Michael J. Fox, Minnesota Historical Society

Title: Raising sheep: helping lam(bs) grow up

Abstract: The concept of integrated public access to all the research collections of the Minnesota Historical Society was envisioned over 25 years ago. What seemed at first to be a straightforward intellectual exercise and managerial challenge has faced instead a range of unexpected technical, social, professional, and organizational obstacles even as the institution's understanding of the underlying objective has evolved. This is a story of the interplay of software, standards, professional norms, management, human nature, organizational systems, and customers. We have not yet arrived at the end of the tale where everyone lives happily ever after.

Presentation:

They say that where you stand has a lot to do with where you sit. That is, one's attitudes and ideas on a topic are strongly influenced by the environment in which you work. So let me begin my talk on raising sheep with some institutional context about my organization, as every good historian or archivist should.

There are several things you need to know about the Minnesota Historical Society as background for my story.

1. The Society is a state-wide educational institution that preserves the evidence of the past and tells the stories of Minnesota's people in order to provide life-long learning opportunities that that help people engage with their history.

2. We have many programs that contribute to that mission. Among them is a research collection of books, manuscripts, the state archives, sound and visual resources including art, and three-dimensional materials. We also operate 26 museums and historic sites around the state. Their geographic distribution, thematic range, and operational variations create some unique challenges. The network includes house museums, two military forts, a lighthouse, a farm, several battlefields, historic landscapes, and three purpose-built museums, one inside the ruin of a National Register flour mill.

3. While the Society's interpretive programs are critical to the successful fulfillment of its mission, they share the spotlight and resources with our research and publication activities that are peer functions both in scope and scale. This is a significant aspect of my story. This is not a museum with a supporting library and education programs, or vice versa. That reality has affected the path that we have trod.

4. In 1992, our staff and programs moved from three public locations and five storage facilities into a single, new building in Saint Paul, the Minnesota History Center. An important aspect of our vision for that move was to be the introduction of "central reference." This was to be a broader concept than perhaps its name implied: centralized
collections storage, centralized public access to all collections through a common reading room, a single reference staff to service that area, and an integrated catalog to all collections. The idea was one-stop shopping for the researcher who formerly had to navigate five separate reading rooms to consult the full span of our collections. The phrase "central reference" became our code name for what today we would describe as library, archives, and museum collaboration. Here, I want to focus particularly on that aspect of central reference that relates to the integration of the process of resource discovery across our many collections.

My tale today is an historic one in three parts. It recounts how we tried to make that happen, identifies the barriers to success we encountered (I hope that you may find them instructive or comfortingly familiar), and concludes with a description of where we are today.

Let me begin with a description of what we did: a brief history of trial and error towards realizing the vision of central reference. To clarify a complex series of events, I will divide this recitation of milestones in our efforts into four phases.

The first might be called "We've Got a Theory, Now How Do We Make It Happen?" When the idea of central reference was first discussed around 1986, integrated access meant nothing more that the simple idea of interfiling cards from a number of catalogs into a single set of drawers. It seemed brilliant and cost-effective solution but when I studied it more carefully when I arrived in 1987, it became clear that it was impractical given the variety of different cataloging practices for different collections over the years. These included inconsistent card formats and information layouts, differing descriptive content, and a variety of standard and non-standard vocabularies for access points (name authorities, topical headings, etc.). So when we moved in 1992, we simply lined up all the card catalogs in a row in a single new reading room. At least they were all in a single place though rather like a restaurant with a wine menu, a main menu, a table-top card with specialty drinks, a menu of daily specials, and a dessert menu.

Phase two could be characterized as "MARC: the Great Integrator."

In the late 1980s, we began using OCLC to catalog our published collections. With the introduction of the MARC Format for Archives and Manuscripts Control, we began to add MARC catalog records for our manuscripts and state archives collections into the RLIN database of the Research Libraries Group. Over time and after considerable retrospective conversion, these descriptions made their way into a statewide public catalog called MnPALS, where they may be found today. Today that network employs the Aleph library software.

Well, this worked so well that we said, "What the heck, let's try it for our material culture holdings as well." MARC for museums. We received a grant from NEH and cataloged about 12,000 items in the three-dimensional collections, extrapolating library and archival descriptive cataloging rules and MARC tagging convention. The entries were loaded into RLIN and exported into MnPALS.
When we were done, we looked around to see if any other museums were interested in this approach. We got a big yawn and a lot of responses of the "not-invented-here" variety. Some of them were very reasonable. Software developed for libraries has never been well suited to the back of the house operations like the management of acquisitions or conservation for anything other than print materials. But equally significant seemed to be the antipathy of many museums towards sharing information on holdings with the public. Research was for scholars; collections were for professionals to interpret through exhibits. And anyway, if we let people know what was in our collections, they'd just try to steal it. This was a dead end; NEH didn't want to extend the grant and take this experiment further.

Just as we became stuck with MARC, new technology models propelled us into phase three: "Best of Breed, or, Every Format Its Database."

This new wonderment, the web, appeared and with it the idea of disseminating information from multiple sources, including relational databases that would be separate from the standard OPAC. We switched our focus as to how we would improve access to the collections. Our emphasis now would be on the use of web-based technology for disseminating information rather than concentrating on improving access through the use of a single information store like MARC and the OPAC. We outsourced the creation of such a database that would handle the "unique" access and management requirements of our photo and art collections, or was it the unique requirements of the curator? I forget. The library director was adamant that we would only create a web-accessible catalog since no one would actually want to see the photos themselves online. But I was disobedient (and silent) until the first iteration of the unimaginatively named Visual Resources Database appeared with 1,000 images. It was an instant hit.

Then we tackled our extensive newspaper holdings. The library director, my boss, really disliked the messy business of serials cataloging, especially title changes, and who could blame her. So we built, you guessed it, another custom database for newspaper holdings. All along we realized that we needed a back of the house system for all collections to track operations like acquisitions, storage, and conservation and to provide public access for the museum collections. We acquired the EMu software and it has been our workhorse every since, taking on additional responsibilities as we go.

Most recently, archival finding aids in EAD-XML were converted to HTML files and delivered from the web site or links in the catalog.

That's where we were three years ago: multiple systems, separately accessed to discover collection metadata and surrogates. Let me stop the tape there (I'll bring you up to date later) and consider the issues that kept us from achieving our goal of central reference for those 20 years.
What were the underlying issues that affected the decisions had made over those years? What seemed at first to be a straightforward intellectual exercise and managerial challenge, encountered instead a range of unexpected technical, social, professional, and organizational obstacles, even as the institution's understanding of the underlying objective of the undertaking evolved. Let me group them into three categories.

First: management and organizational structure. Recall that our library and archives research program and our museum were roughly equivalent in terms of institutional significance and resources. There were three players whose cooperation would be essential for establishing of central reference: the executive director, the library and archives director, and the museum director who had responsibility for the three dimensional collections. Unfortunately the museum director was not really interested in integration. It has never been clear to me how much this was a matter of territoriality, personality, priorities, or professional disinterest in supporting research by the general public. Since the Society's director did not force the issue, nothing would ultimately happen with respect to the museum collections, except for what could be done under the radar by operational staff, until two successive deputy directors for museums moved on and administrative responsibility for all collections was centralized. Strong leadership is required to overcome resistance resulting from human nature or organizational silos.

Second: professional norms. We all understand that different sectors of the memory business and the collections they preserve and share really do have differences. Among these are disparate attitudes towards collections as public resources. While the web is minimizing these differences, I recall a time when librarians referred to unpublished materials as "non-books" and the cataloging rules treated manuscripts as a form of defective publication. Librarians and archivists have a strong tradition of helping the patron through reference and other reader services. But I do remember a prominent archivist arguing that our first responsibility was to the records, not the users. Many curators continue believe that their role as experts is to explain their collections to visitors through interpretive exhibits. And I can recall disagreeing not too long ago in a session at the Society of American Archivists annual meeting with a colleague who thought that archivists ought not to do exhibits as it compromised their neutrality toward the historic record.

These differences play out in disagreements as to what we ought to say about our collections and how we describe them to facilitate resource discovery. Is it to transcribe information from the work at hand, to analyze the contents of a collection, or to describe the physical characteristics of an object, say a hammer or a work of art? Is it context or content and of what nature? The puzzled researcher confronts these differences when the online display simultaneously presents information not only about books but also firearms and photos and fine art.

We thought we were doing something called cataloging but then we discovered metadata and the discussions became a whole lot more complex. The issues took a lot of time to work through but we learned an important lesson. We had to consider the larger professional contexts in which were worked and not just what might be done just within
our own institution. We had to deal with real differences across disciplines. From a 35,000 foot perspective, the road to integration seemed easily navigable but at ground level the path was rutted, muddy, and mine-strewn. We learned the hard way that resistant to change is especially powerful when human caution is backed by force of professional authority and long-standing traditions.

Third: technology. Frankly, even if we had resolved the social and organizational issues 15 years ago, we simply could not have moved forward without the evolving technologies that have finally made it possible to achieve our vision. Clearly the biggest technology leap came with the web. There were two reasons. Its ubiquity of service freed us from proprietary networks of limited scale. But more significantly, I believe, has been the impact of applications like Google as integrating technologies, ones that bring disparate information together with minimal overhead and a less rigid approach to the form and content of data. Many will view what I am will say next as heresy but it is informed by almost 40 years of labor in this vineyard. To the great consternation of those who thought that the world wanted a giant catalog of highly structured metadata, we have come to understand that the public is served reasonably well, and more sustainably so, by something less structured and centralized.

This brings me to the present and the fourth but hardly last phase of the story. I call this the "Federated Period." Our current approach to integrated access to our information resources is formed by four ideas.

- a changing vision of what integration involves
- the realization that the perfect is the enemy of the good and that simpler solutions can be effective and sustainable
- the need to limit our exposure to too many disparate platforms, also a sustainability issue, and
- a simpler, more flexible and extensible technical solution that begins to knot the pieces together simply

1. A changing vision of what integration involves. Central reference assumed that the user wanted seamless access to information about collections. Today, we realize both that the user wants integrated access to resources themselves and that these resources include a broader range of information sources that we create or acquire than we previously realized.

- Sometimes this information takes the form of pointers to collection materials through the mediation of structured metadata.
- Sometimes it is consists of direct access to the contents of specialized indexes that we acquire or create of ourselves: vital records, census data, military records, cemetery data, biographical and geographical information.
• Sometimes it is facilitated by the searching of the full-text text of primary source materials such as newspapers, journals, books, and manuscripts.
• Sometimes it requires the display of digital surrogates for all forms of our holdings.

As a leading cultural heritage institution in the region, we expanded our vision in another way to imagine a web portal that would integrate access not only to our own disparate collections but also to those of our history partners in Minnesota, North Dakota and South Dakota. A major grant from a regional foundation supported the work of solving the organizational and technical barriers to the creation of what is now known as the Great Rivers Network.

2. The single, all-encompassing catalog will never happen. This may be heresy to you. To my thinking, its last gasp may be the web 2.0 fantasy that we will be able to cross-map all the different metadata schemes to achieve a giant semantic web. In some ways, this seems like the ultimate cataloging Ponzi scheme. I may be wrong but who can explain to me where the resources will come from to realize such a goal?

3. The need to minimize the number of systems we try to support. Our newspaper holding are now in the library catalog where they belonged in the first place. The visual resources database is gone and its content migrated into EMu. We have been chastened to realize that building one's own customized systems requires substantial ongoing maintenance of the software and hardware platform, tasks better shared with other users either through commercial products or open source tools with community support.

4. Our new technical vision. It would utilize a suite of components that would allow us to ingest data from a variety of sources, especially customized databases, and handle their indexing, searching, and display while also performing real-time, federated searches of standardized systems like our OPAC. We initially acquired the IDOL software from Verity with the anticipation that a single tool could do both. After three years we abandoned it as too complex and too expensive for our needs and technical capacity.

In our new model, we employ a combination of Solr/Lucene to handle the indexing of information extracted from one set of data stores and web services to search and display information from those sources and to perform federated searching of our standardized systems such as our OPAC. With this simple and relatively low-tech approach, the single search box at the top of home web page returns the results of a search of the Society's web site itself, from collections catalogs in Aleph and EMu, and from various full-text sources and indexes of the types of data that I previously enumerated. Work continues on testing and refining the display of search results but the components are in place in a way that provides a great deal of flexibility. This aspect of the project is one area where the needs of our particular audience necessitates a different interface than that which might be desired by a more academically oriented institution, for example. A search of the Great Rivers Network dot com site expands the number of sources searched to
include several CONTENTdm sites and more custom databases at partner history organizations in Minnesota and North and South Dakota.

This is a different reality than the one we first envisioned but isn't that the way life goes. A consultant once told me that there are four possibilities when one starts a technology project: you can fail to achieve your results; you can achieve your goals and then discover that they are not actually what you actually needed; you can achieve your objectives and discover that they are just right; or you can end up with something new and better than you imagined. Our users and our vision for them of integrated resource discovery persist; technology comes and goes.