

**Computers in Libraries Vol 24 (No. 1), January 2004**

## **Scouting Upstream for Pre-Publication Information**

**By Terence K. Huwe**

The more ambitious and powerful our digital utilities become, the more important it is for digital librarians to play locational and custodial roles. What's more, it has become far more important for us to have a voice in the very early formative stages of content creation, both in published form and as datasets. A recent experience I had in Sacramento with government information, and its sometime uncertain future, confirmed how critical and important it is to engage in digital library development—before the product is completed.

I recently joined a state advisory board that includes lots of economists and policymakers, but I was the first academic librarian to join. During a recent meeting, the head of a public-private venture that publishes occupational information gave an interesting presentation on their print-plus-PDF products, which focused on county-level data.

The publication, in existence since 1999, was threatened with closure due to shortfalls in the California state budget. Unfortunately, publishing county-level datasets was becoming too expensive. Yet the program had conducted a survey of its users, which are predominantly small businesses. The survey confirmed that there was a strong desire for county-level data, and less interest in metropolitan statistical regions (MSAs) or state level aggregations. The information need was a local one; the resource, which was as good as anything of its type I have ever seen, was at risk.

I was seated next to the presenter, who was a very nice (if worried) fellow. He indicated that his program has accumulated a data set of occupational and wage information for the five years they have published their report. My librarian "ears" perked up at hearing that. What plans did they have for the data? Would it be "persistently" available? I know researchers who would be very excited to get their hands on these multi-year data. But there were no plans yet for future activities. It reminded me that on several occasions since I became a labor specialist, I had learned of valuable data sets that were either lost or about to be lost—and they had no advocates or acquirers in sight.

**Join Advisory Boards to Save Stuff**

Digital librarians have a new “bulleted item” in their job descriptions, whether they know it or not: saving stuff. The problem is, there is so much “stuff” that difficult choices need to be made. And, as this example illustrates, the point at which stuff must be saved has moved further upstream, i.e., prior to the point of formal publication.

Of course, there are many success stories. The California Digital Library’s “Counting California” archive of state data is ambitious and user friendly, but it doesn’t include this series and many other datasets that lie undiscovered (<http://countingcalifornia.cdlib.org/>). I had a lunch date the same day of this meeting, with a colleague who works at UC Davis, and we discussed this publication. Interestingly, it’s the good ol’ campus OPAC that gives the best access to this title, which is held both at UC Davis and UC Berkeley, with e-links to the state URL included in the record.

It strikes me as significant that the radically evolving OPACs we use so heavily not only provide the best access, but continue to present knowledge organically and topically, with multiple “communities of practice” in mind. Academic libraries, with local collections and local communities of scholars, are the digital archivists as last resort.

The challenge—and it’s a big one—is to discover resources before it’s too late. Large-scale collaborative projects, like Counting California, have many strengths, but speed and agility isn’t always numero uno. The sheer size of state data guided Counting California’s emphasis on the most frequently used resources as its core collection. It falls to special librarians, who have deep knowledge of discrete disciplines and information producing agencies, to follow a tandem course, and save valuable, specialized data before they are lost.

The challenge is greater when authorship is shared, too. My occupational series example is a joint venture, including local partners in counties throughout California, central state agencies, Worker Investment Boards, and other government bodies. Who “owns” the data? If I had not been there for the board meeting, it would have been almost impossible to learn about the imminent changes until it was too late. As it stands, I am in a position to open a dialogue among many stakeholders, about the data’s future—because I joined the advisory board.

## **Datasets and OPACs**

Library catalogues have come a long way toward being a one-stop access points in the last ten years. Nonetheless, the best synergy for effective research still requires review of Web sites and data centers around the country, because the research process will take you wherever necessary. For this reason, our OPACs continue to require training, interpretation, and frequent referrals to other collections. OPACs are designed to manage print collections, and increasingly, digital files, but when it comes to data sets, they operate best as gateways. For this reason, special librarians often make it their business to keep abreast of changes in government.

Despite the growing popularity and power of e-links in catalogues, I continue to see a need for robust review and discovery of Web resources outside the OPAC. Every intellectual discipline needs extensive pathfinders and Web guides, and they need to be updated frequently. Even though the “semantic Web” is in development, the current Web has gotten a lot smarter, with a lot more references to solid content by reputable authors.

But sometimes it just falls to us to make a data rescue quickly, if all the seeming permanence of government Web files vanishes. To be ready to do so, we need to join advisory boards. I have found that education works both ways in such venues; in some ways it’s been more of a revelation for me to hear how government researchers do their work than it is for them to hear about my patrons’ needs.

## **No Board? Then Create One**

Even when there is no advisory board in existence, it may be possible to lobby for the creation of an advisory board. Government researchers recognize the challenge of staying in touch with their constituencies, and it may be possible to cultivate an inside contact to help you advance the idea. I am not referring to a library-focused advisory body; I am referring to an advisory body that goes in the other direction, where we, the digital experts, are giving our counsel where it is needed—further upstream in the publishing process.

## **Find the Data Managers Behind the Series**

Nicholson Baker famously made the case for archiving old newspapers, and even convinced a few people along the way (while upsetting lots of others). But do

data retain value in the same way? If so, do the producers know that they do? It was clear to me that in the case of my occupational data example, the people in charge had not considered the elasticity of the data, how it could be used in a wide variety of ways, and how they could partner with other institutions to preserve it. Most of all, the leaders of the program didn't have much in the way of contact with archivists or librarians. That's why I see an opportunity for librarians to bring their value points into sharp focus, by taking the time to join the advisory bodies that guide government agencies.

## **Two Strategies**

I see two strategies we can employ to extend collection development upstream from the OPAC. First, as I've suggested, join a board, or create a board. You don't have to be a library director to make a difference, either. Indeed, joining the professional life of your information producers is a career-building action. Moreover, in a digital era, it's no longer enough for information professionals to wait for publications, print or digital, to arrive and then be catalogued. We need to find ways to join in the dialogue about how data are formed and packaged, and how metadata can interface with the content. It's not hard to find social scientists and economists who grasp the importance of metadata standards; they make easy allies. Once we join them at the table, our professional culture begins to permeate theirs, strengthening the library's position.

Second, it's important to network with the line staff who handle data in information producing organizations. These include both managers and support staff—the ones who often know more about how to get things done than the bosses. This may be a nitty-gritty, getting-to-know-you, “cold-calling” process sometimes. That's good: it builds character, creates potential new users for your collection, and lays a foundation for political alliances that may help you in other arenas. Regardless of the politics of the agency, you are then in a position to know the status of the data at all times, and swoop in for an acquisition when the time is ripe.

So, you might say that your digital wisdom is requested, upstream from the OPAC. By creating personal relationships with the people who are designing and publishing material, we can increase the odds that the data and knowledge we need to catalogue will arrive downstream in a robust format.