

Local Creation/ Global Use

Bibliographic Data in the International Arena

Glenn Patton

OCLC has grown from the original group of Ohio academic libraries to 27,000 libraries located in North America, Europe, Asia, Latin American, and South Africa. Each of the records in WorldCat (the OCLC Online Union Catalog) is a local creation that is available for use across the globe for different purposes. Common issues that must be faced with the expansion of a bibliographic utility include cataloging standards, subject access in languages appropriate to the user, local needs versus global usefulness, and character sets. Progress has been made with the cooperative creation of an international name authority file and the uniform application of ISBD principles. A method of linking various subject vocabularies and an improved infrastructure of MARC formats and character sets are needed. Librarians need new automated tools to provide preliminary access to data available in electronic form and to assist them in organizing and storing that data.

At one point, I had thought of titling this article “It Takes a Village to Build a Bibliographic Database.” Although I eventually rejected the title, I find the “village” metaphor still useful as I attempt to put the varied topics of the other articles in this volume into the practical context of a bibliographic service and its member libraries. As I am writing this from OCLC’s perspective, my comments may or may not apply to other bibliographic services.

In OCLC’s case, that village has become progressively larger as we have grown from the original Ohio academic libraries to 27,000 libraries in 64 countries. Each of the 39 million bibliographic records in WorldCat (the OCLC Online Union Catalog) is a local creation—the work of catalogers in one of those libraries—that can be enriched by catalogers in other institutions. These records are available for use across the globe for many different purposes, the same “generic tasks” that are the foundation of the IFLA Functional Requirements described by Madison in this volume.

Those many different purposes, which bibliographic data in large, shared databases serve, mirror the “cradle-to-grave” life cycle of a village. Bibliographic records are used for initial collection development and selection decisions—that is, bibliographic conception, to support the acquisitions process, and to provide the basis for cataloging and for recording holdings. Bibliographic and holdings data also support resource-sharing activities and can support collection management decisions that lead to weeding—bibliographic death, if you will.

As with life in any village, life in this “bibliographic village” has its ups and downs. There are disagreements; there are conflicts. The village may grow (sometimes dramatically). The environment can change. I’d like to look at

aspects of village life and identify some common threads in how the village has adapted that relate to what we have heard today.

The Village Is Ohio

The original OCLC village consisted of a group of Ohio academic libraries who came together to share bibliographic data in hopes that they would reduce their costs. Even within that relatively homogenous group, there were needs for both shared, standardized data and individual, local flexibility. Those needs were reflected in various ways. Advisory committees formed to set standards for input of records, and fairly early on committee members identified the need for various levels of record content. They also identified the parallel need for the ability of one library to add to and to enhance records created by another library when those additions and enhancements would support the common good of other village members.

The strong emphasis on building a shared database to support both cataloging and resource sharing also brought with it efforts to convert older cataloging into machine-readable form. Village members spent much time discussing how to integrate cataloging created under older rules, as well as older classification numbers and subject headings, with current cataloging as efficiently as possible and without requiring complete recataloging.

The need for local flexibility manifested itself in the ability to make local copies of master records that could be edited as the library saw fit. Flexibility was also evident in the formatting of catalog cards—still the primary record delivery mechanism in those early days—with literally hundreds of options for call numbers with associated locations and oversize stamps.

The Village Is the United States

As the village expanded to include both U.S. libraries outside of Ohio and libraries other than academic ones, these sometimes-conflicting needs for standards and for flexibility continued to grow. Other types of libraries introduced the need for support of additional classification and subject heading schemes. Other types of libraries and other geographic areas brought different viewpoints to the discussions of advisory committees.

The Village Is North America

In the mid-1970s, OCLC member libraries began to encounter Canadian serial records as part of the CONSER

Program, which introduced the element of bilingualism to the village. This was especially evident in name headings for corporate bodies in both English and French.

The Village Expands to Europe

In early 1981, the village was in the midst of the implementation of AACR2, a significant environmental shift that made fundamental changes in village culture. During that same period, OCLC opened an office in Birmingham, England, and began to work with libraries in the United Kingdom. The expansion of the village, coming as it did at the same time as the implementation of AACR2, was perhaps somewhat less traumatic than it would have been under earlier cataloging rules because we were no longer dealing with two different sets of rules. Nonetheless, differences in cataloging practices and differences in the implementation of options present in the rules proved, to paraphrase the poet Dylan Thomas, that we were a village “up against the barrier of common cataloging rules” (Rees 1993).

Addition of records from the British Library to the shared database brought this “separation” to the attention of a broader range of villagers and consternation ensued. Because of differences in cataloging practices and variations in the applications of the rules, these records did not fit well into existing workflows for copy cataloging and thus were in conflict with village goals. After much discussion (some of it heated) and evaluation, calm was restored and villagers adapted their workflows to suit the characteristics of these records.

As OCLC subsequently began to work with a group of university libraries in France, we were able to take advantage of their decision to use the French translation of AACR2 to provide a common ground. Integrating records from these libraries into the shared database, however, brought some new challenges to the village in the form of notes and subject headings in the “language of the cataloging agency,” French. One portion of the village of course, had encountered this challenge before in the Canadian serial records that I referred to earlier, but the effects of the challenge had not really been evident outside the “serials” portion of the village. Records for current European publications created by the French academic libraries were more likely to be used by other village members. Again, some consternation ensued but villagers adapted and incorporated these records into their workflows.

More recently, the village has expanded into central and eastern Europe and villagers have encountered cataloging rules other than AACR2. Since those rules are, however, founded in the principles of the International Standard Bibliographic Description, the transition has been fairly easy. Expansion into this area also brought with it the need

to deal with records in UNIMARC format and in different Latin-alphabet character sets. While making that transition in the structure of the record was not as easy to deal with as the content, the shared database is now richer. More about the character set issue below.

The Village Expands to Asia

As the village expanded to include Asian libraries in 1986, villagers came into contact with non-Roman scripts and other aspects of that “barrier of a common language.” At this point, the village encompasses at least four meanings of the term “football”!

This expansion also moved the village into twenty-four-hour operation. Since village boundaries now covered so many time zones, round-the-clock operation became essential.

The Village Expands to Latin America

More recently, in 1995, OCLC began to work with libraries in Latin America. Here, too, the advantage of having existing translations of AACR2 already in use in Latin American countries has helped tremendously in integrating the descriptive portion of bibliographic records into the village database. New village members, however, continue to point to the need for name headings and subject access in the “language of the cataloging agency”: Spanish or Portuguese. Longtime village residents, who can also benefit from these bibliographic descriptions, want to use them most efficiently with headings in the language of their cataloging agency: English.

The Village Continues to Expand

As the village continues to expand to encompass catalogers in South Africa, it is perhaps too soon to tell what additional challenges may appear on the horizon. After more than twenty-five years, however, villagers have probably encountered some version of the challenge before. These new village residents are already using AACR2 and have recently adopted USMARC so it is likely that they will feel at home in no time.

Further expansion continues to introduce character set challenges. Village catalogers have dealt with some of these challenges by a combination of vernacular data and transliteration or by transliteration only. Other villagers have been more or less patient with this in the past, given that technology up to a few years ago has not been easily able to display either non-Roman script or to print it on cards. Users have, however, been bemused that librarians cannot agree even

on transliteration systems—as is demonstrated by the “U.S.-versus-the-rest-of-the-world” split between Pinyin and Wade Giles and the various transliteration schemes for Cyrillic that are in use in the United States and Europe.

Common Issues

How do these little snippets of village history relate to the topics discussed in the other articles in this volume? Let's pull out some common threads for consideration.

Cataloging Standards

Since the introduction of the International Standard Bibliographic Description in the 1970s, it has been widely adopted both as a set of cataloging rules and as the “foundation” for national cataloging practice. What has not always been obvious is that, while many sets of cataloging rules (including AACR2) acknowledge that foundation, the ISBD principles have actually been mixed with other practices carried over from previous rules and extended in various ways.

The result is subtle variations and minor differences that often have major impact in automated systems. In the “good old days,” a catalog card with a minor typo might well have been filed in the right place simply because the filer's eye read the heading as if it had been typed correctly. Similar variations in a machine-readable records—for example, a word that can be abbreviated according to one set of cataloging rules but not according to another set—might well result in the two records not being recognized as representing the same bibliographic item.

Subject Access

Another common thread in the village history that we read about in this volume is the need for subject access, as well as name headings and descriptions, in languages appropriate to the user and using terminology appropriate to those users. Thus far, village members have accomplished that by maintaining multiple parallel subject access points appropriate to the language and terminology of their users. That method works but it is certainly not the most efficient one to make appropriate relationships and distinctions.

Local Needs vs. Global Usefulness

I began this article by noting the distinction, recognized by the original Ohio villagers, between shared standardized data and individual local flexibility. This model (that is, the master bibliographic record and local copies modified to suit each villager's needs) has served the village well. It is, however,

stretched thinner and thinner by the need for information in multiple “languages of the cataloging agency” and the desire to share that information with other groups of villagers.

Character Sets

MARC has facilitated transliteration by providing parallel fields where data can be represented both in the original and transliterated forms. Technology, however, moves on, in this case fueled more by global business interests than by bibliography. Microsoft now issues software to work with different scripts, and as a result, users of the village’s catalogs are less willing to accept transliteration in the library. Villagers also encounter materials in languages that cannot be represented in character sets currently implemented in USMARC.

What Do Villagers Need for the Twenty-First Century?

Finally, I would like to consider the question “What do villagers need for the twenty-first century?” In a paper for the 1990 Seminar on Bibliographic Records held in conjunction with the IFLA General Meeting in Stockholm, I listed a number of areas in which progress would be of benefit to OCLC villagers (Patton 1992). When I reviewed that brief list recently, I was both pleased that some progress has been made but also sobered by the fact that topics covered in the preconference papers have created a longer list. So here is today’s version of that list:

- International name authority file

On top of my 1990 list was “cooperative creation of authority records.” We have certainly made progress with that effort with more and more villagers participating in the Name Authority Cooperative Program. We are also beginning to make progress toward ways in which the established forms of name appropriate to various languages can be linked and manipulated to produce displays appropriate for various languages.
- Continued reliance on the structure of the ISBDs

Second on my list was “uniform application of ISBD principles.” I noted in my 1990 paper that, as my OCLC colleagues and I had gained experience in working with and evaluating bibliographic data created under rules other than AACR2, we had become convinced that the degree of conformity to ISBD principles was an accurate measure of how well that data could be integrated into the shared database. If anything, the experience of Project REUSE and of the “rules harmonization” project currently underway with the Russian Library Association and our colleagues at the National Library of Russia reinforces

the validity of this view. Further work to align national cataloging rules more closely with the ISBDs and with the IFLA Functional Requirements, while not perhaps the most exciting work, will certainly lead to bibliographic data that is more readily transferable in the global environment.

- Method of linking subject headings

The need for a way to link various subject vocabularies was not included in my 1990 list, but I realize now that it certainly should have been. The techniques described in Tillett’s article in this volume for linking established forms of name headings should be investigated for subject thesauri as well.
- Improved infrastructure of MARC formats and character sets

Also on my 1990 list was the “continued exploration of MARC format capabilities to handle multilingual data.” Much work had already been done at that point to accommodate both vernacular and romanized data in the same record, and Aliprand shows in her article in this volume how the future use of UNICODE can assist us. A logical extension of that capability might be a way of storing data in several languages that share the same character set with coding to allow manipulation. This kind of capability could allow a system to display, based on a single bibliographic record, a record with Spanish notes and subject headings or French notes and subject headings, as appropriate to the user.
- Automated tools

Since 1990, villagers have witnessed an explosion of data available in electronic form. Some villagers have joined to try to provide the controlled access of cataloging rules and subject vocabularies to the flood of information. However, the flood continues to grow faster than they can attempt to provide that controlled access and the “digital indigestion” that Madison described is a result.

Villagers need new tools to help them harvest data that is of use to the village, to provide preliminary access to that data, and to assist them in organizing and storing it for the village’s use.

Conclusion

Each of us could probably make a similar list based on what we have read in the other articles in this volume. You might even have already considered topics that seemed to you to be worth further investigation or functionality that would make life in your village easier.

Many of you as villagers might remember some of the events that I described above. Some of you may well

remember that initial “grass hut” that was the foundation of the village: the shared database that has grown to become the WorldCat we know today. Villagers recognize it as their greatest asset as they enter the twenty-first century.

That initial group of Ohio villagers probably did not foresee that the village would grow to international proportions in only a generation and a half. In another generation or two, will we consider “What in the Universe . . .

Knowledge Access Management on the Intergalactic Level”?

Works Cited

- Patton, Glenn. 1992. The commercial sector. In *Seminar on bibliographic records*. Munich: K.G. Saur.
- Rees, Nigel. 1993. *Dictionary of phrase and fable*. [S.l.]: Paragon Books, 1993.