

Appropriate technology and the catablog

Robert S. Cox, UMass Amherst

In many respects, the Department of Special Collections and University Archives (SCUA) at UMass Amherst is typical of mid-sized archives throughout the United States: we labor under the profound sensation of being underfunded, understaffed, and overworked. Yet beginning in 2004, we sought physick for our ailments, and even with the relatively scant means at our disposal, the physick has begun to have its effect. Even as we dramatically increased the rate and quality of our acquisitions, we managed to increase the rate of processing, increase our exposure to the public, and -- ultimately -- increase collection use. Here, I would like to focus on only one part of the regimen we pursued, a part that might be viewed as a case study in appropriate technology, the application of tools and techniques that harmonize with the culture, resources, and technical capabilities of the communities in which they are applied. Rather than reach for the most advanced, most complex, or most expensive solutions, we selected those that made maximal use of our limited resources while establishing a sustainable, expandable base upon which to build.

The first stage in our course of physick was addressed to the psychosomatic aspects of our disease, and though critical to our success, it lies somewhat beyond the scope of this paper. Frustrated with what we perceive to be a tendency toward linearity of thought and dichotomization that forms the practical framework of much archival writing, we made an effort toward developing what we call a "field theory" of archival praxis, an integrative approach to theorizing the varied roles and functions of archival institutions in the broadest sense, based upon the recognition of the interconnections cross-cutting what are often viewed as discrete archival functions and stake holders. In briefest terms, a field theoretical approach considers all aspects of archival work as

organically linked, methodologically diverse, and connected to various publics of researchers, records creators, donors, administrators, and archival colleagues.

At a practical level, UMass suffered from the characteristic malady of the late twentieth century archival world, our version of the neuralgia of a century before, and just as fictional. Our symptoms ran from an unnatural pallor in our finding aids to flux in descriptive practice to outright hidden collections, and our sequelae included reduced patronage and a profusion of type I (false positive) and type II (false negative) errors on the parts of those seeking information. To be sure, we were far from the worst in what we had done, our predecessors left us with a poorly defined but otherwise sound foundation, but at the same time we were far from where we wished to be. Our greatest fortune, perhaps, was being immune to the scourge that afflicts large segments of the archival community, paralysis by grant: the persistent feeling that one cannot manage additional projects, particularly large ones, without a significant transfusion of money. Although we are glad to receive fiscal nostrums wherever they can be obtained, we were encouraged by the recognition that small forces acting consistently over a geological (or archival) time scale can have a profound impact, and secure in the knowledge that we can deliver very little.

Thus it was in 2005 that we launched a collection survey project adapted from the protocols pioneered by David Moltke-Hansen and Rachel Onuf at the Historical Society of Pennsylvania, with the triple goal of gaining comprehensive intellectual and physical control over our holdings, establishing priorities for collections management, and most importantly, exposing all of our collections to the public. Conducted almost single-handedly by Danielle Kovacs, our Head of Collections, over the course of two years, the survey included a hands-on examination of each of several hundred manuscript collections, whether processed or not. In each case, Kovacs established a standardized title and provided accurate estimates of extent (size) and chronological range, applied collection-level controlled vocabulary terms (LCSH), and generated a two-paragraph abstract that

corresponds to the abstract in an EAD record and that summarizes the background and scope and content notes, respectively. When necessary or desirable, she performed a level of preservation triage (e.g., replacing shipping containers, stabilizing fragile materials, and isolating problematic materials) and on occasion, when conditions were particularly favorable, she performed a degree of series-level organization. In keeping with the spirit of the survey methodology, we evaluated our collections for research potential, depth of documentation, and other factors as a means of generating priorities for future processing, and we compiled and maintained the results in a standalone MSAccess that we have used, in various iterations, for ten years.

So far, so good. Because access, even more than administrative oversight, was the critical motive for the survey, we immediately sought ways not only to make the results publicly available, regardless of the processing status of the collections, but to maximize their accessibility. After an informal canvas of researchers suggested that few (i.e. none) appeared to use our OPAC as a significant tool for discovery, and considering both the power of open web access and the limitations of the bibliographic utilities for handling manuscript and archival collections, we looked to our website and to "appropriate" ways of overcoming the limitations of flat html.¹

Our solution was to exploit the indexing capacity of blogging software to create a "catablog" -- a blog-powered integrated collection management system -- that we have dubbed UMarmot (UMass Archival Records, Manuscripts, and Other Things). After experimenting with several packages, we selected WordPress, which provides cheap, effective, flexible, and relatively comprehensive integration of manuscript and archival resources, and of course, it offers the potential for a degree of interactivity with researchers. While implementing a full-fledged content management system would conceivably offer greatly flexibility, the ease of development and

¹ Because it requires so little additional effort, we continue to generate collection-level MARC records for input into our OPAC, and thence into OCLC, but these are decidedly a secondary priority. Although we have not quantified the results, the overwhelming majority of questions we have received regarding our collections in the past four years has come either through researchers following footnotes or simple web searches.

maintenance provided by WordPress has proven a tremendous advantage in a department in which we have had (until recently) no employees dedicated to technical matters. In building UMarmot, we relied upon the library's technology staff only once -- for installing WordPress on our web server -- and we could easily have forgone that. All other work in designing, developing, and populating the site was undertaken by two staff members in our small department in the course of their normal duties.

The structure of each catablog entry is quite elementary, with the content exported directly from the high-level descriptive metadata generated during our survey or, for processed collections, culled from the EAD record. In some cases, we augment these basic records with the inclusion of relevant graphics (for visual interest) and by providing links to a variety of external resources and complete EAD-encoded collection descriptions.² Although WordPress makes available a large number of stylesheets generated by the community of users, many of which make for attractive sites, we decided to invest in adding our own css coding, both to enhance the visual clarity of our catablog entries and to demarcate discrete elements of the record so as to facilitate translation into Dublin Core or other metadata standards, should we wish to export from the blog.³

Most importantly, we make use of "categories" and "tags," in the language of WordPress, which are essentially controlled vocabulary terms (typically subjects) that are compiled and indexed to collocate blog entries. In UMarmot, the "categories" correspond to the major collecting areas defined in our collection policy, rendered in the terms most frequently employed by our researchers, while the "tags" (which are indexed separately from the categories) provide other fields of potential interest, including genre and (for archival collections) functional role and departmental affiliation.⁴

² We serve approximately 400 EAD finding aids from our website in both static html and pdf, mirroring these through the website of the Five Colleges Archival Consortium, where the html is generated on the fly.

³ Practically speaking, we use our MSAccess database as the platform from which the export, rather than the blog, per se.

⁴ Though not currently displayed as such, the categories are hierarchically structured to reflect coordinate relations between collecting areas.

Since the conclusion of our survey, we have begun to enter records for new collections at the point of arrival, investing a small amount of time to enhance our accession records to make them suitable for public consumption. Now that the process has become routine, we typically make information about new collections available to the public within one to two weeks of arrival.

The resulting catablog is a fairly intuitive interface in which researchers are able to explore collections in layered fashion: any single collection can be located in several different ways, reflecting different research styles, including free-text searches across the entire corpus, alphabetical browsing by collection title, selection by generic category (subject content), or by using a variety of mediated approaches that provide an interpretive gloss, such as exhibits, "featured collections," or research guides, all of which are integrated into the catablog. The WordPress calendar, which we have not implemented, offers the further possibility of viewing collections by date of entry, so, for example, researchers can efficiently locate recent acquisitions or recently processed collections.

For all the ease of development and maintenance, blogging software does have some obvious limitations as a collection management system. Search options are rudimentary, complex collections are not particularly well served, and it takes some planning and even more manipulation to translate records into other environments. At this stage, we have come to the tentative conclusion that the benefits of "Web 2.0" technologies are largely outweighed by their hype. In larger institutions, particularly those with highly focused collections and a high throughput in acquisition and processing, enabling researchers to subscribe to an rss feed might have a limited appeal, while in other contexts the ability to cull rss feeds from other sources might prove desirable for some institutions (e.g. those embedded in consortia or focused communities of interest).

Our experience with user generated commenting, however, has been decidedly mixed. Disregarding the ubiquitous spam, user comments have ranged from unimaginative to unhelpful, veering on the positive side from the occasional reference inquiry to requests to purchase materials

and to comments that in any universe might charitably be termed off topic. We have not yet enabled public editing of our pages, nor are we entirely convinced of the benefits of social tagging and user-generated metadata, as we continue to labor under the impression that it is our professional obligation, and our obligation to our donors, researchers, and administration to provide thorough descriptions of our holdings.

In many respects, our survey might be seen in the context of the recent trend in archival circles toward so-called "minimal processing," but we depart from that trend -- at least as it is practiced by many of our peers -- in significant ways. Our field theoretical approach to conceptualizing the process of processing places a rigorous insistence upon viewing the survey within a long trajectory. The minimal descriptions that form the basis of UMarmot records ("pre-descriptions," in our terminology) are considered little more than temporary placeholders, and even the full EAD finding aids ("descriptions") are considered ephemeral and open to revision ("post-description"). Our finding aids, in other words, are intended to be contingent and dynamic, changing in response to historiographic and cultural currents, and we feel strongly that many processed collections are as inaccessible as any unprocessed collection.

More importantly, we choose to emphasize *maximal*, rather than minimal description, aiming to produce descriptions of our collections as thick as we can possibly deliver, taking into account the strictures imposed upon us by staffing and funding. While we take this to be merely a semantic distinction from the original intent of "minimal description," we are concerned that as minimal standards are becoming institutionalized, they are becoming ends, rather than means, and cutting corners has come to replace care, context, and archival discretion. Put simply, we do not aspire to the minimum, and considering the full range of researchers and publics who might benefit from our collections, we feel it incumbent upon us to recognize that just as some access is better than none, so too is richer, better contextualized, and better informed access better still.

Sources

Elizabeth Yakel, Seth Shaw, and Polly Reynolds, "Creating the next generation finding aids," *D-Lib Magazine* 13:5/6 (2007). <http://www.dlib.org/dlib/may07/yakel/05yakel.html>.

Krause, M.G. and Yakel, E. "Interaction in virtual archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid." *The American Archivist* 70:2 (2007): 282-314.

Max J. Evans, "Archives of the people, by the people, for the people," *American Archivist* 70:2 (2007): 387-400.