

Is MARC Dead?

A panel at the American Library Association Meeting, July, 2000

In preparation for this program I followed up on some discussions and some earlier talks on this topic in hopes of finding a clear thread among those who are critical of the MARC format. Instead I found an eclectic series of complaints, among which are:

- MARC is not a good database format
- It's not flexible enough
- It isn't good for display

I also heard:

- It's too complex

And

- It's too simple

Other complaints seem to be about what we do with the records. One person felt that the last name of a personal name needs its own subfield. Another was unhappy with the collection of items that could be found in a single notes field.

Based on these statements, I can't tell if our problem is with the *container* or its *contents*. Nor if it is with the *record* or our *systems*. In fact, I can't get a clear picture of the perceived problem at all.

Like a classic dysfunctional family, I feel that we are talking around the problem rather than about the problem, and for the next few minutes I'd like to do some therapy on our issues and see if we can't find some underlying problems that give rise to our unhappiness.

Out of the Mainstream

The first basic problem that I can coax out of these motley complaints is that MARC is not a mainstream format. Instead, it is particular to libraries, and only libraries. This means that Internet Explorer will never have a "display MARC" plugin, and we'll not be able to buy a standard database management system that has an "import MARC" wizard. We suffer the curse of the early adopters, because when the MARC format was developed there was no World Wide Web for us to conform to. As a matter of fact, modern computing was still a good ways off.

So is this a big problem? No, it isn't; and it isn't because we have an enormous advantage over most other institutions, and that is because we have a very large body of highly codified, standards-based data. In other words, we have excellent, well-coded content, and that content can therefore be rendered in any number of record formats through the use of a translation program. I know this because I have worked with many sets of bibliographic data that are not produced in the MARC format. I can tell you that we spend about five percent

of our time understanding the record format and translating it to our structure, and the other ninety-five percent of our effort is spent on the content -- content that was not created using cataloging rules, that had changed greatly over time, and that followed no perceivable standards.

If your content is good, if it is consistent, you have what you need to feed different record formats or different systems. If your content is not based on standards and if your coding of the content is irregular, no record format can save you. Content is everything in this business, and we have it.

Flexibility

Many critics of MARC complain that the format is not flexible -- and I say that is true, and it is good. We think of the origins of MARC as being related to card production, but it's more than that: MARC was developed as a machine-readable holder for sharable bibliographic description. Whether the output would be as printed cards or downloadable records, the purpose of MARC is to allow libraries to share cataloging for items that they hold in common. It contains that information that is true about that item regardless of what library stores it on their shelves.

This quality of being sharable is absolutely essential to our way of life today. Libraries could not return to doing their own cataloging for each regularly published item they receive. The use of shared cataloging has freed us to put our efforts into other services and into non-traditional types of information.

Anything we do to the MARC record to make it less sharable endangers the whole ecology of today's libraries and we should protect it from changes that effect its sharability. Let me give two examples:

The 856 field

The 856 field seemed like a good idea at the time that it was proposed. But the information in the 856 is not part of the bibliographic description and does not have the same longevity and stability as the other information in the MARC record. Note that when the 856 field was proposed the World Wide Web had not yet happened and the 856 was developed with subfields for the phone numbers of dial-up bulletin board systems. The 856 field, in today's world, is not a sharable field because libraries tend to have different routes that they take to get to online resources, whether it's because they run a proxy server or because they have a URL that identifies them to the vendor.

Library system developers have recognized that the URL needs to be highly flexible and they have developed system services that don't rely on a URL inside the MARC record. I refer to the SFX technology, Jake, and the work we have done at the University of California. Separating the URL from the bibliographic record is a much better solution to the problem.

Multiple Versions

A few years ago many of us spent hours at conferences like these discussing a concept called "multiple versions." The idea was that if you had more than one version of a work, such as a hard copy, a microform copy and an online copy, you could create just one bibliographic record for the work with links from small "version" records. The discussions about multiple versions broke down long before we got to the point of trying to fit it into the MARC record; they broke down when we realized that multiple version records represent a point of view of an individual library and therefore do not result in a sharable bibliographic record.

We must not underestimate the importance for libraries of sharing bibliographic information. And we have to protect this quality of the MARC record, in part by recognizing that it is a container for bibliographic description and it should remain so.

Library Systems

Many of the problems that people seem to have with the MARC record are actually problems with library systems. Most complaints about the lack of flexibility are not referring to desired enhancement to the bibliographic description but to a need to do something entirely different within a library system. We find ourselves in a situation where we have only one system option, our integrated library system, and many things we would like to do. And like the saying "when all you have is a hammer, everything looks like a nail," when all you have is a library system, everything looks like it needs to be input into the MARC record.

The primary example of this is the development of the Community Information format. This is really a testament to the creativity of librarians. Having a system that could accept only MARC bibliographic records, they figured out a way to enter information about community resources into those records and create searchable community information files. It was a kluge, a clever kluge. What is astonishing to me today is that we took that kluge and made it a standard. Had this need come up for the first time today we would probably expect libraries to create searchable web-based databases.

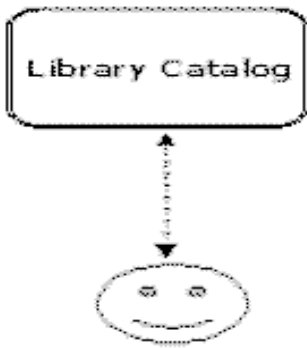
Unfortunately, the success of the Community Information format may have led us to believe that all of our data processing needs should be resolved using the MARC record. The Classification format is an example of data that probably should have used a different data structure but we seem unable to consider anything other than MARC. This isn't the fault of the record format, it is our own failing.

It's the System, Stupid

When we talk about the need to "modernize" the MARC format, I think we are really talking about the need to modernize our library systems. It is no longer acceptable to have a closed, contained library system that does not interact with the rest of the networked world. However, how we connect to that world is still evolving.

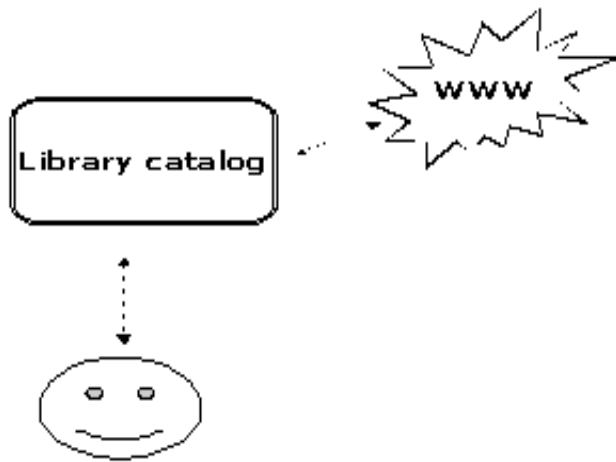
I have some slides that illustrate the kind of evolution that we have undergone so far, and where I think we need to go. You will see that the MARC format does not hinder this development because its niche of providing sharable bibliographic data for regularly published works is intact.

#1



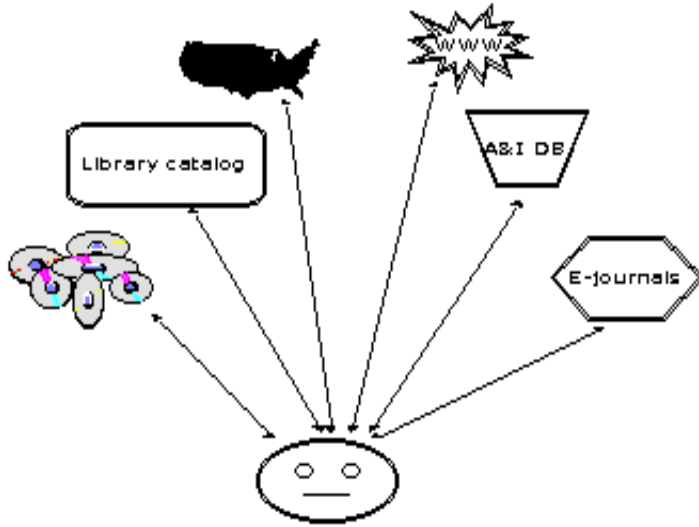
This is a diagram of where we started and this was our situation when the MARC record was first being used. The library of those days had walls and the catalog was a closed box. Note that the user is generally a happy user.

#2



In the early days of using the Web, we added links to networked resources from the online catalog. The problem is that the World Wide Web grew enormously and it quickly became impossible to try to filter the world of the web through the narrow gates of the catalog. Besides, our users are onto us: they know where the web is and that they don't have to go through the library catalog to find it. This model is quickly crumbling.

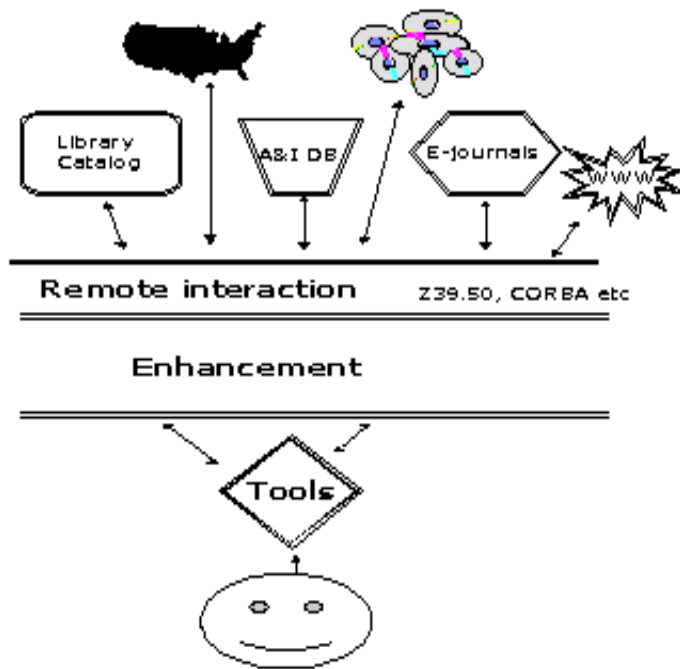
#3



In recognition that there is information outside of the library catalog, some of us have moved to this model, presenting our users with a host of different information resources. But note that my user here is less happy than in previous models. This model is complex and leaves the user confused and overwhelmed. It isn't easy to make sense of data in different formats, from different sources, etc.

In this model the library catalog has become just one of many finding aids for information. It can still be fed efficiently with national MARC record output. It is unlikely, however, that resources from other communities, including the GIS (geographical information system) community, vendor abstracting and indexing databases, social sciences data, image resources, will use the MARC record format. Each type of data has its own data processing needs. This doesn't mean that we can't search or display these data resources in a single system, however, and the next slide shows my "vision" for what the future system might look like.

#4



This diagram shows that we can think of all of these resources as a single complex system. The interesting action takes place in the layer called "enhancement". This is where a miracle occurs, and I call it a miracle because we don't yet know how to do this well. But the goal of this layer is to broadcast a search to a variety of information resources and gather the results into a form that will help the user make sense of the heterogeneous retrieved set. This can be done by creating sets of similar items, it could employ relevance ranking, or it could allow the user to select some items and find "like" items within the set.

But I'll go even further than this and say that I think we also need to rethink the user interface. It isn't much help to our users for us to throw a lot of data onto the screen and stop there. Users need tools for dealing with information. They need to store items for future reference, they need to know what they have already seen and what is new to them, they need to compare past results with current results. The beginnings of this toolbox might be in our current personal bibliographic systems, but our goal should be something on the order of Vannevar Bush's famous "Memex" machine -- the personal information space with the ability to link and describe and annotate. Only in our vision, the Memex is not the size of a desk, it probably fits neatly in the palm of your hand.

Conclusions

My conclusions are relatively simple:

- We must have a sharable, stable bibliographic record. Its record format is not of great importance, its content is key.
- No single record format is going to serve all of our data processing needs or all of the information communities that we will interface with.
- The problem, dear colleagues, is not with the record format but with our systems. We must turn our attention to system design if we want to move forward.

[©Karen Coyle, 2000](#)



This work is licensed under a [Creative Commons License](#).