16

International standards for global information

Alan Hopkinson
Middlesex University. London, UK

INTRODUCTION

In some contexts, the phrase *international standards* is used to refer only to those standards which have been endorsed by the International Organization for Standardization (ISO) which are sometimes referred to as *official* or *de jure* standards. Here the phrase will be used to refer to any convention which can be called a standard and which has international acceptance. The MARC standards which underpin many of the library world's cooperative activities are not formal international standards though the record structure on which they are based, that described in International Standard ISO 2709, is. The Dewey Decimal Classification scheme is not, nor is the Universal Decimal Classification, though by a quirk of fate it is actually a British Standard. The Anglo-American Cataloguing Rules are not nor are the International Standard Bibliographic Descriptions though some countries have adopted them as national standards. Nevertheless we will be discussing these in this chapter. It is also worth noting that in the time of the Cold War, library professionals in Eastern bloc countries were much more able to gain governmental permission to accept standards which emanated from ISO and UNESCO, than from IFLA, and even less able to adopt standards from foreign national institutions such as the Library of Congress. Thus 'official' standards and those developed by UNESCO had a certain kind of influence which in today's different political climate is waning.

Librarians like standards. Fewer librarians speak out against them compared with members of other professions. The main argument *against* standardization is that standards stifle innovation. This is not often voiced as a concern in the library

world though it is in the world of computing or telecommunications. This can be seen in many discussions for example in Espina and Markman's conference paper entitled 'A conceptual assessment of tradeoffs between technological innovation and technological standards' delivered at EMS - 2000 where they discuss the dichotomy between uniqueness, an inherent characteristic of technological innovation, and commonness, a central aspect of technological standards, and conclude that local and regional economies are more likely to be able to take advantage of new product innovation when new technologies are compatible[1]. That is to say, in the technological field there is a running debate as to whether standards are a good thing, but the library world seems to take their value as read.

The standards that come to mind first and foremost are those for cataloging and classification. In the world we now call the book trade, librarians are not alone in favoring standards as title pages of books have for generations included a certain amount of standardization and 'correct spelling' has been valued for a number of centuries. We will see how the book trade and that part of it we call the library world have worked together increasingly on standards development.

Many of the standards we value now as requisites in enabling our international bibliographic computer systems were devised before the advent of computers. This means that the library world was well placed to enter the technological revolution because it already had standard building blocks. Of course one role of standards is to prevent everyone reinventing the wheel. Before computers, it was good to use a classification scheme that someone else had developed in order to save time developing one oneself. Indeed, let the experts develop one and everyone else can benefit from their expertise. Even though there was not then the interaction between libraries as there is now with their information systems linked to each other via the internet, using common standards was felt to have value.

The history of classification schemes and cataloging codes highlights how librarians were thinking globally a century ago. Librarians used metadata long before the word was invented. The advent of computers brought in new tools and libraries have consistently adopted these to foster globalized systems.

Formal international standards as developed by ISO have played a valuable part here and the places where difficulties arise are often in those areas where standards are lacking. UNESCO has attempted to fill gaps in standardization, gaps which have appeared often to be disadvantaging less developed countries. Other UN agencies have been involved along with organizations like Canada's International Development Research Centre (IDRC) and the Development Centre of the Organization for Economic and Cooperation and Development. More recently the European Union has become involved in sponsoring research and development and has contributed to the development of standards and promoted their use in projects across the member states.

---

[1] Espina, M.I.; Markman, G.D. "A conceptual assessment of tradeoffs between technological innovation and technological standards" *in: Proceedings of the 2000 IEEE Engineering Management Society. EMS – 2000*. Piscataway, NJ: IEEE, 2000 p.581-5

International Organization for Standardization

By its very nature, the International Organization for Standardization (ISO) is dedicated to globalization. This organization based in the UN quarter in Geneva does not work alone but through its national member bodies. In the US the national body, the American National Standards Institute (ANSI), delegates its role in the information and documentation area to NISO, the National Information Standards Organization. In most other countries the ISO member is the National Standards body, AFNOR (Association française de normalisation) in France, BSI (British Standards Institution) in the UK, DIN (Deutsches Institut für Normung) in Germany, DS (Dansk Standard) in Denmark, ELOT (Hellenic Organization for Standardization) in Greece, to name those which have been among the most active in the area which is generally called information and documentation and covers not only libraries but also archives, museums, publishing and the book trade.

Countries develop standards through their national bodies with the work being done by representatives of industry, professional organizations and, in the library field, by staff of national institutions co-opting, as necessary, technical experts. They are supported by secretariats who know the rules for making standards and usually have technical expertise themselves. Often a standard emerges which is of interest to other countries or is of interest to the proponents to make international so that the system it supports will be applicable internationally. Sometimes a standard is proposed initially as an international standard. This was the case in the year 2000 with a proposal to develop a standard for data elements for radio frequency identifiers in libraries. A transmitting device can already be placed on a piece of hardware and it can be detected even when boxed. The trade standards for the encoding and specification of the data elements are not necessarily the best for library use where each individual item is unique and requires its own identification. Library security systems could detect books in people's briefcases. Stock checking and even the detection of books mis-ordered on the shelves could be carried out by walking between the shelves with a receiver without having to remove books to read barcodes as at present. In a library, it is important to know which individual copy is being identified whereas in the book trade each copy has identical value. The impetus for this standard came through Danish Standards from the technical team of a multinational company based in Denmark. Obviously, the task of developing and manufacturing the RFID would be more economical if the work applied everywhere and not just in one country or in one sector. ISO represents all sectors, library as well as book trade and it will be important for whatever is eventually agreed to be acceptable to both sectors.

ISO is divided into technical committees. TC 46 is the technical committee for information and documentation and covers most of the standards used in the library field. ISO works with the International Engineering Council (IEC) through a joint technical committee (JTC 1) to produce standards in the areas of office

automation and information technology. This committee oversees, for example, the development of character sets, but because librarians were the first group to have special requirements they developed their own standards within the TC 46 committee. These are discussed later. Subcommittees of ISO/IEC JTC/1 include JTC 1/SC 2 Coded Character Sets, JTC 1/SC 6 Telecommunications and Information Exchange between Systems, JTC 1/SC 23 Optical Disk Cartridges for Information Interchange, JTC 1/SC 28 Office Equipment and JTC 1/SC 36 Learning Technology.

ISO TC 46 celebrated in 1997 its golden jubilee with the publication of a short history by George Richardson[2]. In 1947 there were a number of standards already being proposed for areas which are important in our international computerized systems today which depend so much on standards. The consistent abbreviation of titles of periodicals is vital in automated information retrieval systems, though perhaps we do not abbreviate so much today because of the lesser need to conserve storage space in electronic data. Transliteration of Cyrillic characters: then, a standard to prevent the need for everyone to reinvent the wheel: now vital to ensure in our global systems that everyone transliterating produce the same forms of names. International standards were also mooted for microform readers and for the size of catalog cards. But to indicate that TC 46 then, as now, was at the forefront of technology, it is worth noting that in 1947 standards were being proposed for sizes and quality of photocopies and microcopies. Some of these standards clearly already recognized that librarianship was global. Catalog cards could be circulated around the world. Microforms could be read by machines anywhere. As time went on other standards were added to the portfolio and by 1971 there was a subcommittee entitled 'Automation in Documentation' which had working groups under it on ISSN, ISBN, coding of country names, terminology and thesaurus construction. After adoption of the ISO standard ISO 2709[3] *Format for bibliographic information interchange on magnetic tape*, which is one of the key building blocks of global librarianship, much work was to follow on standards for intercommunication using forms other than tape and standardization of the data content of messages to make international transactions possible in areas like interlibrary loan.

We will refer frequently in this chapter to specific standards that have been established by ISO.

---

[2] Richardson, George. *The first 50 years of ISO/TC 46: a short history.* Berlin, ISO/TC 46 Secretariat, 1997

[3] *Format for bibliographic information interchange on magnetic tape.* Geneva: ISO, 1981 (ISO 2709-1981) now called *Format for information interchange*. ISO: Geneva, 1996

BIBLIOGRAPHIC CONTROL AND ITS CONTRIBUTION TO GLOBALI-ZATION

Bibliographic control is the process of the classification and cataloging of materials held in libraries and similar repositories. There are a number of different schemes available for classification by coded schemes and subject headings. There exist also thesauri used less by libraries than special information services. Libraries are free to make their own choice or even develop their own. The inclusion of the most popular classification schemes and subject headings in MARC records has facilitated their use without further editing by those using them. Though none of these systems are international standards except UDC, they are all standards of a kind. The main difference is that they are each developed by a group of persons employed for the purpose rather than by committees attempting to provide a solution to a problem by seeking a consensus view from interested parties.

Library Classification

Library classification is required today mainly to organize the storage and shelving of library materials. Most libraries use the Dewey Decimal Classification (DDC) or the Universal Decimal Classification (UDC).

The Dewey Decimal Classification was developed by Melvil Dewey in 1873 while working as a student assistant in Amherst College library and published in 1876. It is used in 135 countries.

The UDC was developed in 1895 as a French translation of the fifth edition of the Dewey Decimal Classification. It was first published in French from 1904 to 12907 and it has been published in 23 different languages. The scheme was intended originally as a classification of knowledge rather than as a scheme to be applied to books. It is universally applicable across language and scripts since it includes only numerals and punctuation.

Both these schemes have contributed to the globalization of librarianship. Even users who does not know the working script of a library may still find their way round the shelves by means of the numeric schemes.

Subject headings

Library of Congress Subject Headings (LCSH) devised by the Library of Congress for their own use and now used around the world have been translated widely. The scheme is now in its 24th edition[4]. During the 1970s and 1980s, the

---

[4] *Library of Congress Subject Headings.* 24th ed. Washington, DC:, Library of Congress, 2001

British Library developed and used PRECIS[5] (Preserved Context Index System) which was an attempt to use the power of computing to assist in the generation of subject strings, and additional access points to them. The British Library used this until 1990. Then after a brief flirtation with COMPASS, a watered-down version of PRECIS, they adopted LCSH in pursuit of globalization.

Thesauri

Thesauri have made less impact on libraries than the other systems. They tend not to be so globally applicable because they cover very specialized subject areas though standards have been prepared to assist in their development and to prevent creators of thesauri from having to reinvent the wheel.[6]

Library cataloging

Cataloging rules as we know them, are considered to have originated with Panizzi's rules, a set of 91 rules established in 1841 by Anthony Panizzi who was Keeper of Printed Books in the British Museum[7]. He was first to define what catalogers call the heading and to stress the need for uniformity and consistency in cataloging. It was he who selected the title page as the authoritative source of data. Charles C. Jewett of the Smithsonian continued this work for his library[8] and Cutter generalized the concept of cataloging codes making his 'dictionary book catalog' applicable to small and medium libraries[9]. The first attempt at an international code was in 1908 when the Joint Code, known as AA (1908),[10] was established as a result of cooperation between the Library Association in Britain and the American Library Association. It seems incredible that librarians were thinking globally so early when the distance between the United States and England must have

---

[5] Austin, Derek. *PRECIS: a manual of concept analysis and subject indexing.* London: BNB, 1974

[6] *Guidelines for the establishment and development of monolingual thesauri.* Geneva: ISO, 1986 (ISO 2788). *Guidelines for the establishment and development of multilingual thesauri.* Geneva: ISO, 1985 (ISO 5964)

[7] Panizzi, Anthony. "Rules for the compilation of the catalogue*" in The Catalogue of printed books in the British Museum Library.* London: British Museum, 1841

[8] Jewett, Charles Coffin. *On the construction of catalogues of libraries, and their publications by means of separate stereotyped titles: with rules and examples.* 2nd ed. Washington: Smithsonian Institution, 1853

[9] Cutter, Charles Ammi. *Rules for a printed dictionary catalog.* Washington, DC: US Bureau of Education, 1876 (Special Report on Public Libraries, Part II).

[10] American Library Association. Catalog rules: author and title entries; compiled by committees of the American Library Association and the [British] Library Association. Chicago: American Library Association, 1908

made co-operation difficult and the advantages of common practices less important by far than in these days of instant electronic communication. Various issues arose relating to the concept of corporate author (which was lacking), to name but one example, and revised ALA cataloging rules were published as a draft in 1941 and finally in 1949.[11] Over the next ten years, communication began to improve and international conferences were becoming feasible and, looking with hindsight, it was inevitable that the impetus that had built up for globalization in cataloging should result in an international conference.

In 1961, the International Conference on Cataloguing Principles was held in Paris. Choice of headings was the main preoccupation of this conference[12] which resulted in international agreement on the general aims and principles of author and title cataloging and the publication of *Statement of Principles*, better known as the *Paris Principles.*[13] With this conference we enter the modern era particularly as automation was to influence the globalization of cataloging a short time later. Many new national and regional cataloging codes based on these principles were developed, such as the *Anglo-American cataloguing rules (AACR)* published in 1967.[14] Since the *Paris Principles* had concentrated on headings, it became clear that some standardization needed to be applied to the bibliographic description. This need for standardization of the bibliographic description also proved evident from other initiatives such as the Shared Cataloging Program in the United States as well as the increasing development of computerization in cataloging. The Shared Cataloging Program of the Library of Congress aimed to take descriptive cataloging from national cataloging agencies in countries of publication, and accept it as it stood with attention paid by the Library of Congress staff only to reconciling headings. Though it was clear that if you standardized the access points, descriptions in different styles and with different content could stand together in the same catalog, nevertheless it drew attention to the differences. Computerized processing opened the door to machine-manipulation of bibliographic descrip-

---

[11] American Library Association. *A.L.A. catalog rules: author and title entries*. 2nd. ed. Chicago: American Library Association,1949

[12] International Conference on Cataloguing Principles. *Report of the International Conference on Cataloguing Principles, Paris, 9-18 October 1961*. London: Clive Bingley, 1963

[13] International Conference on Cataloguing Principles. *Statements of principles adopted by the International Conference on Cataloguing Principles.* Paris: October, 1961. Annotated edition with commentary and examples by A. H. Chaplin assisted by Dorothy Anderson. Sevenoaks, Kent: IFLA, 1967.

[14] *Anglo-American cataloging rules*; prepared by the American Library Association, the Library of Congress, the Library Association, and the Canadian Library Association. Chicago: American Library Association, 1967

tions, revealing further needs for standardization. In 1966, IFLA's Committee on Cataloguing initiated a project for an international standard for the descriptive content of catalog entries and commissioned Michael Gorman to make a comparison of practices in different national bibliographies. His report was presented to the International Meeting of Cataloguing Experts which was held in Copenhagen in 1969. The purpose of this meeting was to revise, consolidate and build upon the Paris Principles and look at the question of a standard bibliographic description. Participants agreed on the desirability of this and appointed a small working party to prepare it. This still thought primarily in terms of non-computerized systems when it reported in October 1969: 'The purpose of SBD [standard bibliographic description] is to help in standardization, to help the international exchange of bibliographic data and to help in the comprehension of bibliographic records even when the record is in an unfamiliar language. The SBD is not concerned chiefly with the question of machine readable records. It is concerned with bibliographic records in manuscript, typescript or in printed form and with helping in the transfer of such records to machine-readable form, in that tagging, coding, etc., of bibliographic records is made easier by a generally accepted standard.'[15]

The first edition of *International Standard Bibliographic Description for single-volume and multi-volume monographic publications*[16] was published in 1971. The national bibliographies of France, Federal Republic of Germany and Britain had already agreed in principle to adopt ISBD. It was quickly translated into French, Spanish and Russian. A similar document for serials was proposed in 1971 and this was followed by official and unofficial proposals for other kinds of materials. Although they had the same intentions, their development and supervision were uncoordinated. The original was expanded as ISBD(M) and published in 1974, followed quickly by ISBD(S) for serials. The chaotic development of proposals led to the development and publication of *ISBD(G): General International Standard Bibliographic Description*, published in 1977. Cartographic Materials *ISBD(CM),* Non-book materials *ISBD(NBM)* and Serials *ISBD(S)* were published that year, and Antiquarian *(ISBD(A)*) and Printed Music *ISBD(PM)* the year after.

The main features of the ISBDs are the definition of data elements and their grouping into areas. Data should be usually taken from the document being described: the standards define the exceptions. The areas are

1.Title and statement of responsibility area
2.Edition area

---

[15] Curwen, Anthony G. *ISBD manual: a guide to the interpretation and use of the International Standard Bibliographic Descriptions*. Paris: UNESCO General Information Programme and UNISIST, 1990 (PGI-90/WS.16), p.3

[16] International Meeting of Cataloguing Experts. Working group on the International Standard Bibliographic Description. *International standard bibliographic description, for single volume and multi-volume monographic publications.* London: IFLA Committee on Cataloguing, 1971.

3.Material (or type of publication) specific area
4.Publication, distribution, etc. area
5.Physical description area
6.Series area
7.Note area
8.Standard number (or alternative) and terms of availability area

Additionally, the standard defines prescribed punctuation before the areas and the elements: this helps those who do not understand the language to distinguish between, for example, the title and the person or institution responsible for the work. The standard also indicates which elements are mandatory and which are optional. As stated earlier, it was not one of the aims of ISBD to concern itself with machine-readable records, though authorities such as C. Sumner Spalding hoped that it would.[17] After the publication of ISBD(G) there were moves to change the punctuation but radical changes would have been necessary so none was implemented. For example, a comma is used in the edition area to separate elements. A comma may appear in the text of an edition statement, so comma will not be distinctive enough to indicate it precedes the additional edition statement. One of the examples in the revised ISBD(G) indeed illustrates this point, using a comma in the usual grammatical way rather than in the specialized way as prescribed by ISBD:[18]

The nether world [GMD] : a novel / George Gissing. - Facsimile ed. / edited, with an introduction, by John Goode

Another problematic area is that of parallel elements in other languages, denoted by a preceding equals sign where it is not always possible to match up parallel subtitles with parallel statements of responsibility. The punctuation used was heavily influenced by that available on typewriters for roman script and its applicability for non-roman scripts has been questioned.[19]

The ISBDs continue to be reviewed every five years with subsequent revisions. New ones have been added to those mentioned above; *ISBD (CF): International Standard Bibliographic Description for Computer Files* replaced by *ISBD (ER): International Standard Bibliographic Description for Electronic Resources. ISBD(S)* is under revision as *International Standard Bibliographic Description for Serials and Other Continuing Resources (ISBD (CR))*. Overall they have made a significant contribution to the globalization of the bibliographic record though

---

[17] Sumner Spalding, C. 'ISBD: its origin rationale and implications'. *Library journal,*15 January 1973, 121-3

[18] *ISBD(G): General International Standard Bibliographic Description: Annotated Text.* Rev. ed. München : K.G. Saur, 1992 (*UBCIM Publications. New Series*, 6)

[19] Har-Nicolescu, S. 'Romanized and transliterated databases of Asian language materials', in *Automated systems for access to multilingual and multiscript library materials,* Munich, Saur, 1987, 13-29

they retain an element of Anglo-American bias, and their early adoption by AACR in turn enabled the adoption in the original or in translation of the Anglo-American rules around the world, making them the *de facto* international cataloging standard.

The IFLA Committee on Cataloguing was also instrumental in continuing the work that the Paris meeting had begun on headings. *Anonymous Classics: List of Uniform Headings for European Literatures*,[20] *Form and Structure of Corporate Headings*[21] and *Structures of Corporate Name Headings: Final Report* are representative of a selection of the publications that have been produced by IFLA over the years.

Other bibliographic descriptions

One of the most notable systems of bibliographic description alongside ISBD was the International Serials Data System (ISDS) which was based on an international centre set up by UNESCO for the international control of serials and the maintenance of the standard numbering system known as International Standard Serial Number or ISSN. This was set up in 1974 to allocate ISSNs and serial key titles in an attempt to tame the jungle that was serials. There had been an earlier attempt to set up a standard coding system for serials known as the CODEN which still survives alongside the ISSN and is maintained by the Chemical Abstracts International CODEN Service. The ISSN was quickly adopted by Ulrich, the largest directory of periodicals. The ISDS International Centre published the *ISDS manual* which contained rules of description and methods for encoding the data into machine-readable form. Originally the rules were at variance with those of ISBD(S) and there was a long-running controversy since national libraries which were also ISDS centers had to prepare two separate records for serials. In the early eighties, this was resolved by resolving minor inconsistencies. UNESCO was instrumental in encouraging these minor changes.[22] They also produced *List of Serial Title Word Abbreviations*[23] which contains the words of the titles of serials processed by the ISSN network and their abbreviations. The List includes 45,500 words and their abbreviations in about 50 languages. The words are abbreviated in accordance with the ISO 4 standard for which the ISSN International

[20] *Anonymous Classics: List of Uniform Headings for European Literatures*. London: IFLA Committee on Cataloguing, 1975

[21] Form and structure of corporate headings: recommendations of the Working Group on Corporate Headings. London: IFLA UBC Office, 1980.

[22] Anderson, Dorothy. 'Compatibility of ISDS and ISBD(S) records in international exchange: the background' *International cataloguing* 12 (2) 1983 14-17

[23] *List of Serial Title Word Abbreviations.* Paris: ISSN Centre, 1999

Centre (as the ISDS International Centre is now known) is the maintenance agency.

Rules for description were also developed by other agencies. ISBDs initially did not concern themselves with analyticals. Later, in 1988, *Guidelines for the application of the ISBDs to the description of component parts* were published[24] though these were not regarded as a normal ISBD. This was produced partly in the hope of being able to contribute to the many different efforts to develop manuals which were produced by and for the secondary services abstracting and indexing agencies which also needed standards for producing their records. By then it was becoming the norm to produce records in machine-readable form and a number of different agencies had provided data entry manuals with a view to machine-readable records being available in files of records with great consistency between those produced by different agencies. It was becoming clear that it was not sufficient to propose a well-functioning bibliographic exchange format without detailed data element specifications amounting to rules for bibliographic description. An example of these is the *Manual for preparing records in microcomputer-based bibliographic information systems*.[25] This was produced by the International Development Research Centre (IDRC), a government body in Canada tasked with increasing globalization in the field of information for development. This manual was part of an effort to develop standards and practices for exchanging data between libraries of institutions working in international development under the auspices of EADI (European Association of Development and Training Institutions) and ICCDA (International Coordinating Committee for Development Associations), which were heavily supported by IDRC. The aim was to enable these institutions to exchange data without the comprehensive records required by the national libraries.

Standard numbering systems

We saw how many features of the ISBD were being developed at the dawn of the automation era and another important initiative began then. These were the standard numbering systems. Around the same time, in 1965, W. H. Smith (the largest book retailer in the UK) made a plan to move to a computerized warehouse in 1967 and wanted a standard numbering system for its books. The British Publishers Association's Distribution and Methods Committee devised the Standard Book Numbering (SBN) system in 1966 and it was implemented in 1967 with the setting up by J. Whitaker and Sons Ltd. of the Standard Book Number Agency which later became the UK International Standard Book Number Agency.

---

[24] *Guidelines for the application of the ISBDs to the description of component parts*. London: IFLA UBCIM Programme, 1988

[25] Di Lauro, Anne. *Manual for preparing records in microcomputer-based bibliographic information systems.* Ottawa: IDRC, 1990

The SBN consisted of 8 numeric digits with hyphens separating the publisher element from a running number and that from a check digit which could be X as well as a numeric digit.

At the same time, ISO TC46, the technical committee on documentation, set up a working party to investigate the possibility of adapting the British SBN for international use. A meeting was held in London in 1968 with representatives from Denmark, France, Germany, the Irish Republic, the Netherlands, Norway, the United Kingdom and the United States with an observer from UNESCO. Other countries contributed written suggestions and expressions of interest. A report of the meeting was circulated to all ISO member countries. Comments on this report and subsequent proposals were considered at meetings of the working party held in Berlin and Stockholm in 1969. As a result of the thinking at all of these meetings, the International Standard Book Number (ISBN), the SBN plus an initial digit indicating country, language group or region, making ten numeric digits in all (the last one being a numeral or X) with three intervening hyphens to separate the elements, was approved as an ISO standard in 1970, and became ISO 2108.

The original standard has been revised as book and book-like content appeared in new forms of media and to enable extensions to country codes, but the basic structure of the ISBN as defined in that standard has not changed and is in use today in almost 150 countries. The ISSN came a little later, using a similar check digit system but consisting of eight characters split into two equal parts with a hyphen between them.

Standard character sets

Libraries have special requirements for character sets compared with other businesses . ISO TC 46 Sub-committee 4 Working Group 1 looks after character set standards specifically made for bibliographic use and also registers character sets with ISO/IEC JTC 1/SC 2 which is charged with registering all character codes. The main work today is done by ISO/IEC, since bibliographic requirements are becoming less specialized as the rest of the world requires the same high standards that libraries and publishing have had for some time. With the advent of UNICODE which has the capability to represent a much greater range of characters than the original sets which were limited to 256 characters, it became possible to represent many languages at the same time in one software package on one computer.[26] This sub-committee is now ensuring that the characters in its standards are represented in the UNICODE pages. The flavor of these character sets

---

[26] It is important to be aware that the most common way for the representation of characters with diacritics is to regard each unique combination of letter plus diacritic as one character in the computer character set. Thus a language with 3 diacritics appearing on one letter of the alphabet and having 26 letters in its alphabet would need 29 characters to represent its complete alphabet. Taking into account all usages of the latin (usually called the roman) alphabet in Europe, one can see how 256 characters would not be sufficient.

can be deduced from the following select list: *ISO 5426-2:1996 – Extension of the Latin coded character set for bibliographic information interchange Part 2: Latin characters used in minor European languages and obsolete typography* is intended to supplement the standard character set ISO 646. *ISO 5428:1984 – Greek alphabet coded character set for bibliographic information interchange* is an alternative Greek character set for library use. *ISO 6438:1983 – African coded character set for bibliographic information interchange* provides characters which include the letters and their diacritics for many of the languages spoken in the African continent. *ISO 6630:1986 – Bibliographic control characters* consists of characters for specialized purposes required in library activities such as indicators of non-filing characters like articles at the beginning of titles. Armenian and Georgian which have their own alphabets have their own standards.

TOOLS FOR LIBRARIES TO SHARE DATA

Libraries began to use computers in the 1960s. The first software systems were built, as were other contemporary systems, with fixed-length fields, for accession number, borrower number, date due for return. The library patron might receive a postcard informing the book was due for return with only the accession number to identify the book. The next more sophisticated development was a larger file containing author, title, publisher, date, classification number, accession number. Each copy had an entry in the table and a link with the earlier kind of system was used to provide a primitive catalog and circulation system. It was quickly realized that this kind of database was not satisfactory. Libraries needed to use computers to reproduce the catalog card with its multiplicity of different kinds of data. The Library of Congress had provided a printed card service from 1898. They looked at ways of automating this. In 1966 they developed a MARC (Machine-Readable Cataloguing) format for the MARC Pilot Project which aimed to develop procedures and programs for the conversion, file maintenance and distribution of MARC data.[27] An important feature of MARC was its compatibility with Anglo-American Cataloguing Rules and its hospitality to classification schemes. Many foreign librarians took an interest in this, most notably those from the British National Bibliography (BNB) who also provided a card service, so the next phase included the development of a standard communications format specifically suited to interchanging data between many organizations across national boundaries.

---

[27] Avram, Henriette D. *MARC: its history and implications*. Washington, DC: Library of Congress, 1975

The establishment of MARC and its record structure was followed by a flurry of activity not only in the world's national libraries but also in international networks and in the not-for-profit sector. A number of organizations took the record structure and developed their own implementations for their own purposes. MARC records were primarily intended for national bibliographies. Abstracting and indexing services wanted to computerize the preparation of their publications and made variations on the MARC format, usually adding different fields appropriate to the kinds of material they cataloged. Computerization also facilitated the collection of data from diverse sources. The record structure was adopted as a US standard, then a British standard and then as an international standard, ISO 2709.[28] This standard record structure helped to make possible the use of common software across these different implementations. INSPEC was one of the first companies to adopt a MARC-like format, then Chemical Abstracts and around the same time the International Atomic Energy Agency of the United Nations (IAEA) set up INIS (International Nuclear Information System) and the UN Food and Agriculture Organization shared IAEA's processing systems for AGRIS (Agricultural Information System). AGRIS is a good example of how standards helped to set up a global organization. It started out as an organization, which received input in OCR[29] from diverse institutions, progressing to receiving data on tape from its regional partners such as the Commonwealth Agricultural Bureau (CAB). Then it moved to distributed data entry on CDS/ISIS, a package that owed its effectiveness to the existence of standards. This was also a package that promoted the use of standards as it enabled more organizations to create records which could then be exchanged with others to build up large databases.

---

[28] *Format for bibliographic data inetrchange. op.cit.*

[29] OCR – Optical Character Recognition: a font governed by standards which consisted of characters made more distinctive from each other in order to enable computers to 'read' them: nowadays not so much used as scanners backed up by computer programs are more able to read normal fonts.

So many new formats were being developed that the international scientific community saw this as a recipe for chaos. Around this time, the International Council of Scientific Unions and UNESCO were jointly developing plans for a World Scientific Information System which was known as UNISIST (United Nations International System for Information in Science and Technology). Countries were encouraged to set up focal points in their Ministries of Science. Scientific information would be freely shared and this could start with bibliographic references. One building block was seen in the *UNISIST Reference Manual*[30] (RM), a standard developed by the UNISIST Working Group on Bibliographic Data Interchange and supported by a short-lived maintenance agency set up in the British Library known as UNIBID (UNISIST International Centre for Machine Readable Bibliographic Descriptions). The RM made adaptations to MARC to create a record structure not for data about books but about journal articles and then inserted into the resulting framework records for monographs and monographic collections; serials were not included since they were the responsibility of another UNISIST center, the ISDS Centre which has already been mentioned in this chapter. Incidentally, this format or rather one based very closely on it is still used in Latin America where it is known as the CEPAL format. The developers of the Reference Manual format came mostly from the secondary services, and many of the requirements they had for their records conflicted with library cataloging codes. This was most self-evident in the kinds of material which libraries cataloged, monographs. Concern was expressed to UNESCO that though the world might be polarized into national libraries and secondary services many organizations straddled the divide particularly in developing countries whose interests UNESCO was - and is - bound to serve. Consultants were often faced when defining an automated system for a developing country institution with a difficult choice as to whether to adopt the *UNISIST Reference Manual* format or MARC. In April 1978 the UNESCO General Information Programme (UNESCO/PGI), the department which had recently taken over the UNISIST program, sponsored an International Symposium on Bibliographic Exchange Formats, held in Taormina, Sicily to try to resolve this. Organized by the UNISIST International Centre for Bibliographic Descriptions (UNIBID) in co-operation with the International Council of Scientific Unions Abstracting Board (ICSU-AB), the International Federation of Library Associations and Institutions (IFLA), and the International Organization for Standardization (ISO), the Symposium was convened 'to study the desirability and feasibility of establishing maximum compatibility between existing bibliographic exchange formats.'[31]

---

[30] *UNISIST Reference Manual*. Paris: UNESCO, 1974. 2nd ed. published 1981.

[31] International Symposium on Bibliographic Exchange Formats. *Towards a common bibliographic exchange format? : proceedings*. Budapest: OMKDK, 1978

The Symposium recognized in its recommendations the need for compatibility to be achieved. Following the Symposium, and as a direct result of its recommendations, the UNESCO General Information Programme formed the Ad hoc Group on the Establishment of a Common Communication Format, which included experts able to present the views of a broad spectrum of the information community. Members of this Group worked at meetings and through correspondence to produce a common bibliographic exchange format that would be useful both to libraries and other information services. At the start of its deliberations the Group decided that the structure of the new format would conform to the international standard ISO 2709, that the core record would consist of a small number of mandatory data elements essential to bibliographic description, identified in a standard manner and that the core record would be augmented by additional optional data elements, identified in a standard manner. (It was decided subsequently that to keep the format simple, the extension of these should be discouraged where there was an existing field which would serve, particularly in areas like notes). It was agreed that a standard technique would be devised for accommodating bibliographic levels (analytic, monograph, multi-volume monograph and serial), relationships, and links between bibliographic entities.

In addition it was affirmed that the Common Communication Format (CCF) should be more than merely a new format: it should be based on, and provide a bridge between, the major international exchange formats, while taking into account IFLA's International Standard Bibliographic Descriptions (ISBD).

Early in its deliberations the Group undertook a comparison of all of the data elements in the *Reference Manual, UNIMARC, ISDS Manual, MEKOF-2*, ASIDIC/EUSIDIC/ICSU-AB/NFAIS *Interchange Specifications,* and the *USSR-US Common Communication Format.* With these six standard formats as a guide, the Group identified a small number of data elements which were used by virtually all information-handling communities, including both libraries and abstracting and indexing organizations. These commonly used data elements formed the core of the CCF. A technique was developed to show relationships between bibliographic records, and between elements within bibliographic records. The concept of the *record segment* was developed and refined, and a method for designating relationships between records, segments, and fields was accepted by the group. The first edition of *CCF: The Common Communication Format*[32] was published in 1984.

---

[32] *CCF: The Common Communication Format.* Paris: UNESCO, 1984

Later, a new manual was published to include those data elements for recording factual information which are most often used for referral purposes. The result was the division of the CCF format documentation into two volumes: *CCF/B for bibliographic information,*[33] and *CCF/F for factual information.*[34] These formats have been much used in the scientific information sector but not generally in the library sector. In India its use has been very widespread where the scientific sector is highly automated but the humanities sectors have tended to shun computerization because of the problems of character sets in a country using many different scripts.

To return to the library sector, even there a lack of uniformity in the use of MARC led to different dialects based mostly on either Library of Congress Practice or BNB practice. BNB made a few 'improvements' to the original MARC format such as an increase of subfields to correspond to the different 'areas' in ISBD. In the late 1970s and early 1980s, a number of national libraries adopted formats based very closely on what became known as LC MARC and UK MARC. The existence of different formats led to IFLA's sponsoring the development of a third major format, UNIMARC, in 1975 which was hospitable to international standards and the ISBDs. Many countries then adopted this as their own exchange format. Otherwise, the idea was that each country's national library would convert its own format into UNIMARC and send to each other UNIMARC records. In practice Canada with a format very close to US MARC (mainly differing because of multilingual requirements) and the British Library with UK MARC, also quite close to the US format, continued to exchange records in their own formats leaving it to the recipient to run computer programs to convert the data. Most countries around the world had adopted either LC MARC (which later when MARBI, the American Library Association's Machine Readable Bibliographic Information Committee, took on an advisory role for MARC became known as US MARC) or UK MARC though they had named the formats for their own countries.

---

[33] *CCF/B : the Common Communication Format for Bibliographic Information*. Paris, UNESCO, 1992

[34] *CCF/F : the Common Communication Format for Factual Information*. Paris, UNESCO, 1992

As long as records were distributed from National Libraries or large utilities (for example OCLC) with their large databases, the existence of numerous formats was not so much of a problem but in the mid to late 1990's smaller institutions began to import records from many different sources into automated library systems. UK MARC users were particularly hit as UK MARC had subtle differences from US MARC. UK purchasers of library systems developed in the United States found they had a need to adapt US library systems for their own uses and this sometimes caused serious problems if it was not correctly adapted. In 1995, the British Library was persuaded to investigate harmonization of US MARC with UK MARC and on 20 July 1995 a meeting took place entitled 'Towards a Common MARC Format' at which it emerged that UK users would like to see three key features from UKMARC retained in the harmonized format:[35] ISBD-related subfield encoding, especially of title information; software-generated punctuation; and the treatment of individual volume information.

---

[35] Ede, Stuart. *Letter to UKMARC Format Users*. 20 January 1997

The Network Development and MARC Standards Office of the Library of Congress was reluctant to make substantial changes to the format as their users had just been through a consolidation exercise when the different formats for the different materials (books, serials, etc) were aligned into one format. Since UK users did not wish to lose those aspects of UK MARC which were felt to be superior to US MARC, the UK opted out of complete harmonization. The Canadian format and US MARC were completely harmonized in 1998,[36] and renamed MARC21 in 1999. Australia adopted US MARC in early 1999 with the installation of a new library system at the National Library at which point the Australian National Library ceased to supply records in AusMARC. In the UK, by 2000, the situation had changed. The British Library commissioned a report from BIC[37] which made suggestions for changes to the format and consultation meetings were held. The UK community had changed its mind and no longer wanted to continue making changes to its own format but instead agreed to adopt MARC21.[38] Bearing in mind the upheaval this would cause, plans were made to phase in the provision of MARC21 records by the British Library over a number of years and UK MARC records would continue to be provided. Why had the situation changed? Probably due to the increasing use of interconnected systems using standards for interoperability such as Z39.50. When the *Bath Profile* was published in June 2000 (following pre-releases of the document), UK MARC was not a preferred format for the delivery bibliographic records. More libraries were moving to MARC21 perhaps because they were purchasing library systems which did not support UK MARC, and suppliers of MARC records such as NetLibrary, which provided records relating to the electronic resources they supplied commercially, were making them available only in MARC21. As far as the British Library was concerned, it benefited them not to have to continue to maintain conversion programs which were inevitably imperfect, for the ongoing conversion of records from the Library of Congress. Globalization of processes therefore demands globalized tools.

TOOLS FOR INTEROPERABILITY

So far we have spoken mainly of sharing bibliographic records by transferring records between systems. The future lies in interoperability under which one system can process data from other systems in a real time on-line scenario which requires for one thing greater consistency between the data held in the systems

---

[36] *MARC harmonisation update [webpage], 1998.* http://www.bl.uk/services/bsds/nbs/marc /marcprog.html

[37] *MARC harmonisation: a report to the British Library from the Book Industry Communication, Bibliographic Standards Technical Subgroup.* London: BIC, 1999

[38] 'MARC21 comes of age'. *Library Association record* (May 2001) 103(5) p259

which is why we need to look first at the developments which have taken place in authority control.

Authority files

Globalization of bibliographic data has required international work on authority files. Authority files consist of authority records, records containing data representing a preferred or established form of heading (author, title, subject, classification mark). Normally the records contain information for the cataloger on variant forms and also on rejected forms linking to the preferred form. The aim of an authority file even in the smallest library is to select a preferred form for each entity and to ensure that each entity is represented by a distinctive name, adding qualifiers to distinguish where necessary. This is to satisfy the bibliographer who needs to know which books are by John Smith –1 and which by John Smith –2. Names of institutions (corporate bodies in the language of cataloging) may be qualified by place, persons by dates of birth and death. Cataloging codes define the form of names of persons and institutions and give options for these qualifications. Subject terms are usually defined in a thesaurus or scheme and classification schemes are defined in their schedules. Qualifying terms which are required to make names unique are not easy to apply in a consistent manner. For example, in a small library's collection there are less likely to be persons with the same name with headings requiring qualification than in a large library. If cataloging codes are identical there should be fewer problems of interpretation. A significant problem for users of AACR was that from 1967 for the life-time of the first edition two versions, a North American and a British version, co-existed. Corporate bodies for example were entered under place in the US. Qualifying dates were applied where known to personal names in North America but in the UK only where necessary to resolve ambiguity. When the British Library and Library of Congress produced a microfiche of records in alphabetical order of main heading in the 1970s it was clear that something needed be done because of multiple sequences. There have been numerous initiatives over the years, but however close the cataloging rules and the principles for their interpretation there is going to be room for ambiguity so probably the only way to achieve a common authority file is to allow one agency superiority over all others and to accept their decisions. Bibliographic records can use a unique number as a key and the inclusion of the ISBN and indeed the CIP (cataloging in publication) record on the title page verso in the early days of the CIP initiative helped to bring global standardization to the bibliographic record. Every library can use the same bibliographic record for a book. That is except for the authority data since here each catalog will wish to establish a unique authority. Even if an International Authority Data Number (ISADN) were established, there would still be no solution to the consistency of names across different authority files. To return to the problems between the Library of Congress and the British Library, a Memorandum of Agreement of the Convergence of Cataloging Policy was signed in 1996. This set the scene for the establishment of a joint 'An-

glo-American Authority File'. Standards for the production of this file were agreed, but policies in some areas were still not agreed, such as the preferred form for the word 'Department' in headings (in the US it is 'Dept.').[39] Nevertheless, authoritative forms of names have been agreed as being the responsibility of the country of origin but so far few national libraries have agreed to take what other countries have presented them with and there will always be authors who belong to more than one country. Barbara Tillett who is chief of the Cataloging Policy and Support Office of the Library of Congress has stated that many questions about authorities still remain unsolved.[40] Working at a global level has not yet been solved when it comes to the establishment of common global authority files, even though many tools are in place to facilitate that goal.

Interoperability of catalogs: Z39.50

It should never be forgotten that MARC was developed in the days of main-frame computers and was intended to facilitate transfer of records between systems on magnetic tapes. In practice, with the onset of internet-based communication, the exchange medium changed and records in the MARC format were sent between institutions in electronic form. Other requirements came to the fore. As a result, a standard was developed by NISO known as *Z39.50 – Information retrieval*. As its abstract states: "This standard specifies a client/server based protocol for Information Retrieval. It specifies procedures and structures for a client to search a database provided by a server, retrieve database records identified by a search, scan a term list, and sort a result set. Access control, resource control, extended services, and a "help" facility are also supported. The protocol addresses communication between corresponding information retrieval applications, the client and server (which may reside on different computers)." This standard was published in 1995.[41] An important feature is that it includes 'profiles' for different sectors of the information community. These specify identifiers for indexes which may be (and in the library community will be) based on MARC fields. The existence of this standard has enabled the setting up of 'clumps'. These are virtual union catalogs. In the past a union catalog could be set up only by adding data relating to different libraries' holdings to records in a catalog in one database. Today this standard enables a client to search many different servers, and the servers may be built on the software from different suppliers. The profiles supported for use within the

---

[39] *Cataloguing policy convergence agreement*. Washington, DC: Library of Congress, 1996. http://www.loc.gov/catdir/pcc/converge.html

[40] Tillett, Barbara B. "21st century authority control: what is it and how do we get there?" http://www.oclc.org/oclc/man/9391ausy/tillet.htm

[41] *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification* Bethesda, MD: NISO, 1995. (ANSI/NISO Z39.50-1995). Also available as an international standard: *Information retrieval (Z39.50) Application service definition and protocol specification*. Geneva: ISO, 1998 (ISO 23950)

standard allow client and server systems to be quite precise about indexes and the means of searching them (for example whether right truncation is possible). Both ends, client and server, can specify between author, title, subject standard number and other types of index using codes in the profiles. When the client has searched a number of external databases and found matches, the records are returned to the client in a number of different formats, one of which is MARC21. The client software then processes the records to sort them and display them on the client's screen. Around the world a number of 'clumps' have been set up to facilitate cross-catalog, cross-system searching. Those in the UK have had possibly greater problems than elsewhere because of the prevalence of two different MARC formats, UK MARC and MARC21. Even if a high level of compatibility is achieved in the indexes, the records returned to the clients for processing will be difficult to match if the records are not cataloged according to the same rules or held in the same MARC format. Records now originating in the British Library or Library of Congress are becoming closer, though not identical, thanks to initiatives mentioned above. However in this context there is still the problem of the MARC formats with their slight but important differences. MARC21 would enter a name as '$aShakespeare, William': UK MARC as '$aShakespeare$hWilliam'. Processing does not usually get below the level of the subfield (represented by the $a, $h combination of characters). On the fly conversion to a common standard format as part of the client processing could be contemplated but requires the client to have up-to-date conversion algorithms. The need for a common standard has precipitated the development of the Bath Profile which recognized the need for a common MARC format and required records to be returned to the client in MARC21 or UNIMARC, excluding UK MARC. this may have been one of the reasons why the UK community agreed in 2001 to adopt MARC21. You cannot have a global system without globally applicable tools to develop it. Until that takes place and UK libraries using UK MARC have converted their data, there will be a good chance of a set of returned records including duplicates, reminiscent of the early attempts to combine US and UK records in a common microfiche mentioned above.

The *Bath Profile* is so named because it was conceived at a meeting held at the UK Office for Library and Information Networking (UKOLN) based in Bath England. UKOLN is a key player in the standards field, supported initially by the erstwhile British Library Research and Development Department, which later became part of the Library and Information Commission and then Resource: The Council for Museums, Archives & Libraries. UKOLN also receives funding from the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils (whose terms of reference include seeking to optimize the use and convergence of information and communication technology with information itself as held in libraries), as well as by project funding from the European Union and elsewhere not to forget the University of Bath where it is based.

The Bath Profile: An international Z39.50 specification for library applications and resource discovery[42] identifies those features of the Z39.50 standard that are required to support effective use of Z39.50 software for a range of library functions, such as basic searching and retrieval of bibliographic records for cataloging, interlibrary loan, reference, and acquisitions. The profile defines both a core set of basic author, title and subject search and retrieval specifications across a variety of library databases, and a set of more complex searches. The functionality and specifications identified in the profile are intended to be incorporated into more detailed national, regional, provincial/state, and local agreements. Its use should make easier and more accurate searching and retrieving information from multiple databases. Users will only benefit from accessing databases which have implemented the profile. Librarians will not have to develop detailed search specifications or indexing rules individually but can benefit from the expertise of the international experts who developed the profile. Customers of integrated library systems may have more leverage when dealing with vendors by asking for support for an international profile rather than a locally developed profile requiring custom development. Existing users of databases compatible with Z39.50 may find that this profile could be implemented merely by adding an index to the database (for example the creation of an author index to a particular specification if the current system only has name indexes). The profile also includes the support of additional character sets to accommodate international retrieval.

Interoperability of circulation

A new standard is being developed by NISO known as NCIP (NISO Circulation Interchange Protocol). This standard will define the various transactions needed to support circulation activities among independent library systems. Circulation activities include user and item inquiry and update transactions, such as hold or reserve, check-out, renew, and check-in. The new protocol is expected to support the circulation of printed and electronic materials and will facilitate direct patron borrowing, remote patron authentication, on-line payment, and controlled access to electronic documents. [43]

NISO Standards Committee believes that moving to a national consensus standard will facilitate the development of open systems required when libraries are increasing their need for interoperability in circulation systems, whether it is using self-check hardware from different suppliers or satisfying inter-library loans by giving permission to a user to access data held outside the user's own library to

---

[42] Bath Team. *The Bath Profile: An international Z39.50 specification for library applications and resource discovery.* Bath, UKOLN, 1999 (An internationally-registered profile of ISO 23950)

[43] NISO. Standards Committee AT. *Circulation Interchange Protocol (NCIP)* [webpage]. Washington, D.C.: NISO, 2001. (http://www.niso.org/committees/committee_at.html)

retrieve a particular electronic document. It is not yet known if this standard will be adopted as an international standard or as a national standard elsewhere, but it will almost certainly achieve the status of a *de facto* international standard since the US market for which it is intended is not isolated from the rest of the world and indeed is the largest sector of a global market.

METADATA

The *Free On-line Dictionary of Computing*[44] defines metadata as: 'Data about data. In data processing, meta-data is definitional data that provides information about or documentation of other data managed within an application or environment. For example, meta data would document data about data elements or attributes, (name, size, data type, etc) and data about records or data structures (length, fields, columns, etc) and data about data (where it is located, how it is associated, ownership, etc.). Meta data may include descriptive information about the context, quality and condition, or characteristics of the data'. Metadata was around in libraries before the term began to be used. The term originated from attempts to define the wider world than just the library world and in this context it was felt necessary to use a more neutral term than 'catalog data'. Metadata is sometimes defined literally as 'data about data,' but the term is normally understood to mean structured data about resources that can be used to help support a wide range of operations. These might include, for example, resource description and discovery [the purpose of library catalogs], the management of information resources and their long-term preservation.[45]

The library catalog is the most common set of metadata records with its elements that describe a book or other library item: author, title, date of creation or publication, subject coverage, and the call number specifying location of the item on the shelf.

The relationship between a metadata record and the resource it describes may exist in two different ways. Metadata elements may be contained in a record separate from the item, as in the case of the library's catalog record: this is the traditional form of metadata. Alternatively, metadata may be embedded in the resource itself. Examples of embedded metadata include the cataloging in publication (CIP) data printed on the verso of a book's title page; or the header in a web page. Many metadata standards in use today, including the Dublin Core standard, do not prescribe either type of linkage, leaving the decision to each individual implementation. Although metadata has been around for some time, it is the second

---

[44]   Howe,   Denis.   *The   free   on-line   dictionary   of   computing*   [webpage].   http://www.foldoc.org/

[45] Day, Michael "Metadata in a nutshell" *Information Europe* **6** (2) Summer 2001, 11.

category, that which is contained in the article itself, which has captured the imagination of librarians. If you have a document in electronic form and its metadata within it, that metadata may be extracted and entered into catalogs, thus reducing the labor-intensive activity of cataloging, both its creation and keyboarding. In the case of web pages which are so easy to produce and are therefore so numerous, the inclusion of metadata by their creators is the only possible way that librarians will be able to catalog the material. Anyone who has attempted to find web pages online using one of today's popular web search services has likely experienced the frustration of retrieving hundreds if not thousands of 'hits' from all over the world with limited ability to refine or make a much more precise search. The wide-scale adoption of descriptive standards and practices for electronic resources will improve retrieval of relevant resources from the internet.

Special efforts to control metadata

Most of the directed efforts to develop metadata have been made under the auspices of ISO. There are probably two exceptions: Dublin Core and EDI. EDI has been adopted by ISO but began as an in-house UN initiative

Dublin Core

The need for "standardized descriptive metadata" has been addressed by the Dublin Core proposals.

The Dublin Core metadata set is especially concerned with resource discovery of document-like objects on the internet and does not primarily apply to metadata of traditional bibliographic materials. It should be borne in mind that Dublin Core does not cover all possible requirements for metadata, but serves as a good and well-developed example. The standards for data on web pages are notoriously free and easy. Standards for indexing are also difficult to achieve anyway, particularly if the indexing is to be consistent across more than one discrete catalog; the web is global so the task of indexing across the web is bound to be difficult. The structure or syntax of web pages is also customarily free and easy, though there are certain constraints. Dublin Core is shorthand for the Dublin Metadata Core Element Set which was agreed at the OCLC/NCSA Metadata Workshop in March 1995. It is intended for the cataloging of electronic resources and it is generally held that it should be the standard used on web pages for the 'catalog record', if indeed there is to be one: 'The Dublin Core is the leading candidate as a lingua franca' for resource discovery on the net'.[46]. It is worth noting that it is intended to

---

[46] Miller, Eric. *Dublin Core metadata.* [Dublin : OCLC, 1995?] http://purl.org/metadata/dublin_core

be usable by non-catalogers (e.g. the authors of web pages) as well as by those with experience with formal resource description models (i.e. catalogers).

Here is an example of a Dublin Core record.

```
<META NAME="DC.Title" CONTENT="Presentation of UNIMARC on the Web : new fields,
including the one for electronic resources - 64th IFLA General Conference - Conference
Programme and Proceedings">
<META NAME="DC.Author" CONTENT="Holt, Brian">
<META NAME="DC.Subject" CONTENT="Bibliographic exchange formats, Permanent UNIMARC
Committee, UNIMARC">
<META NAME="DC.Description" CONTENT="This paper treats two topics, the UNIMARC
information which is held on the web and the treatment of Internet documents by
the UNIMARC bibliographic format">
<META NAME="DC.Publisher" CONTENT="International Federation of Library Associations">
<META NAME="DC.Type" CONTENT="Text">
<META NAME="DC.Format" CONTENT="text/html">
<META NAME="DC.Language" CONTENT="eng">
<META NAME="DC.Date" CONTENT="1998-06-29">
<META NAME="DC.Identifier" CONTENT="http://www.ifla.org/IV/ifla64/110-161e.htm " >
```

In this record, the author's name is inverted: there is nothing in Dublin Core to indicate this is necessary. There could also have been an entry:

```
<META NAME="DC.Subject" SCHEME="UDC" CONTENT="025.3">
```

Incidentally, though this example was created manually from the IFLA HTML page which it describes, it could have been automatically provided. Though UKOLN have developed a Dublin Core generator *DC-dot*,[47] it cannot make as good a job as a cataloger can.

The hope is that search engines and web-crawlers will discover and use the Dublin Core metadata in their indexing, since metadata need to be of a higher standard than other parts of the electronic resource.

Electronic Data Interchange for the book trade

Electronic data interchange or EDI is the direct communication of messages necessary for electronic trading between computer systems using telecommunications networks including the internet. Trading messages are most usually orders and invoices but EDI can develop a very sophisticated information exchange. The current standard in Europe is EDIFACT which is officially *UN/EDIFACT: United Nations Rules for Electronic Data Interchange for Administration, Commerce and*

---

[47] Powell, Andy. *DC-dot: a Dublin Core generator.* Bath: UKOLN, [1997?]. http://www.ukoln.ac.uk/metadata/dcdot/

Transport.[48] This incorporates a number of different rules, for example EDIFACT syntax rules (ISO 9735). EDIFACT was not developed exclusively for the book trade but special applications of it have been developed. In North America, BI-SAC (the Book Industry Systems Advisory Committee) and its Canadian counterpart CBISAC pioneered the development of the BISAC fixed-length formats for book trade transaction messages. BISAC are committed to migration to a common set of EDIFACT standards though this will take some time. So, currently, in the UK, libraries which wish to use EDI need to be able to produce both BISAC and EDIFACT messages. In the UK there is TRADACOMS, the UK national EDI standard which will eventually migrate to EDIFACT. This is managed by Book Industry Communications (BIC). They were set up by the Publishers Association, the Booksellers Association, the Library Association and the British Library, thus spanning all sectors of the book trade. BIC has produced book sector implementations of a number of messages based on the UK Tradacoms EDI standard, which include order, acknowledgement, delivery advice, invoice, credit note, statement and price and availability update. BIC is the secretariat for EDItEUR, the European Book Sector EDI group which is recognized by the European Union, the Western European EDIFACT Board, and the European federations of library, booksellers and publishers' associations (EBLIDA, EBF and FEP). EDItEUR, as BIC, does not devise its own message standards but interprets and sometimes extends existing international EDI standards for application to the book trade. EDItEUR works closely with EAN International, which, based in Brussels, coordinates the World Wide EAN article numbering and bar-coding systems and implements the EANCOM subset of EDIFACT message standards.

ISO initiatives

Any industry can have its standards validated by its national standards body and ultimately seek to secure ISO authentication. The book trade is part of the trade community and the development of its building block standards is by the trading community. Libraries have needed their own standards that other industries have not required.

The Taormina Symposium mentioned earlier was co-sponsored by ISO and was followed by a meeting of the ISO TC 46 Working Group which looked after data elements. One of the outcomes of the main part of the Taormina meeting was a proposal to construct a data element directory collating the data elements of the main source formats. The ISO meeting decided to build on this work when it was achieved but expand it in terms of data element directories for particular library functions. The end result of this was a set of data element directories (ISO 8459)

---

[48] UN/EDIFACT Working Group. *United Nations Rules for Electronic Data Interchange for Administration, Commerce and Transport.* New York: UN, 1995- Available at: http://www.unece.org/trade/untdid/welcome.htm

in five parts.[49] These have had an influence on inter library loan protocol standards but have been used to align work in the other areas they cover. Another important initiative is the ISO ILL[50] standard to give it its informal name. The work on this has to be seen in the context of developments at the time it was started. At that time, in 1978, the work on the ISO OSI Reference model was beginning. This is an abstract description of the digital communications between application processes running in distinct systems. The model employs a hierarchical structure of seven layers. Each layer performs value-added service at the request of the adjacent higher layer and, in turn, requests more basic services from the adjacent lower layer. The Physical Layer, Layer 1, is the lowest of seven hierarchical layers. The major functions and services performed by the physical layer are: (a) establishment and termination of a connection to a communications medium; (b) participation in the process whereby the communication resources are effectively shared among multiple users, e.g., contention resolution and flow control; and, (c) conversion between the representation of digital data in user equipment and the corresponding signals transmitted over a communications channel. The next layers are Data Link Layer, Network Layer, Transport Layer, Session Layer, Presentation Layer and Application Layer. This highest layer interfaces directly to and performs common application services for the application processes; it also issues requests to the Presentation Layer. The common application services provide semantic conversion between associated application processes. Examples of common application services of general interest include the virtual terminal, and computer job transfer and manipulation protocols. This reference model was not devised by librarians but librarians were one of the early users since they were interested in sharing data. The standards being developed by ISO TC46 were for the most part standards for the application layer.

The ISO ILL standard was an important step in getting libraries to be able to use OSI for their processes. ILL is in fact an activity which has been conducted at the global level for many years as British Library Document Supply statistics show: they satisfy many international requests, in the year 1999-2000, out of 4.2

---

[49] *Bibliographic data element directory*. Geneva: ISO, 1988-2000. *Part 1: Interloan applications* (ISO 8459-1:1988); *Bibliographic data element directory Part 2: Acquisitions applications* (ISO 8459-2:1992); *Part 3: Information retrieval applications* (ISO 8459-3:1994); *Part 4: Circulation applications* (ISO 8459-4:1998); *Part 5: Data elements for the exchange of cataloguing and metadata* (ISO CD 8459-5)

[50] *Open Systems Interconnection Interlibrary Loan Application Service Definition*. Geneva: ISO, 1997.(ISO 10160:1997). Includes: *Open Systems Interconnection Interlibrary Loan Application Protocol Specification Part 1: Protocol specification* (ISO 10161-1:1997); *Open Systems Interconnection Interlibrary Loan Application Protocol Specification Part 2: PICS Proforma* (ISO 10161-2:1997)

million requests, almost 1.3 million were from outside the United Kingdom.[51] The standard was developed with a large amount of Canadian effort and was approved in 1991. The standard consists of service definition (ISO 10160) and protocol specification (ISO 10161). The protocol permits messages relating to inter-library loans to be exchanged between libraries using different ILL systems but which support the protocol. It defines the types of services that are available, the sequence in which these can be invoked, the information contained in each service and the format of the protocol messages. The standard is used to support transactions in a number of interlending environments. They can range from simple 'library to library' transactions to more complex transactions within a consortium or involving agencies which act as intermediaries. The British Library's Document Supply Centre was slow to adopt this standard because they already had systems in place developed before the standard was even conceived of. However, by the year 2000, partly due to pressure from users of library systems which supported the standard, they had developed the capability to support it on their own computer systems in the form of ARTISO which was accessible to client software devised to support the ISO ILL protocol.[52] Standards have made possible much more complex inter-lending scenarios than the British Library's partially centralized model, but the standards have also enabled the development of a common software client and customers of the British Library will benefit from being able to use the same software to borrow from the British Library as they use for their other transactions.

IFLA's Functional Requirements for Bibliographic Records

We have seen the efforts that IFLA have put into the ISBDs and UN-IMARC. Mindful that the environment within which cataloging principles and standards operate has changed dramatically, the Standing Committee of the IFLA Section on Cataloguing agreed to prepare a study on the Functional Requirements for Bibliographic Records.[53] Shared cataloging systems had been introduced, economical pressures had forced more agencies to adopt minimal level cataloging. On the other hand, new forms of electronic publishing and benefits of technology such as the possibility of getting authors' own abstracts and material from book jackets incorporated into a cataloging record have altered the parameters of the traditional catalog record. Many times had the need for change to take these factors into account been greeted with requests for more specific information on what a biblio-

---

[51] British Library. *Bibliographic services and document supply: fact and figures.* Boston Spa: British Library, 2000.

[52] "ARTISO" *in* British Library. *Document supply news* **68** (December 2000), p3

[53] *Functional requirements for bibliographic records: final report of the IFLA Study Group on the Functional Requirements for Bibliographic Records.* Munich: K. G. Saur, 1998 (UBCIM Publications New Series, 19)

graphic record was supposed to be for, and what the end users really wanted of it. This was raised at various meetings sponsored by IFLA. The terms of reference that were subsequently developed for the study stated its purpose and scope as follows:

"The purpose of this study is to delineate in clearly defined terms the functions performed by the bibliographic record with respect to various media, various applications, and various user needs. The study is to cover the full range of functions for the bibliographic record in its widest sense--i. e., a record that encompasses not only descriptive elements, but access points (name, title, subject, etc.), other "organizing" elements (classification, etc.), and annotations."

"The aim of the study was to produce a framework that would provide a clear, precisely stated, and commonly shared understanding of what it is that the bibliographic record aims to provide information about, and what it is that we expect the record to achieve in terms of answering user needs".

Additionally, the terms of reference also gave a second task to the study group: to recommend a basic level of functionality and basic data requirements for records created by national bibliographic agencies.

Data elements were taken from the IFLA documents, the ISBDs and UNIMARC and from other sources such as the AITF Categories for the Description of Works of Art and from experts who were consulted as drafts of the report were being prepared. The basic elements of the model developed for the study (the entities, attributes, and relationships) were derived from a logical analysis of the data that are typically reflected in bibliographic records using entity-attribute analysis techniques. Authority and subject areas were not covered in any great depth. The model defines the entities that are the focus of authority records (persons, corporate bodies, concepts, etc.) to the extent that they are present in the bibliographic record. It does not analyze the additional data that are normally recorded in an authority record, such as source of authority nor does it analyze the relationships between and among those entities that are generally reflected in the cross-referencing mechanisms found in most catalogs. Recommendations were made for a basic level national bibliographic record. As far as future work was concerned, the group which led the study felt the need to do similar work on subject data. Also, what had been learned from the analysis of the relationships between the entities could equally be applied to the MARC formats which have always been criticized for not representing well in their structure the links between bibliographic entities. This is in fact more true of MARC21 than of UNIMARC.

THE INTERNET AND THE WORLD WIDE WEB

Librarians were quick to use the internet for such things as email and to replace telephone communication and the fax which itself was not available until the early 1980s. Many areas of the profession have become globalized, not least library automation. Users of software now are very likely to have internet connec-

tions except in some institutions in the least developed countries; software and information on how to use it is readily available on the internet. UNESCO's efforts in standardization in the 1979s and 1980s have resulted in the use of standards in its CDS/ISIS software package. Now librarians and systems staff can ask around the world for advice on implementation. There is nothing new in this. There have been discussion lists on AACR for a decade now. But here we have an example of globalization beyond the countries that were originally included, Western Europe, North America and parts of south-east Asia including Japan. Originally global librarianship involved transfer of information in the form of publications to the developed world usually involving specialist acquisitions efforts for government publications, scientific reports and the like. Now a librarian in India wanting to know how to implement a particular standard can get information on this almost immediately from an expert who may be in Brazil who can tailor his answer to the needs of the requester, can provide that information almost instantaneously and can in fact conduct a two-way conversation on the internet to facilitate this transfer of information.

World Wide Web

The World Wide Web is the ultimate in global information access. It is supported by a raft of standards many of which have been used by librarians to establish their own interconnected systems. At the same time the WWW allows globalized publishing. Anyone with access to a server on the internet can publish on the Web. Finding what you want is globalized chaos which can only be resolved by an increase in recall and a much greater increase in precision. Librarians are in the forefront of this, in attempts to increase precision by indexing data, whether it is by encouraging web page creators to provide their own cataloging or by setting up portals using standards understood universally to channel searches into the right direction.

Librarians have been at the forefront of Dublin Core, both its definition and its promotion.

Librarians have built portals using standard classification schemes such as Dewey Decimal Classification (DDC) or UDC. The UK's Bulletin Board for Librarians (BUBL)[54] is a resource used by far more than librarians and which classifies by Dewey and uses its own system of subject terms originally based on the Library of Congress Subject headings. The NISS[55] *Directory of Networked Resources* uses UDC for grouping the resources to which it points and GERHARD (German Harvested Automated Retrieval and Directory) also uses UDC in a sophisticated way as a kind of thesaurus.

---

[54] http://www.bubl.ac.uk

[55] National Information Systems and Services. Directory of Networked Resources. http://www.niss.ac.uk/subject/

LIBRARY AUTOMATION SYSTEMS, STANDARDS AND GLOBALIZATION

Library automation systems originally consisted of software which controlled library circulation and/or the catalog, sometimes with the addition of software to help libraries with purchasing their materials and integrating into the catalog any information from this activity such as indicating books that were on order. Currently these systems do much more such as providing access to inter-library loan services or to electronic resources that may be anywhere in the world. Whether extensively or minimally applied across library housekeeping functions, these systems rely very much on standards. We will not discuss here those standards that are necessary for or are used by computer systems in general. SQL, Standard Query Language is a convention which allows an SQL-compliant database to be queried using a common language. This is applicable to any kind of database, bibliographic or not. Rather we will look at those standards which are necessary for library automation and see how they have influenced the development of library automation systems and what requirement there are for more standards.

The area of library automation which is most standardized is that of the catalog record itself which has a standardized set of descriptors as prescribed by MARC and underlying AACR which could be regarded as a data element directory. The core of the catalog is then highly standardized. Systems can import records without difficulty from other systems because of the standard record structure ISO 2709. Having imported them there may be problems since as we have seen MARC has dialects. OCLC records in MARC21 format may be capable of being imported into a package developed for UNIMARC, but they may not then display correctly. The software will display field 245 in a particular way but in UNIMARC title is 200 and the subfields are different. The more subtle the difference the more problematic the result may be. A system configured for MARC21 will not display UK MARC forenames in the personal name fields. It will probably not include the personal names in the personal name index. Fortunately most systems have parameters in a table which can be changed by systems librarians, but some systems do not. Some UK MARC users have had to convert their data to US MARC standards to use particular packages developed in the United States.

In other areas there is much less standardization. The MARC holdings format came much later than the bibliographic formats and has not been used to the same extent. Many systems have adopted proprietary ways of dealing with records for items or serial holdings and the way they relate to their bibliographic record. Many libraries have had to migrate their data in recent years as earlier generation systems were no longer supported by their developers. Few have been able to migrate serials holdings. Many catalogs do not include detail on the holdings of back runs of serials and have lost data at previous conversions. Item data, borrower data and the data relating to them are valuable to a library for statistical purposes and for indicating trends. There are no standards here so at migration one some data

are lost. That can happen with borrower data where as yet there are no standards in general use though, as we have seen, the development of the NCIP standard promises a future which may be different. Library automation systems are becoming global in their coverage. No longer do they include only the materials in their library but links to electronic materials on the web using the URL. The boundaries of the catalog are blurred in the case of digitized material as even if the library owns rights to it, it will be physically elsewhere. Some material such as abstracts may be available but the full text may not be licensed in a particular library. The policing of this by the owners of the data including ensuring that data that have been purchased are truly available is difficult. Authentication may be by password or by IP address. IP address is part of the internet protocol heavily backed up by standards and authentication by IP address is a by-product of the way the system works. Authentication by user identifiers and passwords to protect them is something only publishers and libraries require. Publishers or the agents who mount the data on internet computers can each have their own different authentication system but this is a problem for the end user. Where IP does not work because a user is dialing up from outside the usual registered network, user name authentication is vital. This can be solved only by a universal system, not by standards. In the UK the academic ATHENS system[56] is a step towards this but many data owners do not trust it. It relies heavily on an institution taking the trouble to remove user names when staff and students leave the institution and are no longer eligible.

A problem has emerged as librarians have required to add links to electronic resources to their catalogs. This problem concerns their location or URL. URLs can change, resulting in huge efforts for catalogers worldwide, efforts which have to be repeated in an identical way for all systems. A new standard is being developed for an Open URL.[57] One of the problems with the World Wide Web is that data which are regarded as part of it (that data accessible through the internet and readable through an HTML browser) is very much uncontrolled. Even if you want to set up a system to help the discovery of resources, there is no guarantee that the resources will remain at the same address or URL (Uniform or Universal Resource Locator). The Open URL standard seeks to solve this fundamental problem of data accessible through the internet.

The Open URL is a protocol for interoperability between an information resource and a service that offers localized services in an open linking environment. The standard will incorporate a syntax for transporting identifiers and metadata describing information objects. It will also include a syntax for communicating with user-specific data which can combine it with user information taking into account, for example, permission to access and then resolve the data into actual links. The aim of the Open URL is that links should lead a user to appropriate resources. A number of information services and library system suppliers were al-

---

[56] *Athens Access Management System.* http://www.athens.ac.uk

[57] *Open URL Standard.* http://www.niso.org/commitax.html

ready working on developing software to implement Open URLs even before it began to be formally developed as a NISO standard in May 2001.

Until recently there were no standards for indexes. Z39.50 has changed this and we are beginning to see extra indexes created for the sake of compatibility with Z39.50 in catalog systems which are following implementations of this standard.

Commercial systems are now beginning to take into account UNICODE[58] which was mentioned above. Of the internationally available database management systems that are used by library systems, SYBASE has adopted UNICODE. Those library systems running under SYBASE can therefore be adapted to implement UNICODE and store characters from many different scripts in one database or even in one record. Hitherto the American Library Association's character set has been implemented but often systems, for reasons of both hardware and software, could not display all the characters represented. It might for example be possible to display the diacritic characters found in French but not the combinations of roman alphabet and diacritics along with Cyrillic used in eastern Europe. Tools like UNICODE when implemented in UNICODE-compatible database management systems with suitable hardware developed by a wider market than just the library market will enable any character to be displayed on the screen.

CDS/ISIS

At the technical level, intercommunication between two systems in the same city is today no different from intercommunication between two systems across the world. What makes for globalization in a particular sector is the prevalence of the technology and the adoption of standards which facilitate the transfer of data and its recognition at the receiving end. More than any other software, CDS/ISIS has facilitated globalization by enabling the technology to be available at very low cost and the standards used in the developed world to be implemented on it in developing countries.

Originally, only organizations large enough to have mainframe computers could benefit from the standardization that went into making intercommunication of library systems possible. In the 1970's, the International Labour Organization (ILO) developed software to host bibliographic databases. This was called ISIS and not only did it support the databases and the data in a MARC-type format, it also provided the capability to output records which resembled those on catalog cards as well as other format structures. ILO appreciated that this software could be used by other institutions and offered it to UNESCO who took it, adapted it and called it CDS/ISIS. They distributed it to other users. IDRC with UNESCO's permission copied the idea and launched MINISIS for minicomputers, specifically the

---

[58] *UNICODE home page.* Mountain View (CA): UNICODE Consortium, 30 Apr 2001 http://www.unicode.org/

Hewlett Packard 3000 range. Other UN agencies used these. They were under-pinned by the use of standards. When microcomputers were developed, it became potentially possible for smaller organizations to jump on the bandwagon, except there was no suitable software. Many of UNESCO's member states in the developing world called on UNESCO to develop a software package for microcomputers. After a false start when UNESCO sponsored German and Austrian development of the IV+V software, the staff in UNESCO who had developed the mainframe version developed the microcomputer version in conjunction with staff of the Instituto Superiore da Pisa[59]. The main standard used by this package was ISO 2709 the record structure of MARC. Many small or poor institutions around the world were now able to develop databases on CDS/ISIS. CDS/ISIS was mainly modeled on the UNISIST Reference Manual format (being a UNESCO product). Unfortunately that format did not have repeatable subfields which in the event were not implemented until 2001 in the basic CDS/ISIS package. However there were ways round this minor problem and many implementations had been made keeping fairly closely to MARC.[60] Between 1985 and 2001, 20,000 registered copies of the CDS/ISIS software were distributed to institutions in both developed and developing countries through a network of some 138 officially appointed distributors around the world, 87 of which were national distributors.

CDS/ISIS is a database management system hospitable to any format which uses three-digit tags, and by itself does not impose any standardization of the bibliographic record. It comes with a sample database based on UNESCO's own practices which in turn were based on MARC. This was not intended as a standard, merely a database to illustrate the software. In consequence a number of different initiatives were made to provide model databases and enable global transfer of data. The best known was *Manual for preparing records in microcomputer-based bibliographic information systems.[61]* UNESCO themselves produced a number of manuals, for example *International information system on cultural development: CDS/ISIS model data base.*[62] Other institutions used CDS/ISIS for preparing records for input into centralized systems and produced manuals such as the *CARIN and CARIS User manual for CDS/ISIS version 2.3.*[63] This manual has

---

[59] Hopkinson, A. "CDS/ISIS: UNESCO's information retrieval package for microcomputers and the VAX minicomputer". *Encyclopedia of library and information science* **57** supplement 20 (1996)

[60] *UNIMARC Demonstration Database. User Manual with diskette in CDS/ISIS;* prepared by the Instituto da Biblioteca Nacional e do Livro, Portugal for the IFLA UBCIM Programme. Lisbon: Biblioteca Nacional, 1993. 43p.

[61] *op. cit.*

[62] *International information system on cultural development: CDS/ISIS model data base.* Paris: UNESCO General Information Programme and UNISIST, 1994 (PGI-93/WS/16)

[63] *CARIN and CARIS User manual for Micro CDS/ISIS version 2.3.* Rome: FAO, 1989

enabled participants in FAO's Current Agricultural Information System to input their own data into the centralized database on CDS/ISIS diskettes from 1989 and more recently by transfer across the internet.

CDS/ISIS now has two versions of the software developed for use on the web: WWWISIS,[64] a CGI (Common Gateway Interface) application for access from a web browser developed by BIREME, the Latin American and Caribbean Center on Health Sciences Information, based in Sao Paulo, Brazil; and JAVAI-SIS[65] developed in Italy by Renato Ennea which requires a dedicated client to access a Javaisis server. BIREME has worked closely with UNESCO on the development of current versions of the system in a way that would not be possible in the purely commercial sector.

Library Cooperatives

Library cooperatives are the ultimate results of standardization. The best known is probably OCLC in the United States, followed by the Research Libraries Group with their RLIN system. BLCMP in the UK, Dansk Biblioteks Center in Denmark and PICA in the Netherlands are other examples. National Libraries throughout the world are union catalog centers: Portugal which used CDS/ISIS for distributed input and Canada are two examples. Without ISO 2709 and MARC these would be very different. National Libraries would have still been centers of their own countries' networks. Cooperatives would not have had the same infrastructure to build on certainly if they wanted to be international. OCLC and BLCMP were built up on the records produced by their national agencies, the Library of Congress and the British National Bibliography (later to become the British Library) respectively. OCLC has been a global institution from its early days after an abortive attempt at a merger with BLCMP. Marta and Tomaz Seljak write about one of the smaller international cooperatives, COBISS which began as a national cooperative in Yugoslavia and became international when Yugoslavia split up.[66] After losing some of the libraries for political reasons which were outside Slovenia where its center remained, many institutions have returned as the political situation has changed more recently.

These institutions began by exchanging magnetic tapes, progressed to online exchange of data usually in ISO 2709 format using disk images of tape files and now are using Z39.50 and investigating XML. From providing databases of records which could be accessed on-line or were used for producing microfiches,

---

[64] *WWWISIS.* http://www.bireme.br/wwwisis1.htm

[65] *Javaisis [Home page]*. http://web.tiscali.it/javaisis/

[66] Seljak, Marta. and Seljak Tomaz. "COBISS: national union catalogue, online bibliography and gateway to other resources" *New library world* **101** (1153) (2000) 12-20 [Note for editors:Tomaz has z hacek)

they have encompassed union catalogs, circulation systems and are now providing access to electronic resources for their members on a global scale.

AFTER MARC: ONIX?

Many people forecast the death of MARC in the 1970s and 1980s. After all it was a standard developed for re-producing catalog cards via magnetic tape. Gredley and Hopkinson in 1990 thought that MARC tags would be more likely to survive that the record structure.[67] Kokabi more recently wrote that the future of MARC was assured.[68] However, the latest versions of XML standards which are being developed by the bookselling end of the book trade are moving away from MARC-type coded tagging and including natural language descriptions of the data elements in their XML tags.

The World Wide Web uses HTML which is a simplified version of SGML. HTML does not always have sufficient capability as it is so diluted. It has necessitated extensions such as XML which is described more fully in Gartner's chapter. This may well be the core of future standardization efforts.

ONIX stands for Online Information eXchange. It refers to a standard format based on XML that publishers can use to distribute electronic information about their books to wholesale, retail and electronic booksellers, other publishers, and anyone else involved in the sale of books. ONIX is a standard for providing all the information relating to publicizing the book, which would include all the information that a library needs in its catalog and more. For a bookseller, the jacket cover of a book contains useful information about that book: cover design, synopsis, reviews, author biography, etc. The Internet has grown as a popular place to buy books. Online, however, there is no physical book to pick up and peruse. What has replaced it is a web page devoted to the book that can be designed to carry all the rich information of the jacket cover, and more, such as audio and video files pertaining to the book. However, getting that data about each book from publishers to booksellers has been a challenge, complicated by the fact that each major industry database company has had a different format preference for receiving the data. This lack of a standard made it difficult and time-consuming for publishers to format and exchange their book information. Hence ONIX. Throughout 1999, the American Association of Publishers (AAP) worked together with the major wholesalers, online retailers, and book information services to create a universal, international format in which all publishers, regardless of their size, could exchange information about books. The group unveiled ONIX 1.0 in January

---

[67] Gredley, Ellen and Hopkinson, Alan. *Exchanging bibliographic data: MARC and other international formats*. Washington: A.L.A., 1990. p.290

[68] Kokabi, Mortaza. "Is the future of MARC assured?" *OCLC systems and services* **12** (3) (1996) 33-36

2000. Much of ONIX is based on the earlier EPICS (EDItEUR Product Information Communication Standards), a much broader standard for defining products which was developed internationally by EDItEUR, drawing on the combined experience of Book Industry Study Group (BISG) in the US and Book Industry Communication (BIC) in the UK[69].

The ONIX standard defines both a list of data fields about a book and how to send that data in an "ONIX message". ONIX specifies over 200 data elements, each of which has a standard definition. Some of these data elements, such as ISBN, author name, and title, are mandatory; others, such as book reviews and cover image, remain optional. While most data elements consist of text (e.g., contributor biography), many are multimedia files, such as images and audio files. Exchanging these optional fields--excerpts, reviews, cover images, author photos, etc. is particularly innovative.

An ONIX message is a set of data elements defined by XML "tags" that conforms to a specific template, or set of rules, also known as the ONIX DTD (Document Type Definition). The DTD defines, among other things, how to order the data elements, and how the elements are interrelated.

ONIX is the only format for the exchange of bibliographical data that uses XML. XML is text-readable, meaning that humans as well as computers can recognize and read the data. Most tags, which define each book data element, consist of English words or abbreviations; for instance, an ONIX message would list the Publisher's name as follows: "<PublisherName>Macmillan</PublisherName>". These factors make it easier for smaller organizations to design and implement ONIX-compliant systems. There are many XML software applications being introduced which will enable this to be used by small institutions. XML can, of course, also be read by an internet browser.

Catalog records in MARC21 fly around the world from sources such as national libraries, cooperatives and other utilities and ISDS. They are incorporated into other institutions' catalogs. Unless they can be effectively converted into ONIX, MARC will remain the medium of exchange. If a two-way conversion does prove possible, XML formats would prove a suitable candidate to replace MARC, with the advantage that they are comprehensible to the naked eye unlike the MARC exchange format.

CONCLUSION

Printed and other bibliographic materials capable of easy reproduction have from the start been global rather than national or parochial, as is information. Librarians, their custodians, have been in the forefront of standards development

---

[69] *ONIX Product Information Standards* [website]; maintained by EDItEUR jointly with Book Industry Communication... London: BIC, 2001. http://www.editeur.org/onix.html

longer than most industries to ensure that the commodities they deal in are treated in a global way. They have been pioneers in the global use of electronic data via the internet and in the use of email to develop their use of these tools and to foster cooperation worldwide.