Missing Information and the Long Tail: How Distributed Collection Development Assures the Continued Relevance of Libraries

Amy Brunvand
University of Utah Marriott Library, amy.brunvand@library.utah.edu

Follow this and additional works at: http://docs.lib.purdue.edu/atg
Part of the Library and Information Science Commons

Recommended Citation
DOI: http://dx.doi.org/10.7771/2380-176X.4917

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Missing Information and the Long Tail: How Distributed Collection Development Assures the Continued Relevance of Libraries

by Amy Brunvand (Librarian, Government Documents & Microforms University of Utah Marriott Library; Phone: 801-581-8394) <amy.brunvand@library.utah.edu>

Working as a government documents librarian, I encounter online information that I just know someone is going to want in the future, but I feel frustrated because there is no obvious way for me to get copies into our collection. What I want is a strategy for digital documents that is as straightforward as the ones we use for books and journals. I want to make a request and have a system in place to acquire and catalog electronic documents for me. This is not just laziness or egotism—I want every librarian to be doing the same thing because in my view, the future of libraries may depend on it.

Missing Information is a Problem

The Web has not solved the problem of missing information. It has merely changed the nature of the problem. Instead of going missing all at once, information decays slowly over time. Outdated hyperlinks lead to dead ends. Less used documents sink ever downward into deep Web archives, while popular ones bob to the surface of relevancy rankings. Advocacy groups and think tanks put their latest action item on the front page and let older campaigns fade away. Web documents may be altered to express whatever political ideology is currently in vogue. Occasionally, a new Webmaster re-writes an entire Website obliterating older data.

The missing information of tomorrow consists of things that are easy to find today, but ten years from now, who knows what will still be there?

Librarians have not ignored these issues. For example, The Web-at-Risk Project, a collaborative project of California Digital Libraries, University of North Texas, and New York University, is developing a Web crawler to build what is essentially a topical electronic vertical file of Web-published materials. “New Mexico Plus,” an ingenious project at the New Mexico State Library, gathers a locally relevant collection by linking articles from local newspapers to online information sources.

It is dismaying that these projects are still considered experimental. Collecting a digital document should be as normal as ordering a book, yet digital collections are still treated as special collections. The selection/acquisition/cataloging system for print documents has yet to be fully translated to cope with born digital documents. While it is perfectly clear what legal rights one has to buy and lend a hardcopy, it is far less cut-and-dried what may be done with a digital copy. As a result, most digital library projects concentrate on publications such as government documents or historic publications in the public domain.

The Age of Missing Information

In an electronic environment, there is a temptation for librarians to try to collect everything, which can become overwhelming. On the other hand, it is tempting to collect nothing because someone else is already doing it. The way through the dilemma is to revive and reinvent the profession of bibliographers and to develop collection strategies that take full advantage of the local knowledge and diverse interests of a multitude of librarians.

In 1992, Bill McKibben published an essay in the New Yorker titled “What’s On?” which described the dearth of local information on cable TV. The essay, republished in book form as “The Age of Missing Information,” offered the profound

Information Sources about the Legacy Parkway dispute in Utah:

SAFE

- UBEt et. al. vs. United States DoT
- Articles and editorials in statewide papers (Deseret News and Salt Lake Tribune)
- Articles and editorials in regional papers (Standard Examiner)
- Court cases and major newspapers are safely stored in various commonly available databases. No worries here.

Regional newspapers, on the other hand, may be a problem. The main newspaper for the northern Utah region, the Standard Examiner, is not included in any major newspaper database, and you need a password to search the archive on the paper’s own Website. Paper is currently online, but future researchers could be reduced to scrolling through microfilm.

THREATENED

- State Documents (Utah Legislature, 2005 2nd Special Session; UDoT news releases)
- Federal Documents (Draft EIS, Final EIS, Supplemental EIS…)

For local, state and federal government information, availability depends entirely on public funding and political will, especially since the EIS documents for this project are examples of “fugitive documents” which were never included in the Government Printing Office depository library program. The Utah State Library is working on a database of Utah documents which may eventually move some such items to the “SAFE” list.

ENDANGERED

- Legacy Parkway Settlement Agreement
- Advocacy groups (Utah Sierra Club Smart Growth Alternative, UBeT…)
- Publications by Mayor Rocky Anderson

On the Utah Legislature Website, you can read the text of a bill approving the Legacy Parkway Settlement Agreement, but not the text of the agreement itself which was published on the Web as an attachment to a UDoT news release.

The Utah Sierra Club, like any advocacy group, promotes current action items, but the less timely this issue becomes, the less likely it is that information will persist on the Club’s Webpage. UBeT is a coalition that formed specifically to oppose the Legacy Highway. The organization itself may not persist once the issue has resolved, and if UBeT disappears, what happens to the information on the Website?

Politicians are ephemeral by nature. Rocky Anderson is not going to be mayor forever, and when he leaves office his Website and publications are likely to disappear as well.
not only a very large mass of documentation (see chart, sidebar, whatever) but also real-life results. This is a highly focused regional issue of the type that generates valuable long tail collections. The publication trail identifies where relevant information resides and whether or not it is likely to be safe for the future. Briefly, the Final Environmental Impact Statement which the U.S. Dept. of Transportation prepared for a highway project in Utah was contested in court by Salt Lake City mayor Rocky Anderson and two citizen advocacy groups: Utahns for Better Transportation (UBet) and the Sierra Club Utah Chapter. The plaintiffs won their case and hired their own consultant to write an alternative “Smart Growth” plan for the highway. The Utah Legislature then called a special session to discuss the highway, and all parties agreed to a compromise.

It is obvious that from a library standpoint only part of the information stream is well represented, and this is not even a comprehensive list. The Legacy Parkway was a regional controversy for nearly ten years and there were associated Webpages, blogs, emails, discussions, speeches and so on.

Even in the new information environment there are core information streams and there are ephemera. There is material which will naturally be saved in library collections and there is material which will be lost. It’s just that the core has broadened.

If each library can truly identify and represent its own information niche digitally then the community of libraries will build a long tail collection with which no other information source can compete.

Endnotes

1. “Long tail” also refers to a type of statistical distribution with a few high-frequency values and many low-frequency values. A graph of the data looks like a steep ski slope with a long, flat run-out at the end.

Is Access to Government ...

Case Study: Utah’s Legacy Parkway

I’ll use a case study of the Legacy Parkway freeway construction in Utah, which neatly shows the interaction of government planning, journalism and citizen advocacy in generating...