A comparative analysis of the distinct evolution of cataloging and information technology towards the creation of the next generation library system

Cynthia A. Romanowski
Technical Services Librarian, Assistant Professor
Governors State University, University Park, IL, United States
E-mail address: cromanowski@govst.edu

Copyright © 2016 by Cynthia A. Romanowski. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: http://creativecommons.org/licenses/by/4.0

Abstract:

Technology plays an important role in the evolution of cataloging systems and its use is fundamental in disseminating information to the public. Libraries must find innovative ways of meeting their patrons’ informational needs in order to remain relevant. The challenge is that cataloging rules and formats and library systems are out of synchronization and slow to evolve to meet these informational needs. Catalogers from around the world struggle to find ways to create bibliographic records and access points to accommodate new electronic formats while information systems struggle to search and display the information that cataloging rules and formats seek to convey to patrons. Future cataloging systems are slowly evolving in synchronizing these two entities in order to create a global cataloging environment. They are slowly bridging the language barrier by properly displaying diacritics and special characters. They are making progress on creating a discovery layer that retrieves accurate holdings information for an entire library collection. They are also utilizing new cataloging formats that seek to replace MARC21 and are expandable to accommodate all formats. This evolution is creating an international partnership that will help maintain the relevance of libraries around the world.

To write this paper, the author collects filed data as well as secondary source data through a comparative analysis of the progression of cataloging rules and formats and library systems. The first part of the paper will review the history of cataloging rules and formats and library systems to establish a progression timeline. The second part of the paper will focus on the current state of both and how it is impacting the ability to disseminate information to patrons. The third part of the paper will look to the requirements of the future of cataloging systems. This study concludes that cataloging and information technology will ultimately merge into the next generation library system in order to help libraries remain relevant.

Keywords: Integrated Library Systems, Machine Readable Cataloging, Resource Description and Access, Anglo-American Cataloging Rules, Bibliographic Framework

Introduction

In 2010, Online Computer Library Center’s Perceptions of Libraries, 2010 found that 84% of library users are turning to search engines first for their research instead of their library’s website. This was
up 2% from their 2005 results. With libraries devoting 70% to 90% of their budget to electronic resources, this is an alarming trend. Why are patrons turning to resources outside of the library when we are spending the majority of our budgets trying to keep them within our library presence where we can provide the best and most reliable sources? The OCLC study revealed that patrons found search engines faster, reliable, convenient, and easier to use and these characteristics are not generally what we provide with current cataloging rules and formats and integrated library systems (ILS), as they cannot seem to evolve at the same pace.

**MARC and AACR: Historical Perspectives**

Since the creation of machine readable cataloging (MARC) in 1966 with the MARC Pilot Project and the publication of the Anglo-American Cataloging Rules (AACR) in 1967, MARC and AACR developed an entangled relationship over the course of 40 years that eventually slowed the speed at which libraries advanced technologically.

The MARC Pilot Project started in order to take advantage of computers and develop a way to: 1) use staff more efficiently; 2) be more cost effective; 3) eliminate paper files; 4) update current publication lists; 5) facilitate the production of union catalogs; 6) allow for local cataloging modifications; 7) access national bibliographic resources. Out of participating institutional findings from the pilot program, MARCII was created. MARCII provided the ability to allow humans and computers to communicate using computer readable codes that defined data elements within a bibliographic record. It answered the need for standardization by establishing structure, content, and character set coding, for the creation of bibliographic records that could be shared among libraries to help automate them.

Throughout the years, variations erupted, such as USMARC, UKMARC, CANMARC, KORMARC, AUSMARC, and OCLC-MARC, by modifying the MARCII format to include additional fields and subfields. These variations made it difficult to share data globally among the different MARC formats. In 1977, the International Federation of Library Associations established Universal MARC (UNIMARC) to convert one MARC variant to another, allowing libraries to continue to share bibliographic data despite what variant they were using. In 1999, USMARC and CANMARC combined to form the MARC21 standard that many currently use today. As MARC established itself as the underlying standard for format and structure of a bibliographic, author, and holdings record in the technology world, AACR developed into the standard that catalogers used to describe library materials so they can input the data into appropriate MARC fields.

Standardization amongst the cataloging rules on an international scale started with a collaboration between the American Library Association (ALA) and the Library Association in 1904. Together they published the first international cataloging code entitled Catalog Rules: Author and Title Entries in 1908. The cataloging world did not see the next version of an international code until the 1967 first edition of the Anglo-American Cataloging Rules (AACR) that developed out of the “Paris Principles” of 1961 and the Library of Congress (LC) rules and ALA rules of 1949. The 12 principles that served as the foundation for AACR were: "scope, function, structure of the catalogue, kinds of entry, use of multiple entries, choice of uniform heading, single personal author, entry under corporate bodies… multiple authorship, works entered under title, including principles for uniform headings for works and other issues related to serials,... and entry word for personal names." In reality, this international standard was two versions, one for the United States and the other for the United Kingdom. Amendments and changes occurred throughout the years to each version, making information sharing difficult on a global scale because each country adapted them to their own specifications.
The International Standard Bibliographic Description (ISBD) sought to unify the two versions of AACR by establishing a uniformed descriptive framework that was applicable to all library materials, regardless of their format. It standardized the punctuation so that it was easier to identify bibliographic information within a record despite its language, and made bibliographic data from different libraries interchangeable and easier to convert into MARC.\(^8\) The changes that ISBD provided were incorporated into AACR, and in 1978, the second edition of the Anglo-American Cataloging Rules (AACR2) was published, bringing the previous two versions into one. For years the AACR2 served as a viable standardization for cataloging rules for all resources, with revisions in 1988, 1998, and 2002; however, the introduction of more and more electronic resources caused the library world to evolve into a new set of cataloging rules.

**The Impact of MARC and AACR2**

The standardization that MARC and AACR2 provided the library profession allowed vendors to develop library computer programs and systems to facilitate sharing of information amongst libraries, creating a global cataloging environment and technologically advancing libraries.

Bibliographic utilities found ways to accommodate the MARC and AACR2 relationship, bringing libraries into the technological world. In 1967, Frederick G. Kilgour and the Ohio College Association founded OCLC (Ohio College Library Center) and introduced its Online Union Catalog and Shared Cataloging system (now known as WorldCat) in 1971.\(^9\) Ohio University catalogers inputted the first 133 bibliographic records on August 26, 1971, making it the first occurrence of online cataloging.\(^10\) This introduction of online cataloging allowed libraries to retrieve bibliographic records that were already cataloged and caused similar bibliographic utilities to arise. On locally installed terminals, catalogers edited these records to match local holdings and printed the cards. When no records were found, catalogers performed original cataloging, adding to the database.\(^11\) As bibliographic utilities started expanding beyond the United States, the need to accommodate foreign languages and non-US publications grew.

The LCMARC database started to contain bibliographic records in foreign languages that used the Roman alphabet starting in 1973 and transliterated records for languages such as Cyrillic or Chinese in 1979.\(^12\) Bibliographic utilities started to establish their own character sets to accommodate non-Romanized foreign scripts. The existence of different character sets made it difficult to share foreign language bibliographic records between institutions. In 1988, the development of UNICODE established a universal character set that gave catalogers standardization. This multilingual programming standard used similar principles to USMARC and UNIMARC.\(^13\) The Research Libraries Information Network (RLIN) worked with many academic and research libraries and was the first bibliographic utility to allow bibliographic records to be cataloged and retrieved in vernacular scripts of Chinese, Japanese, and Korean. Despite the overall content growth that bibliographic utilities had to offer, their technological development maintained similar types of searching and user interfaces as they did when they started back in 1971.\(^14\)

Until recently, users of bibliographic utilities still searched databases in ASCII and MS-DOS environments. The Internet, as we know it, launched in 1990, with the first fully web-based bibliographic utility introduced in 1997, when Impact/MARCit and ITS.MARC offered their products through web browsers. These utilities took advantage of the technological benefits
that web browsers offered by allowing users to click on hyperlinks and boxes, and with a single retrieval command, users searched all online cataloging records.\textsuperscript{15} Since then, other companies have launched similar web based products such as OCLC’s Connexion, in 2001, with both a web-based and a Windows client.

Whether web or client based, bibliographic utilities’ searching capability has evolved more robustly to include unique numeric identifiers, such as LCCN or ISBN, title and keyword phrases, format, and many more search parameter capabilities. Users have the option to apply Boolean operators to perform searches that combine multiple search parameters.\textsuperscript{16} Hyperlinks take you to authority records within the bibliographic utility for easy validation from within the bibliographic record, and will automatically update the entry to the established form. Bibliographic utilities evolved more quickly this century than they had in the past with the introduction of the Internet expanding searching capabilities; however, the reliance on MARC makes it difficult to truly exchange bibliographic information with non-MARC compatible software. Aside from their influence on bibliographic utilities, MARC and AACR2 standardization also influenced the development of integrated library systems.

Even though integrated library systems (ILSs) did not come about because of MARC and AACR2, they played a huge role in their evolution into the legacy systems many have today. ILSs developed out of a need to automate specific functions of the library like circulation, and companies would add on different modules like acquisitions or an online public access catalog (OPAC). In 1975, the first library system, later becoming VTLS, was developed by Virginia Tech and gave libraries the ability to do online cataloging and to circulate library materials. Data Research Associates (DRA) was the first vendor to launch an ILS that consisted of both cataloging and circulation modules and an OPAC in 1978. Internationally, the Automated Library Expandable Program Hebrew University (ALEPH) was launched in 1983 in Israel. The popularity of ILSs sparked out of a need to have an OPAC because the card catalog required a high maintenance cost and was functionally limited when searching for materials.\textsuperscript{17}

Since the library field predominantly used MARC, vendors designed their ILSs around the ability to input and output data in MARC.\textsuperscript{18} How the data was inputted affected library patrons’ ability to navigate the OPAC in order to search and find materials.\textsuperscript{19} Providing access to materials made the cataloging module the backbone of the library system. It was the place where catalogers could create, update, and manage the library’s bibliographic database by importing records from bibliographic utilities or other library databases via Z39.50, or by performing original cataloging using templates within the ILS.\textsuperscript{20} In the 1980s, ILSs gave libraries the ability to communicate electronically amongst their cataloging, circulation, acquisitions and serials departments as well as to their vendors and patrons.

Acquisitions clients in the 1980s used Book Industry System Advisory Committee X12 standards to electronically send book orders to vendors. Overtime, ILSs adopted the 1987 UN/EDIFACT international standard which in turn expanded electronic data interchange capabilities and more procurement data could be exchanged with vendors. Orders could be electronically requested by selectors in the vendor’s database and claimed or cancelled by the Acquisitions staff. Serials modules kept track of journal holdings and gained the ability to make electronic claims for missing journals to vendors using the Serials Industry Systems Advisory Committee X12 standard.\textsuperscript{21} OPACs reflected these orders with statuses and displayed up to date journal and book holdings to patrons.
Patrons could find materials through the OPAC instead of the card catalog using command searches and menu driven interfaces. Vendors gradually added on keyword and Boolean searches on top of the existing author, title, and subject options. In the 1990s, ILSs took advantage of the web and additional functions to better interact with patrons within the OPAC. In the late 1990s and early 2000s, patrons gained the ability to renew and to place holds on materials through the OPAC, and to access the library’s subscription databases. The OPAC became more graphical, with format icons and book covers, to make it easier for patrons to find what they were looking for. Search limits via location and format could be applied. With the explosion of technologically innovative features that occurred in the 1980s and early 1990s within ILSs, momentum slowed and by the early 2000s, ILSs no longer advanced at the same pace as non-library technology.

**Catching up with RDA, Next Generation ILS & BIBFRAME**

With the development of new cataloging rules and technologies such as RDA, Next Generation ILS and BIBFRAME, library technology is trying to bridge the divide with information/web technologies in general. The stagnation of libraries is due to the inadequacies of our integrated library systems because they are based on MARC and AACR2, which in turn are based on physical collections. Many libraries today are increasing their electronic resources, more specifically their electronic book (ebook), streaming video, electronic journal (ejournal), and database collections in order to meet the needs of their patrons. This means that current Acquisitions and Serials clients and OPACs are failing to accommodate the demands that these types of materials are placing on them. As processors of physical materials, Acquisitions and Serials clients streamlined acquisition workflows while OPACs freed patrons of the card catalog and made them a little more independent at the library. With the introduction of electronic resources, acquisition and serial workflows are no longer as streamlined as they once were because current library systems cannot properly handle the management of electronic subscriptions. They are not designed to track “the licensing terms associated with each subscriptions, [maintain] current information on what titles and data coverage apply to each of the aggregated content products, authentication details, and the technical contact information.” As technology advanced and the average library user became more tech savvy, they also put more demands on the ILS.

In 1971, the ebook was developed, and by 1993 it evolved into the more familiar formats, making it easier for libraries to incorporate into their collections through their websites in the late 1990s. This incorporation raised patrons’ expectations for on demand information. Today’s patrons want immediate access to books and full-text journal articles. This trend, along with budget cut backs, has sparked patron driven acquisition and prompted more libraries to switch to ejournals. Since current Acquisitions and Serials clients do not handle these materials very well, some libraries opt to make the additional purchase of an electronic resource management (ERM) system to keep track of licensing and pricing.

Patrons are faced with an OPAC that is not user-friendly. It is built on MARC language that catalogers use. MARC may not cater to patrons’ searching abilities because they expect the same type of searching capability they get from a search engine. It is unable to take advantage of modern technology because of its lack of scale and flexibility. Patrons may also not retrieve valuable information because AACR2 bibliographic records can be inadequate in their description of electronic resources, often lacking in proper access points. The intertwined relationship of MARC and AACR2 inhibits libraries from advancing technologically due to the difficulty in separating the one from the other. At their inception, integrated library systems were built around physical collections. RDA’s introduction has
slowly started the process of changing the way libraries manage their print and electronic resources within the existing integrated library systems in order to compete with the World Wide Web.

The influx of digital materials into the library world created a much needed revision of AACR2. Describing relationships between entities was difficult because the format of the resource impacted the description of the resource. Libraries needed a better way to describe these types of items so that library systems more adequately house, manipulate, and display their bibliographic data. Since AACR2 was based on physical resources, it did not adequately accommodate the electronic resources format. The Joint Steering Committee for the Revision of AACR created a draft entitled Anglo-American Cataloging Rules 3, but the feedback from libraries made it immediately clear that an entirely new approached was needed and this led to the development of Resource Description and Access (RDA) that we currently use today.

RDA officially replaced AACR2 as the bibliographic description rules for all library materials in 2013. Using the entity-relationship conceptual framework of the Functional Requirements for Bibliographic Records (FRBR), the goals of RDA aim to have a consistent and flexible standard so that the bibliographic data can be utilized by information communities outside the library realm and to be more global minded in its usage by removing cultural bias. Another goal is to provide the ability for users to see multiple editions, translations, and formats by a particular author and their work through the support of Work-level and Expression-level bibliographic records. Finally, RDA aims to provide guidelines for the description of digital resources that helps enhance the usage of the online catalog and improves searching and browsing capabilities. Moving away from the vocabulary of the card catalogs, its intention is to use terminology that makes sense to the library patron. This vocabulary is controlled so that it is available in linked data environments in order to utilize the Uniform Resource Identifier (URI) that helps link related materials. Its design works with legacy library system formats, such as MARC, and with more common formats, such as extensible markup language (XML) that allow it to interchange data over the Internet. RDA puts the library world one step closer to being more technologically advanced because it starts to take advantage of the Semantic Web and linked data by using Resource Description Framework code (RDF). This allows bibliographic data to be easily manipulated which opens the door to endless possibilities in sharing data beyond the library realm; however, further research is needed to gauge its effectiveness with end-users.

Since the introduction of the Internet, libraries have tried to incorporate new technologies into their current library system. They want to provide the most reliable information to patrons, but find themselves with very little money and resources to do so. They look to vendors for technological change to help them meet the voracious need of information by their patrons. “Libraries increasingly expect web-based interfaces that eliminate the overhead of installable desktop clients and support efficient and ergonomic workflow.” What are the characteristics of the next generation library system, or library services platforms, that are going to help make libraries relevant again?

With the introduction of RDA, possibilities to expand well beyond the realm of libraries and into the World Wide Web exist. RDA builds on technology made available from the Semantic Web and linked data. Information can be linked to other information on the web via the relationships that RDA builds amongst the data. Describing the actual item and not the
format will provide seamless access to electronic resources, creating access for streaming video, audio, electronic books, or electronic journals regardless of where the item is housed or which vendor owns the item. The new system would allow controlled and uncontrolled language where users can input words, including misspelled words, and still be able to find what they are looking for. Discovery layers will provide the patron with one stop searching capabilities that are more reliable and of better quality than any results they would otherwise receive via a search engine. The results from this search will be more accurate because of the improved relevance ranking and federated searching capability.

The new library system will be a “more comprehensive model of resource management” that will accommodate both print and electronic materials as well as the metadata and business rules associated with them. With a comprehensive system, libraries will have the key functionalities of an ERM that could manage licensing agreements for their electronic resources in order to help aid in the tracking of database and electronic journal subscriptions and copyrights. This will enable libraries to streamline processes, efficiently manage resources, and effectively manage staff.

Free of being held to a solitary vendor, the new system will be based on a format that is flexible and expandable to allow for future enhancements and make it easier to upgrade and add new features when technology advances. This new open source software will be customizable so that libraries, no matter what types, can adapt it to more adequately meet their needs. It will also be mobile to appeal to those patrons that are constantly on their tablets, smartphones, and other wireless devices. This will turn libraries into mobile information sources and meet the demands of mobile patrons. More importantly, it will incorporate the bibliographic description framework called Bibliographic Framework (BIBFRAME).

It has been 50 years since the MARC Pilot Project and what was once ahead of its time, has now become a huge anchor putting libraries technologically behind. Initially designed for printing cards, MARC is not suited for modern technology. Building on MARC’s “ability to promote contextual relationships using field data in closed catalogs and in open, Web-based discovery catalogs”, BIBFRAME Initiative aims to make these types of catalog’s contents as consumable as other materials accessible on the Web. The BIBFRAME 2.0 Model is made up of three classes: Work, Instance, and Item. Each class takes different bibliographic information, such as author, places of publication, and item location, and organizes into more useful information. The hope is that BIBFRAME frees “the library bibliographic information from its information silos and allow it to be integrated with the wider web of data.” Even though BIBFRAME is designed to work with RDA, the goal is independence from any specific set of cataloging rules. This independence will avoid future entangled relationships and will prevent libraries from falling behind technologically because of an ill-fated relationship like MARC and AACR.

**Conclusion**

Cataloging rules and formats and integrated library systems have not been on the same timeline. One seemingly moving ahead of the other, yet neither keeping up with library users’ needs, causing libraries to remain behind the capabilities of search engines. Libraries established themselves as technologically advanced in the 1970s when most people did not have computers in their homes. As technology evolved, so did library patrons along with their technological expectations of their library. Patrons now have access to a plethora of
information with the Internet via smartphones, tablets, and personal computers, and expect to get this same amount of information and more from their library.

While libraries have slowly started to update and/or rid catalogs of records in AACR2, MARC still remains, and the stagnation of libraries continue. The need to provide proper access to library materials through cataloging is fundamental to the success of the next generation library system. While things have been slow to evolve, library service platforms have been making progress in the right direction by addressing issues that the relationship between MARC and AACR created, making us more internationally engaged. However, libraries need to take the lead and dictate our needs to vendors or we will be without a voice in the information world.

Acknowledgments
A special thanks to Dr. Michel Nguessan, Mary Konkel, Alexis Sarkisian, Dean Lydia Morrow-Ruetten and all my colleagues at Governors State for your help and encouragement.

References
5. Ibid., 79
12. Ibid., 21.
15. Ibid., 22-24.
16. Ibid., 24.
17. Ibid., 36-37.
   http://search.proquest.com/docview/1662652540, 446.
21. Ibid., 52-53.
22. Ibid., 48-50.
36. Ibid., 80.