TOPIC

EVOLUTION OF BIBLIOGRAPHIC DATA EXCHANGE

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
INTRODUCTION FROM THE GUEST CONTENT EDITOR

It is with great pleasure that I introduce this issue on the Evolution of Bibliographic Data Exchange—full of thoughtful and informative articles describing new metadata initiatives across the library landscape. As guest content editor, I am fortunate to have a set of keen observers who have applied their knowledge, experience, and critical thinking skills to help bring us insight into the background and compelling forces for change. I am grateful to all of the authors for taking the time and effort to share their ideas with ISQ’s readers.

As is clear from the themes throughout the issue, the success of the web as a research tool has dramatically changed the library’s role in the exposure of library catalogs. As librarians have increasingly professionalized and improved the core mandates of selection, acquisition, preservation, and description of library collections, there has been a corresponding fracturing and loss of effectiveness in another of our responsibilities: exposure. The user has generally moved away from the library catalog as the tool used early in the research process—it is now used, if at all, as a source for availability or fulfillment in the last mile of the research process.

A companion theme throughout this issue is the widespread recognition that our current model for data exchange between library organizations has outlived its usefulness and is ripe for replacement with something with lower barriers to entry for library developers and partners.

Imperatives for Data Exchange

The rise of new metadata initiatives reflects the need to respond to this change and to increase our effectiveness in the exchange and management of library metadata.

As we proceed, we need a metadata model that allows us to achieve the following outcomes:

1. Effective exposure of library collections on the web
2. Efficient sharing of data between libraries and library organizations
3. Promotion of data quality to enable effective library workflows

The key word in the first item is “effective.” Our goal should be to find methods that will maximize the full disclosure of unique and commodity library collections on the web. That includes taking risks with the formats and methods that the web search engines prefer. It also suggests that we respond to the technical requirements of the web by aggregating data whenever possible and using canonical identifiers to make our assets efficiently identifiable in the linked data ecosystem.

The second item echoes the work that libraries have been quite good at over the last half-century—collaborating on standards that facilitate data sharing among library organizations. Our willingness to achieve near-universal adoption of data exchange standards is one of our greatest assets. We can leverage that collaborative spirit as we design the next generation of exchange standards and shed the inefficiencies and high barriers to entry of the current MARC 21 model.

Finally, the last imperative encourages us to look broadly at the data we manage (books, journals, collections, articles, etc.) and welcome new models for managing all of the data assets we care about. Data quality doesn’t just mean accuracy; it also means breadth and depth of data. Catalog librarians and library systems developers are comfortable managing the books and journals that information seekers use. It is well past time, though, to recognize that library users care about more than just books and journals. They also care about collections, parts of books, articles, and parts of articles including tables and charts. We must be willing to address the management of the metadata
describing those things and accept the need for new shared methods for exchanging this data. We should also accept the possibility of allowing social input to the management of library data. We shouldn’t automatically assume that it’s unacceptable for end users to make assertions about our metadata. Social input could improve the accuracy of both the metadata itself and the relationships between elements, such as manifestation clustering and collection memberships. We have a tremendous opportunity to build on our expert communities of practice and the vast potential of motivated end users.

Themes: “One size can’t fit all”

The articles in this issue all echo two common themes. First, that the mandate to effectively expose our data on the web calls for changes in the way we describe and manage that data. In our feature article, Lars Svensson from the German National Library reminds us that: “The bibliographic world still very much mirrors card catalogs. The problem is that the card display was not built around the concept of pivot points (e.g., authorities) but for sequential display organized according to certain criteria (title, headings)… To enable a better integration into modern, web-based workflows—be it the identification of a book for private reading or the construction of a bibliography for a PhD thesis—it is important that library (meta) data is not only available on the web, but really an integral part of it.” Jackie Shieh from the George Washington University echoes that reminder and writes: “In the last two decades, information professionals have been under pressure to remain relevant in the world of web data. Information professionals, in particular those who provide bibliographic description, have had to rethink and retrain themselves in the face of a new data service model for the records that they create and curate.” Richard Wallis (OCLC) endorses the call for change in his description of the Schema Bib Extend Working Group.

The second theme to emerge in this set of articles, and the one that I hope is a contribution to the dialog about library data exchange, is best expressed in the BnF Director Gildas Illien’s response to one of my interview questions: “In the past 40 years, be it with MARC or other formats such as Dublin Core, we have experienced the limitations of trying to answer all functional and community requirements with a single format or implementation scheme. **One size can’t fit all and doesn’t need to.** [Emphasis added.]...I would say we are ideally looking for a scenario where we could meet the joint requirements of a) internal metadata management, including the management of legacy data not only for descriptive purposes, but also for digitization, rights management, and long term preservation of collections; b) rich bibliographic data exchange services with no loss of granularity in description; and c) standard data exchange and exposure on the web the people and search engines use.”

Paul Moss from OCLC is more emphatic on this point when he writes: “The library is not in a position to define its own standard for interoperability with those [search engine] 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 suggestion that we will need multiple exchange models or layered exchange models for different use cases offers a pragmatic recommendation for the way forward. Our task is to develop these models in an orderly and efficient way. If we do, we can maximize the potential of libraries to provide the information needs of library users while avoiding ineffective and costly responses to current demands.

A Modest Hope

It is my hope that this set of thoughtful essays provides you with some insight into the landscape of new metadata initiatives. Indeed, it is my hope that this is a useful continuation of the dialog on how we can improve data exchange and that we see more recommendations and experiments inspired by the pragmatic and optimistic spirit of these authors. doi: 10.3789/isqv25no4.2013.01

Ted Fons | Executive Director, Data Services & WorldCat
Quality at OCLC