Biz of Acq-Where's Web?

Aline Soules  
*California State University*, soulesae@gmail.com

Michelle Flinchbaugh  
*University of Maryland Baltimore County*, flinchba@umbc.edu

Follow this and additional works at: https://docs.lib.purdue.edu/atg  
Part of the [Library and Information Science Commons](https://docs.lib.purdue.edu/atg)

**Recommended Citation**  
DOI: https://doi.org/10.7771/2380-176X.5660

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.
Biz of Acq — Where’s Web?

by Aline Soules (Librarian and Professor, University Libraries, California State University, East Bay, 25800 Carlos Bee Blvd., Hayward, CA 94542; Phone: 510-885-4596; Fax: 510-885-4209) <soulesae@gmail.com>

Column Editor: Michelle Flinchaugh (Acquisitions Librarian, Albin O. Kuhn Library, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250; Phone: 410-455-6754; Fax: 410-455-1598) <flincha@umbc.edu>

Column Editor’s Note: Aline Soules

This interesting article on considerations for handling free Web resources in the library. I was delighted to receive this article because the decision to add, or not add, free Web resources to the library collection is one that I’ve been dealing with in the course of my ordinary acquisitions work for some time. At my institution, acquisitions began receiving occasional requests for items available for free on the Web, and lacking a policy or philosophy, we began determining on a case-by-case basis what to do with those requests. If we think the item requested would likely remain on the Web on a long-term basis, we catalog the site. In other instances, where we suspect the material will be removed, such as television news streaming video of political debates, we recommend that the faculty member add the resources to courseware instead. Eventually we will need to develop a carefully considered collection development policy for free materials, and this article brings up many of the issues that we’ll need to discuss. — MF

Introduction

We’re all familiar with games like Where’s Waldo or I Spy. The only reason it’s tricky to find what you’re looking for is that it’s obscured by everything around it. Similarly, it’s easy to enter a term on the Open Web and get a million results, but will you find what you’re looking for?

If you want facts, the answer is probably yes, but what about searching for Websites relevant to your scholarly interests without also retrieving far too many extraneous results? The ability to enter your own natural language terms and get back exactly what you want is far from perfected yet. Integration with other types of materials is possible if the Open Web is included in federated searching tools and you search through those tools, but not guaranteed if a search is initiated on the Open Web. If you do find something useful, what is the best way to save it? Simply, can the Open Web really give you what you want?

- indexed in commercial databases
- cataloged in online catalogs
- gathered on subject guide pages by librarians and in various forms by users themselves
- developed by groups with like interests
- referenced on Facebook, Twitter, and YouTube
- affiliated with specific books, such as supplements to textbooks
- replicated in print and electronic books
- linked from anywhere to anywhere

The good part about this is that these practices follow Weinberger’s principle to “put each leaf on as many branches as possible,” for the theory being that these sites will be more easily and more often found.

Searching for a Web Site

But is this always the case? The Walt Whitman Archive, for example, is a noted Website of primary sources (it even provides access to a recording of what is thought to be Whitman’s voice). It is accessible through the Open Web, but also indexed in the MLA International Bibliography, linked from numerous other Websites, and included in some libraries’ online catalogs. If you know to look for “The Walt Whitman Archive,” you’ll find it easily through any of these sources; however, unless you know that specific title, you are more likely to search “Walt Whitman” or even “Whitman,” as so many of my students do. If you enter Walt Whitman as a search phrase in these various sources, what do you find?

I conducted searches in these sources on July 18, 2010.

The Archive showed up as the second entry in a Google search resulting in about 2,650,000 results. It appeared right after the Wikipedia entry (which will users pick first?). In Google Scholar, it did not appear in the first three screens (30 entries) of the 43,600 results. Will users go past three screens? As far as three screens?

In the MLA International Bibliography (searched on the EBSCOhost platform), it did not appear in the first three screens (30 entries) of the 3,835 “all results.” On looking at the “source types” list on the left navigation bar, I had to click “more” before the category “Websites” appeared as an option in this particular search. On clicking “Websites,” the Whitman Archive was the first entry of 21 results culled from the original 3,835. Having this particular archive (there are others) indexed in these various places, therefore, doesn’t necessarily increase the likelihood that users will find it.

As for the catalog, California State University, East Bay does not have an entry for the Archive, but a search in the LINK+ consortium of 51 academic and public libraries to which Cal State East Bay belongs yielded 2,795 results and it was not in the first screen of 50 results, although a search for “Walt Whitman Archive” yielded 5 results, of which this specific archive was the second one listed. In other words, this Archive may be included in all these sources, but it is quickly buried among the large results set of a simple author search, even if the search includes both the first and last names of the author. This leads to the question of whether it is helpful to include the Archive in all these sources or whether including it causes the user to be confronted with too many results overall, with Websites (and potentially other useful items) getting lost in the overall list. Further, in spite of the fact that the Wikipedia entry came up first on an Open Web search, the Archive was prominently displayed as the second choice and, in the end, proved to be the most effective in drawing this source to my attention.

To Collect or Not to Collect?

The first decision libraries face is whether they will collect Websites at all. If I were to base such a decision on my own experience with the Whitman Archive, the answer would be quite clear, but a real study of many Websites is required to understand the full impact of such searches. If librarians do decide to collect Websites, what policies should they establish for collecting them and including them in various locations? Sites may be collected for their subject suitability, but there are other considerations. For example, if the general criteria for selection include the anticipated number of users, how is that assessed for a Website before it is collected? Also, once collected, how can currency be assessed long term? Although a book has a built-in aging process, a Website’s aging process is predictable. Not so with a Website. Furthermore, while a Website like the Whitman Archive doesn’t have a cost attached to it (donations are accepted), there are acquisitions and budgetary implications in terms of staff time, record maintenance, etc. If a decision is made to create a record for a Website, is it better to catalog it or provide a link through selectors’ discipline-specific subject guides or provide it as a separate set of library Web pages?

Library Created Web Sites

In addition to collecting some Websites, librarians are putting their own collections of information on Websites, which are subsequently cataloged. One example of an individual contribution is Bailey’s Scholarly Electronic Publishing Bibliography, a free eBook that includes over 3,600 references to...
The indispensable reference for all who work with words

The Chicago Manual of Style
SIXTEENTH EDITION

“The new go-to publishing manual is here! Responding to the demands made of publishing professionals in the digital age, the 16th edition continues to provide information for word lovers but has enhanced its guidelines to include electronic workflow and processes. . . . This venerable bible also has a new pastel blue bookjacket to alert users to the style, usage, and technology updates within, though the familiar orange cloth lies below. A worthy, welcome addition to every library collection as well as professional wordsmiths and educated readers of all persuasions.”

Library Journal

1008 pp., 4 halftones, 54 line drawings, 15 figures, 12 tables, 1 musical example

ISBN: 978-0-226-10420-1 • Cloth $65.00

The Chicago Manual of Style Online is available to libraries by subscription. For rates and information, go to www.chicagomanualofstyle.org. Free institutional trials are available.

The University of Chicago Press

against the grain / november 2010

<http://www.against-the-grain.com> 67
appropriately for our situation? Will we even find what we're looking for? Will we have so many results that we don't go past the first few screens and, despite "relevance" ranking, lose valuable possibilities or not find what we're looking for at all?

**Discovery**

The answer may now lie in "discovery" services. The problem with commercial options, however, is that in order to ensure that users stay inside their discovery product, vendors include many pieces of information that might be better if excluded, e.g., Websites. In the California State University system, however, David Walker has developed an open source interface to Metalib by ExLibris. The possibilities can be seen in presentations Walker has made available on the Web. While the traditional concept of a catalog as the centerpiece of library holdings has been shifting for some time, this really makes the catalog one piece in the puzzle and a small one at that. A library could include the Open Web as a search option, along with the catalog, databases, or any other source, all on an equal footing. The implication is that Websites do not have to be collected, cataloged, or acquired in separate transactions. They can simply be searched along with everything else.

The positive aspect of this idea is that librarians can sidestep difficult decisions about what Websites to collect or not collect for inclusion in the catalog or subject guides. The Open Web results would show up as part of the results set. The negative aspect of this is, once again, the size of the results set, unless the user inserted specific and, probably, multiple terms. In the discussion portion of a Webinar on Xerxes 2.0, Walker indicated that users are expected to "interface" with their results. That suggests that they will refine their searches, think of other terms, add terms, etc. But do they? Users, particularly the students with whom I work, often use terms that are far too broad, e.g., "I have to write a paper on women," after which they enter "women" as their search term. Users are also used to looking at the first one or two screens of results and either picking something or deciding that there's nothing there, if the results don't mesh exactly with what they think they need.

**Full Circle**

This type of behavior is what leads librarians to "collect" Web information in the first place, whether in their catalogs, on their subject pages, or elsewhere, in order to try to help users find key resources. If the searches for the Whitman Archive are any indication, however, the most successful search for the Archive was actually on the Open Web, not in the library catalogs or even Google Scholar. Perhaps it's time to re-think the idea of collecting Websites altogether. This brings us back, full circle, to the question of whether some Websites should or should not be collected and drawn to users' attention as part of the collection of scholarly materials available.

Another alternative is to create a multi-tiered approach and make that available through the discovery tool. Currently, users can specify in database searches whether they want to search a specific type of material, such as scholarly or peer-reviewed items only or just articles. This is not presented preferentially, however. It's simply a choice of equal value. Students who are told by their professors that they must have three scholarly articles, for example, will introduce that limit to their searches, if they are aware of it, in order to meet the requirement. An alternate possibility is to structure a discovery tool to search categories chosen by librarians first, then, in a second tier, offer them the Open Web. Some might argue that this is a role that librarians shouldn't play, but we already play that role by what we collect in the first place.

**Conclusion**

The Open Web has useful information. To find it, we need the following:

- continued development of discovery tools, which are in their early stages
- willingness to recognize the value of the Open Web to the point where we include it in discovery tools
- increased attempts to help users understand the meaning of their results list, and
- concerted efforts to encourage users to interface with and manipulate results to better effect rather than just picking the first items on the list.

That is where we should put our efforts rather than trying to compensate for this transitional period by collecting Websites, creating records, and expanding library Web pages, all of which make more work for ourselves than we can ultimately manage.

---

**Rumors from page 56**

Porter's famous play. It seemed appropriate in the times we find ourselves in right now. And we have tried to stir up the program this year in the spirit of *Anything Goes!* We have Tech Talks on Saturday morning which will offer publishers and vendors 30 minutes each to demonstrate their most innovative products during concurrent sessions. There will even be food to keep us alert! The fabulous Beth Bernhardt and Leah Hinds (could call them the Minnesota Twins except neither one of them is from Minnesota, oh well) have cooked up the idea of Happy Hour concurrent sessions this year! Sounds like a unique idea, yes?

**Endnotes**


11. Walker, David. Xerxes 2.0. Webinar. July 28, 2010. This Webinar was given to California State University users and is not available on the Open Web at the time of writing.

---

<http://www.against-the-grain.com>