

ARTICLE
EXCERPTED
FROM:

INFORMATION STANDARDS QUARTERLY
WINTER 2013 | VOL 25 | ISSUE 4 | ISSN 1041-0031

ISQ

MARC 21

BIBFRAME

TOPIC

EVOLUTION OF BIBLIOGRAPHIC DATA EXCHANGE

AUTHORITY

ARE CURRENT BIBLIOGRAPHIC
MODELS SUITABLE FOR
INTEGRATION WITH THE WEB?

A TRANSFORMATIVE OPPORTUNITY:
BIBFRAME AT GEORGE WASHINGTON
UNIVERSITY, AN EARLY EXPERIMENTER

INTERVIEW WITH GILDAS ILLIEN , BnF

SCHEMA BIB EXTEND

ANNOTATION

AUTHOR

NISO
How the information world
CONNECTS

OP

[OPINION]

A judgement formed about something;
a personal view, attitude, or appraisal



Paul Moss

PAUL MOSS

Replacing MARC: Where to Start

The MARC format for transmittal of bibliographic records has been an unparalleled success for interlibrary communication. This success, however, has also brought about a world that deals exclusively in MARC and is inherently bound by its limitations. Several efforts over the years have been made to break free of these perceived limitations, but these efforts often miss the crucial mis-step of MARC-like thinking: that a library interchange format should be the only way to ingest, expose, or build systems around bibliographic data. In order to begin a transition away from MARC, each function MARC serves should be examined independently and may be replaced by a different technology.

The MARC mindset

Over the years, usage of the MARC format has expanded into every facet of libraries and how they operate. For a library to ingest data from outside parties, it requests and even demands MARC records. When a library wishes to expose its collections whether it be in an exported file or via Z39.50 or other means, the basis for the exposed data is MARC. Many library application vendors have chosen to accept the limitations of MARC at the core of their applications by making it their fundamental data model. Everywhere you look in the library and its systems you can find some evidence of MARC data or cataloging rules applied to the data.

It is somewhat reasonable, given the expanse of uses of the MARC format

that any intended replacement of this format would assume that it must be a replacement suited to all of its use cases. This does not necessarily have to be the case though. Modern technologies very often espouse a very clear separation of concerns, such that each component may work together and even be separately improved without affecting the other.

Use cases

There are quite possibly too many different uses of MARC to cover them all in a brief article. I will focus on data exposure, core data model, and library data exchange.

➔ Data Exposure

For these purposes, I am using “exposure” to mean making library

data available to non-library services on the web. The goal of this kind of exposure is clear. Libraries want their users to find the research materials they seek wherever they are. It is a commonly accepted idea that many library users will go to Google or Wikipedia to begin their work. People will tweet links to interesting material to the world or share their research with colleagues on Facebook. This is the world of the web as it exists today and this is the world that the library must break into if it is going to be able to continue to offer services its users care about.

Interestingly, these services often have predefined ways of sharing metadata. Google has recently pushed its schema.org

initiative (along with Bing, Yandex, and Yahoo!). Twitter and Facebook have ways to create Cards, or small snippets of a page, that will be meaningful to users.

These are the de facto standards of the web. Data exposure to non-library services should follow these de facto standards. The library is not in a position to define its own standard for interoperability with those players, but rather should accept that the price of getting their materials in front of users is to do what is necessary to get where the users are. The systems that expose library data must include mechanisms to expose that data using these de facto web standards. Today it is schema.org; tomorrow it will be something else. Library data management and exposure systems must be prepared to follow the trends of the web.

➤ Core Data Model

The core data model seems to be largely where MARC replacement efforts are focused. The MARC record format is one intrinsically based on a model of collapsing all information pertaining to a particular book or other item into a single set of fields which make up a record. There are various reasons why this can be problematic. A study by Tom Delsey for the Library of Congress summarizes this challenge by saying:

In the past decade, the rapid evolution of digital information media and communications networks has posed significant challenges for the continued development and viability of the MARC format. Adapting the format to the demands of this new environment entails more than simple incremental enhancement to format specifications; it requires extensive re-examination of the underlying logical structure of the format and its application.

There is enough consensus in the industry that this must change, that adding my words to it feels like just piling on. Due to the prevalence of MARC formats inside of different facets of the library, making a wholesale change to the data model will be extremely difficult without separating the data model from the rest of the system(s) which use the data.

➤ Library Data Exchange

There is still a need to transmit data between libraries and/or library vendors. And there is still a need to improve upon the way that is done today. One of the problems here is that most providers of books and other materials to libraries do not use MARC as a fundamental data model. This presents problems for libraries to accept their data.

Take, for example, the recently developed KBART recommendations for interchange of electronic resource data. This set of recommendations can be loosely summed up as: Put your data in a spreadsheet and please use this set of column headers. It may be overly simplistic to describe the full richness of library cataloging, but it has a key feature: it does not, in any way, proscribe how to design the producer or the consumer applications. This benefit means that disparate systems, created for different purposes and with different technologies, may talk to one another.

A proposal

What would a standard for interchange of library data look like if it were *only* that? This, I think, is the proper purview of a MARC replacement at this stage. By removing the



People will tweet links to interesting material to the world or share their research with colleagues on Facebook. This is the world of the web as it exists today and this is the world that the library must break into if it is going to be able to continue to offer services its users care about.

CONTINUED »

requirement to be the future of bibliographic description for every purpose and focusing simply on the problem of moving metadata around, we may achieve a state which allows us to transition away from MARC as a representation of bibliographic data.

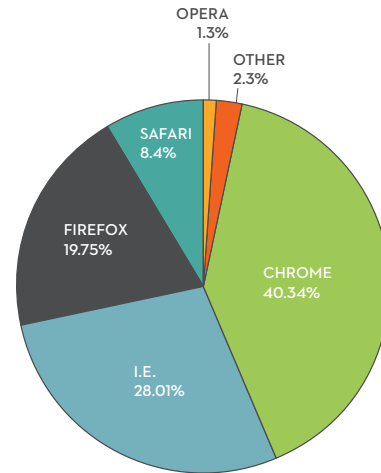
Consider a simple example: A list of books packed into a box and shipped to a single library. The current practice is for the library to obtain, either from the book vendor or third party service, a full MARC record describing these books at roughly the same time the box is received. This creates a coupling between the library system and the supplier of these records. If either party chooses to alter their end to support some alternate representation, then a translation between that format and MARC must occur. A small but very powerful change could be made to this transaction which breaks this coupling. If, instead of transmitting a MARC record, a simple list of identifiers (expressed as URIs) was passed, then the description of which books are in the box is no longer tied to the MARC format.¹ The identifiers may point back to a central service like WorldCat or to a service provided by the vendor, if available.

The difference between these two scenarios is subtle. By abstracting the format out of the equation for simple data interchange use cases, both parties may now be free to adjust their preferences for format in a semi-independent manner. Actually retrieving a usable format of a record or other carrier for including in a local catalog can be done through HTTP content negotiation or other mechanisms. (UnAPI is an example of a more complex mechanism.) This changes the expectations of each party from an agreed upon MARC requirement to one where each expects a range of different formats to be supported and a preferred one decided only at the time the record is required. This type of decoupling is very similar to what allows internet users to update their browsers on an irregular and, importantly, different schedule from the rest of the people browsing the internet.

Altering the mechanism for interchange of bibliographic data in this way could allow a new data model, such as proposed by the current state of BIBFRAME, to be adopted in parallel to existing models for transmitting MARC records. By decoupling the systems, the ecosystem of libraries and vendors and other parties can start to adopt new models alongside old without causing significant disruption.

Conclusion

It is a good thing that libraries are rethinking how we transmit data among ourselves. MARC is unquestionably an artifact of an earlier era. However, in replacing it libraries must understand that it isn't just the complexity of an old format which must be replaced but rather the reliance on a single data format for everything. We must accept that potential users of the library can easily get their needs met elsewhere and instead of fighting to get a library-specific standard supported by the Googles of the world, we should focus on making it so that they don't have to.



This type of decoupling is very similar to what allows internet users to update their browsers on an irregular and, importantly, different schedule from the rest of the people browsing the internet.

In the meantime, we can define our library specific data exchange format without the requirement of being the future representation of all bibliographic data everywhere by simply following the modern concept of a separation of concerns.

|OP| doi: 10.3789/isqv25n4.2013.03

PAUL MOSS (mosp@oclc.org) is the Product Manager of the WorldCat knowledge base with OCLC.

¹ This is a very specific description of one way to decouple these systems. It is, however, only an example. There are other possible implementations that may achieve this end.

Delsey, Tom. *Functional Analysis of the MARC 21 Bibliographic and Holdings Formats*. Library of Congress, January 4, 2002.

www.loc.gov/marc/marc-functional-analysis/source/analysis.pdf

Knowledge Base And Related Tools (KBART)

www.niso.org/workrooms/kbart

MARC

www.loc.gov/marc/

UnAPI

unapi.info/

WorldCat

www.worldcat.org

Z39.50 standard

www.niso.org/standards/z39-50-2003/



RELEVANT
LINKS