

3<sup>rd</sup> Meeting of the Working Group on the Future of Bibliographic  
Control

## Economics and Organization of Bibliographic Data

Library of Congress  
July 9, 2007

Speaker

Mary C. Little  
Director of Technical Services  
Queens Library  
mary.c.little@queenslibrary.org  
[www.queenslibrary.org](http://www.queenslibrary.org)



My thanks to Dr. Griffiths, Judi Nadler and the rest of the Working Group for their invitation and the opportunity to offer a public library's perspective. I also want to thank the previous speakers and contributors to this project. Their ideas and those of many others are reflected here. The work, observations and conversations so far offer valuable input. I'll be watching to see what the Library does with it.

As I set down notes for this morning I tried not to be daunted by the charge for this session:

to better understand

- the economic
  - human, technology, money
  - small scale and big picture
- organizational needs and challenges for
  - stakeholders
    - libraries; mine or yours
    - providers; vendors in the US and elsewhere
    - structures and standards creators
  - frameworks,
    - individual and collective
- creating, maintaining, sharing, supporting
- metadata, structures, and standards of

bibliographic control.

- Now
- foreseeable future
- imagined future

Wow, I don't think I can do that in 10 minutes.

I can talk about where my library is today, some of the things on the current landscape we find interesting and to some degree where I hope we can end up. At the same time the undercurrent in my mind keeps asking, is control of bibliographic data possible or even desirable in the imagined future.

You've been given a snapshot of Queens Library as landscape for my comments. 2.2 million, cultural and linguistic diversity, catalog in thirty-three

languages, circulated over 20 million items last year and the year before and have nearly 45,000 visitors a day.

The Working Group's first question is all about the economics and it's the key question for me. Everything we do has to fit into the resources at hand or those we can create. We're driven by mandates that evolve over time; most grow up from inside our own organizations. Today Queens Library catalogs every title that comes into the collection, that's new. A response to contemporary ILS systems. Though our ILS plans are still evolving, we still want to know what we own and give customers access to it. These are the management and access roles of bibliographic data expressed in previous sessions.

Expediency demands that more work be pushed outside the organization. Although talent runs deep in my department we don't know all sixty-five languages and can't spend much time with each of the 70,000 titles we cataloged last year. My role is balancing quality against several facets.

- Am I creating data I'm not likely to use during the record's lifetime?
- Are we the first to catalog it?
- Do I have enough data for the ILS to function today? What about tomorrow?
- How long will a type of material last in our collection?
- Is it likely to be replaced once worn or lost.
- Is there an outside data source that I can batch with a reasonably high product on the other side?
- Will my customers be able to find it in the catalog? In my ILS? On the shelf?

Today the English language stuff is easy:

Here's what I think most libraries are doing or are about to do. On-line selection in vendor websites feeds data to catalogers. Records are extracted in batch, massaged to identify those destined for local practice, further massaged and uploaded into the catalog. 90% of English language titles are identified in prepublication and in the catalog before the item arrives. Maintenance costs lay in touching records more than once. About half of our records receive some form of maintenance during their lifetime. Problems are reported by internal customers with resolution pushed high in the workflow.

Other than English? Vendors, whenever possible supply electronic order lists or reports. We negotiate data elements with increasing success. The best outcome to date and a model for other international providers, comes from a large Chinese bookseller in Queens. They provide monthly lists of high demand titles with enough data to catalog. Best of many possible worlds: customers have high demand titles when they're new, fewer single copies, data at low cost per

Queens Library  
Economics and Organization of Bibliographic Data  
July 9, 2007

title, no cataloging delays, vernacular access, batch processes, K level, vendor knows what to buy, library knows what to budget. We just added standing order series romance fiction – I'm very excited. I'd love it if more than one in a hundred providers could do this. The other end of the spectrum is no data at all, just the books. In lots of cases I'm happy with that. It's hard to buy current, popular Arabic books today, as a public library we don't find much.

All of this points to the trade-offs we've been making for years. Uncataloged collections; localized browsing collections were preferable to cataloging backlogs; they offered availability to low cost, high turnover materials that customers discovered by simply walking around. That presupposes your collection and your customers are in your library and are willing to take what they can find. That's less and less true today.

Our tradeoffs are all about quality and quantity. Maximizing external data for AV materials offers level 7 records, it meets customer's needs, tracks inventory and allows circulation of objects that will live for less than two years and not be replaced. Batch processing AV order records frees time to batch process pending order files for Chinese series fiction through some Perl programming. Access points to CD's are marginally impacted while a browsing collection never seen outside the local library gets visibility and at a level that can be put back into the collective.

The vendors feel our push. Our favored audio vendor just announced an OCLC partnership, success. I can move most work on audio titles to a clerical function with some degree of comfort. We pushed for over a year and then focused on another vendor. Collectively this is our big stick. We've not shown it much in the past, it's time. The big ILS vendors didn't see our demand for open source coming, it'll be curious to see how they respond to the mounting cries for stable core systems where we can hang best of breed components.

I'm wrong to think that the data structures that fit my needs will be "The" model. I'm lucky so far that we're ahead of the curve. Our international partner libraries are surprised at what we demand and what we do with it. If I can have K level for a dollar or two with vernacular, I'm there. That selection and materials management is easier and faster all the better.

In the distant future we need to be able to mine the data in other systems. Is that the eXtensible Catalog, impressive and really exciting? What data do I need to hold onto today that represents the foundation for this kind of interoperability?

The Group asked about commonalities. We capture the same data and maintain it locally for expanding formats. Our landscape, like yours, is a growing and complex linguistic environment. Supporting discovery in other languages and especially non-roman is the challenge.

In the near future we still hold local data, we've added access points and expanded service by accepting the costs of partnerships. I hope we know enough to be feeding that future collective. We've called on our communities to tag us. Especially valuable for me, I'll get to say things like "Tag me in Hindi, please". We're waiting for that ideal ILS and the promise of open source. Well OK, with open source lots of us aren't waiting. We're looking outside libraryland. We're watching projects like Pines and the eXtensible catalog. We demand faceted searching, give me FRBR. The landscape is forever changed.

The distant future is all about global underpinnings as the key to discovery. Today these proprietary systems are owned by the world's national libraries, private corporations, publishers, associations and other consortia. If this data were held in common, without local practice, ubiquitous and collectively managed we'd be focused on working across each other's systems rather than housing and massaging lots of data for every item we owned. The existing authority systems are such a valuable foundation to integrate social tagging and data mining into. Dropping the barriers changes the discussion on social tagging; it's gaps and timeliness. Unfettered use of authority sets from the whole spectrum of publishing and literature fills the gaps and starts us with a fairly level field.

Language will always be a key concern in Queens. When we're successful in opening systems and pulling in authority work from around the world, my needs become your needs. Vernacular. Unicode to work in all the public pieces of our ILS. A word cloud that discovers and returns in non-roman character sets. Easy to use input devices. Translation to lay over any page or serve as intermediary.

What role for Library of Congress in a game that is already afoot. LC has not been in the forefront of development, though this working group and the urgency of the project is a good sign that LC is indeed interested. The cries for open source solutions and interoperability demonstrate few concerns about the existence of historical data sets or new ways to gather input. In the foreseeable future standards and structures box up what we already have. My naive undercurrents tell me in the imagined future, our data isn't boxed.

Where will the current systems crisis take us?  
Feeding that "Where", is LC's opportunity today.

- Providing the content
- Working with other national libraries
- Exemplifying that new model, new paradigm, for openness

Thank you