards. The libraries hold a tradition in the creation of bibliographic catalogs, with the use of well-tried and effective tools. However, problems arise in the new environment, which tries to incorporate more and more types of documents.

The most serious problem concerning the location of the Web resources still is the lack of use of effective controlled vocabularies (like DDC, LCC, LCSH) for subject access. The automatic production of metadata needs to be investigated in depth, mainly due to the reduction of cost that its application brings about. Nevertheless, at least for now, we shouldn't overlook the fact that more complete results are attained only with the supplementary application of human and automatized methods. (Greenberg, Spurgin and Crystal, 2005) These descriptions can be generated by the producers of the documents, at the point of posting the resource on the Internet; they can be automatically generated by intelligent software agents or produced by the librarians. Afterwards, the librarians can correct or enrich the record. (Lee-Smeltzer, 2000)

11. Conclusions

Libraries and communities of information services can play a significant role, not only in the supply of bibliographic services, but also in the services that offer direct access to all the information contained in a digital object. It can work for the integration or the adaptation and the adjustment of the traditional tools in the bibliographic

---

[8] The Text Encoding Initiative, http://www.tei-c.org/

description area. It ought to research the development of automatized and the production of new tools, such as the FRBR (Functional Requirements for Bibliographic Records) recommended by the IFLA and is about a new view of the cataloging function in the new environment and the new needs.

In order for the libraries to handle effectively the Internet collections, it is necessary to put into practice simple organizational structures and to present them in such a way, that they can be useful and intelligible to the users of both conventional and digital libraries. Passing from the conventional to the digital information environment and covering the current hybrid environment, that is to say the environment where institutions, tools and practices coexist, we have to be able to turn to make productive the advantages and the possibilities that it offers. Keeping in mind that this has to remain stable, by adjusting the ones having this potential, by setting aside the old practices and the tools when the old ones are not in the position to satisfy the new needs, but at the same time by creating new tools when need arises.

Bibliographical references

[Last access to all documents: 3/05/2007]
1. AACR JCS, A brief history of AACR (2005) [available at: http://www.collectionscanada.ca/jsc/history.html]
2. AACR JCS, RDA: Resource Description and Access, (http://www.collectionscanada.ca/jsc/rda.html)
3. AACR2 Homepage, http://www.aacr2.org/access.html
4. Arms, William Y. (July/Aug. 2000), Automated digital libraries: how effectively can computers be used for the skilled tasks of professional librarianship?, *D-Lib Magazine*, 6 (7/8) [available at: http://dlib.org/dlib/july00/arms/07arms.html]
5. BookMARC, *The UNIMARC and MARC21 manual in XML: automating validation, explanation and help systems for bibliographic records*, [available at: http://www.bookmarc.pt/unimarc/]
6. Carvalho, Joaquim (2005). An XML representation of the UNIMARC Manual: a working prototype, *71st IFLA General Conference*, [available at: http://www.ifla.org/IV/ifla71/papers/199e-Carvalho.pdf]
7. Coyle, Karen (Feb. 2005). Understanding metadata and its purpose, *The Journal of Academic Librarianship*, 31 (2): 160-163
8. Coyle, Karen (Jan 2005). Catalogs, card and other anachronisms, *The Journal of Academic Librarianship*, 31 (1): 60-92
9. Grant Campbell, D. and Karl V. Fast (2004), Academic libraries and the Semantic Web: what the future may hold for research-supporting library catalogues, *The Journal of Academic Librarianship*, 30 (5): 382–390
10. Greenberg, Jane, Kristina Spurgin and Abe Crystal (2005) *AMeGA Final report*, http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf
11. Hawkins, Les (Oct.-Dec. 2000), Refinement of cataloging tools, *Serial Review*, 26 (4): 37-41
12. Hawkins, Les (Spring 2003), Program for cooperative cataloging tools related to the 2002 Revision of AACR2, *Serial Review*, 29 (1): 40-42

13. Hunt, Leta (Li) and Philip J. Ethington (Nov. 1997), The utility of spatial and temporal organization in digital library construction, *The Journal of Academic Librarianship*, 23: 475-483

14. IFLA, *Functional Requirements for Bibliographic Records: final report* (1998). München: IFLA Section on Cataloguing, K. G. Saur [available at: http://www.ifla.org/VII/s13/frbr/frbr.pdf]

15. IFLANET, *Family of ISBDs* (2005) [available at: www.ifla.org/VI/3/nd1/isbdlist.htm]

16. Lee-Smeltzer, Kuang-Hwei (Janet) (2000), Finding the needle: controlled vocabularies, resource discovery, and Dublin Core, *Library Collections, Acquisitions, and Technical Services*, 24 (2): 205-215

17. Manola, Frank and Eric Miller (eds) (2004) *RDF Primer*, [available at: http://www.w3.org/TR/rdf-primer/]

18. McCallum, Sally (2004). *Recent Developments in MARC with XML* [available at: http://www.loc.gov/acq/conser/MARC-overview.html]

19. OCLC Research Activities, *IFLA's Functional Requirements for Bibliographic Records*, [available at: http://www.oclc.org/research/projects/frbr/default.htm]

20. Taylor, Arlene G. (2004). *Wynar's introduction to cataloging and classification*, 9th rev. ed. Westport, Conn.: Libraries Unlimited, 2004

21. The Library of Congress, *MARC standards*, http://www.loc.gov/marc/

22. The Library of Congress, *MARC XML, MARC21 XML Schema official website*, http://www.loc.gov/standards/marcxml/