November 2005

Deep Linking -- Beyond Journal Articles

Chuck Hamaker

University of North Carolina at Charlotte, cahamke@email.uncc.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Hamaker, Chuck (2005) "Deep Linking -- Beyond Journal Articles," *Against the Grain*: Vol. 17: Iss. 5, Article 7.

DOI: <https://doi.org/10.7771/2380-176X.4549>

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.

Maximizing the Value ... from page 20

until the last minute to begin a research paper, can rarely interest the student in the complete range of e-serials that cover the topic. As a result, every library has paid for electronic databases, ejournals, and other resources that do not get the use they deserve. Often these end up being cancelled because the cost per use is too high to be affordable (Moore-Jansen, Williams, and Dadashzadeh 2001; Tenopir 2001, Hiott 2004).

Bibliographers, collection management librarians, and library instructors could work together in a process aimed at increasing the student's knowledge of the entire range of e-resources for the best academic performance. For example, at regular intervals, bibliographers and library liaisons could review the usage statistics for e-serials and identify likely candidates for cancellation. Library instructors would then take the lead in demonstrating one or more of these under-utilized serials in their classes. With usage data from before and after the instructors' intervention, it should be possible to detect if there was a positive or negative effect on students' use of the targeted e-serials.

If there are generally positive results with this technique then libraries could regularly identify their lowest performers and give them wider publicity through instructional programs. If an e-serial's use fails to improve after this "probationary period", then collection managers may feel more confident that librarians are canceling products that are unsuitable for their users and not just unknown to them.

Students are noted for wanting to use only products providing full text (Tenopir 2001, Dilevko and Gottlieb 2002, Black 2005). Yet citation databases often provide the best resources for a particular research topic. Student preference for full-text resources have been attributed to multiple factors, including the convenience of electronic full text and to the common situation where photocopying print materials costs the student directly but printing out electronic text does not (Black 2005). In addition, it is also well documented that students are very confused and frustrated by the process of moving from identifying a reference in one source to retrieving the full text for that reference from another source. (Black 2005, Labelle and Nicholson 2005).

If BI and information literacy programs are effective in making students more comfortable with the library research process, then usage statistics for citation databases will rise from previous levels. If so, libraries could again target low-performing citation databases for a little extra attention in instructional programs and use usage data to monitor the results.

Even relatively mundane activities such as system administration can benefit from input from usage statistics. Wake Forest University recently added a wireless component to its hardwired network system. The wireless system includes connections for guest, faculty and staff, and students. Guest users have no access

to networked printing and file sharing nor to the library's small collection of locally-mounted CD databases. In the first months of the dual system, large numbers of students were using the guest wireless system rather than the system intended for them. As a result, they were unable to access local databases. The only way that Reynolds Library personnel were alerted to the problem was by the sudden tripling and quadrupling of the usage statistics for those databases. Databases that previously averaged 10-12 sessions per month were abruptly showing 50 sessions per month, usually with a single IP address, which represented one user, attempting 2 and 3 login attempts within a period of 60-90 seconds. Once librarians began advising students of the need to use the student wireless for these databases, the usage statistics dropped significantly though not back to previous levels. The pattern of multiple attempts in a short time is still apparent, indicating that not all students received the message.

The biggest challenge in maximizing the utility of e-serial usage data may be simply recognizing new opportunities for using this data. As new technologies appear they can exert both positive and negative effects on the use of e-serials. For example, as bibliographic management software such as **EndNote** or **Reference Manager** becomes more prevalent on campuses, users may respond by focusing on e-serials that work best with those systems. Libraries could monitor its e-serial usage data for such indications and develop appropriate responses. Developing useful statistics requires a significant investment in time, personnel, and money. Diligence and creativity in using e-serials statistics can enable a library to reap the full benefits of its investment in a statistics program.

Deep Linking — Beyond Journal Articles

by **Chuck Hamaker** (Associate University Librarian
Collections and Technical Services, Atkins Library, University
of North Carolina at Charlotte) <cahamake@email.uncc.edu>

Usability has become the single most critical function for e-resources. Usability must be improved dramatically for libraries and the electronic resources they support. Deep linking must become ubiquitous from catalogs, from Abstracting and Indexing services and from other identification tools for content beyond journal articles. Its lack is a major obstacle to the full use of networked electronic content.

For music CD's **Gracenote** formally known as **CDDB** working with the music industry worldwide has compiled metadata for consumer use on 3.6 million CDs and 46 million tracks. **Apple's iTunes** supports 30 second previews for everything and links at the artist, album, and track levels for commercial use. Some similar

metadata center or centers, some clearing house of stable syntax or permanent networked identifiers, much as **CrossRef** provides for journal articles, needs to exist for all networked e-content supported by libraries, vendors and publishers of e-content. **CDDB's** goal is to "enable consumers to better manage, enjoy and discover digital media." That sounds similar to the goals for the library, vendor and publisher community with networked digital content: discover, use, manage, and preserve.

Site and title level linking is currently the standard for library catalogs (**OPACS**). Such linking fails to support robust exploration and retrieval - how well systems answer queries and deliver content. Title level linking is inadequate compared with the deep linking available from **OpenURL** supported indexes and from famil-

Endnotes

- Black, Steve. 2005. Impact of full text on print journal use at a liberal arts college. *Library Resources & Technical Services* 49 (1):19-26, 56.
- Dilevko, Juris and Lisa Gottlieb. 2002. Print sources in an electronic age: A vital part of the research process for undergraduate students. *Journal of Academic Librarianship* 28 (6):381-192.
- Hiott, Judith. 2004. Collecting and using networked statistics: Current status, future goals. *Library Quarterly* 74 (4):441-454.
- Jewell, Timothy D., Ivy Anderson, Adam Chandler, Sharon E. Farb, Kimberly Parker, Angela Riggio, and Nathan D. M. Robertson. *Electronic resource management: Report of the DLF Initiative*. Digital Library Federation 2004 [cited April 1, 2005]. Available from <http://www.diglib.org/pubs/dlfermi0408/> <http://purl.oclc.org/DFL/dlfermi0408>.
- Labelle, Patrick R. and Karen Nicholson. 2005. Student information research skills: Report on a Quebec study on information literacy. *Feliciter* (1):47-49.
- Moore-Jansen, Cathy, John H. Williams, and Mohammad Dadashzadeh. 2001. Is a decision support system enough? Tactical versus strategic solutions to the serials pricing... *Serials Review* 27 (3-4):48-61.
- Tenopir, Carol. 2001. Database and online system usage. *Library Journal* 126 (16):41-45.
- Watson, Paula D. 2003. Acquisitions of e-journals. *Library Technology Reports* 39 (2):28-41.
- 2004. Raised expectations/rising prices. *Library Technology Reports* 40 (6):40-41.

continued on page 24

lar Websites like **Google** and **Yahoo** or the metadata **CDDB** provides, directing users to granular content. *[The OpenURL is a NISO standard for programming search terms into a URL (Uniform Resource Locator) — the address of a file on the World Wide Web. The OpenURL permits a query to a database for a known item when the exact address for that item is unknown.]*

Site level linking is an answer for the Web of ten years ago. Content level linking represents the next logical stage of library, publisher and vendor supported navigation of premium content. The absence of track linking for music and chapter linking for books is a disservice. It is an obstacle to library users identifying and reaching specific content purchased for their use.

Library catalog records for networked eBooks, music and many other formats routinely include formatted and unformatted contents in the **MARC 505** field. Formatted contents have the option of adding a sub-field **u** (Uniform Resource Identifier) for each content item. The laborious solution to access these contents is the URL manually appended to every chapter or music track. That assumes stable deep link capabilities which do not yet exist for most networked content, with the exception of journal articles.

Fully networked navigation solutions are needed for all library supported networked content. We need service providers like **Bowker**, **Marcive**, **Blackwell's**, **Baker and Taylor**, **OCLC** or others to step up to provide stable reference systems. As an example, the URL could be provided at the contents level in **MARC** records for library catalogs for eBooks, in cooperation with content providers.

Library catalogs with contents fields in their records already offer users significant indexing of books and music and a host of other networked content. We must use what we have rather than wait for some magic wand to make our content viable in the current Babel of networked retrieval. Any networked content providers who want a place at the resources table as something other than a poor cousin to journals must provide the kind of flexible access now common in article retrieval systems. If they don't they will lose even more as journal based electronic networked resources continue to commandeer new spending in libraries.

Integrated Library System (ILS) vendors must also support granular content discovery and navigation in **OPACS**. Maybe what we need as a short cut through the inertia, is the ability to put an **OpenURL** icon at the end of each individual chapter or track title anywhere they are indexed. Is capturing the metadata associated with book chapters or music tracks rocket science? **CrossRef** supports **Digital Object Identifiers (DOI's)** — permanent identifiers for e-content — at the chapter level, that could be used as well, but not many have been deposited for book content. **Bowker's** entry into this field may accelerate book publisher use of **DOI's**. *[The DOI is a publisher assigned unique identifier paired with metadata and linked to the current address of an electronic item. CrossRef is the primary system in the US for this service. The DOI is most commonly used to uniquely tag individual journal articles. It could be used for much more extensive linking for many e-formats.]*

Isn't it about time for:

1. Library system vendors to support adding **OpenURLs** and/or live **DOI** links to contents notes in **OPACS**?
2. eBook vendors, publishers and other e-content providers to develop and support predictable syntax (either **DOI** or **OpenURL**) to each subtitle in a networked resource?
3. **OpenURL** system vendors to support linking syntax for all types of networked materials, not just journal articles.
4. Publishers and vendors to deposit **DOI's** for book chapters?

This is an opportunity for content providers that will allow users to navigate to specific content, thus making it more usable. It isn't enough to build a great database or a fantastic search engine for a resource anymore. Linkage coming from outside rapidly is becoming a primary means of identifying and accessing specific content.

Much has been written about the uptake of eBooks. For many libraries, the usage data is not as compelling as for journal articles. Part of the reason libraries are not buying eBooks as readily as e-journals is navigation.

We already have in depth indexing of many networked electronic titles; we just don't have simple uniform means to get users to that depth.

Standard metadata used for the **OpenURL** could be harvested from **MARC** records: ISBN, provider and chapter title for networked eBooks. In thousands of catalogs worldwide, networked eBooks already have content listings and records for electronic music includes track titles. Why can't library users just click and go directly to the content? Changes need to be made to our systems to permit deep links to the book chapter level or specific piece of music or to any identified/indexed section of an electronic resource.

Similar linking systems should be available for all e-content supported by libraries and their vendors. Librarians want to show our users "all the other" resources. If all they want is access to all electronic tracks of a movement in a concerto, or chapter in a book, we should make that content easy to identify and navigate.

It isn't just library **OPACS** and other library identification systems, but also content vendors that need to change. eBook databases have to be accessible from an external URL call at the chapter level for such navigation. **Safari Bookshelf (O'Reilly)** has a basic syntax for ISBN and chapter title. Why don't all eBook vendors? In music databases **Alexander Street Press** currently offers records for track level access to its Classical Music Library. But so far most libraries have not opted for the 40,000 or so bibliographic records they provide that would support such access. Electronic content vendors who do not support linking schemata such as the **OpenURL** or **DOI** or other stable reference system are losing to content providers who do.

Article indexing systems and journal publishers, vendors and aggregators have led the way by supporting the **OpenURL** standard making deep linking ubiquitous in the journal literatures. We have learned that a whole way of thinking about access to content can change, and change quickly. To keep the networked generation seeking and using library resources, we must have this capability for all electronic content.

If libraries and content providers and vendors don't change and change quickly, we will move from a slow crawl to a rapid march towards obsolescence. Libraries will become content mausoleums, which only draw students and faculty to the library to use computers, coffee shops or group study rooms. The library collection as an accessible easily navigable virtual space is essential if this system is to continue to have a collecting and support role for our educational institutions. All those other amenities may be enough to keep libraries relevant as place, but they aren't why institutions invest in libraries. **OpenURL** linkage for eBook content or predictable stable syntax, access to individual songs, to cataloged media and other item level content, such as quadrant coordinates for maps, is essential. Support for granular access for all networked e-content is critical to our common survival.

Response from Angela D'Agostino, Bowker

Chuck Hamaker's article wonderfully articulates a vision **Bowker** has been discussing for some time. As a **DOI Registration Agency** we support the assignment of **DOI's** to chapter level content. We believe this linking is necessary not only for digital content but also for print content while publishers gradually make the transition. Exposing the content housed in print form within the library is just as valuable as its electronic content. Think: "Search Inside The Books In Your Library." Cross referencing books mentioned within bibliographies for example is valuable even if it "resolves" to a shelf location.

Playing out a scenario from **Chuck's** article, we could, for example, assign **DOI's** to our entire table of contents database which is made available through our **Syndetic Solutions OPAC** enrichment service. By adding these to the **OPAC** record, a user could link theoretically to the full-text. We estimate less than 5% of those titles are available in electronic format. Our task then, together with the rest of the academic library community, is to help publishers define the value proposition of electronic content and interlinking. Will they sell more books? Make their content more visible? Clear more rights? In short, will they make more money as a result of this interlinking capability? If we can demon-

continued on page 26

strate that the answer is yes — with Chuck and the rest of the academic library community's support — then our mutual vision can become a reality.

Response from Jean Bedord, Consultant, Senior Analyst, Shore Communications Inc.

Chuck Hamaker thoughtfully addresses a major issue in current library **OPACS** which librarians, vendors and publishers would be wise to heed. As an adjunct faculty member at **San Jose State University** teaching online searching to future librarians, "findability" using online techniques is expected by my students as distance learners who do not have physical access to the main library. Right now, they can electronically "find" information in journal articles, but are challenged by books, eBooks and music albums, as well as other sources, which have standard catalog entries only at the title level.

But patrons want relevant pieces of information, and they want consistent, standard means to find needed pieces of content, not just title specific search capabilities, which assumes prior knowledge of the relevant title. "Findability" of the breadth and depth of the library collection needs to be reflected in the virtual space that our patrons expect to access 24 hours a day, seven days a week, just as they do the open Web. It's a new generation of library users, with different expectations, and library vendors and services have to evolve to remain relevant to their needs.

Response from Ron Boehm, ABC CLIO

I couldn't agree more with **Chuck's** points about the need to provide deep linking into e-content, and the need to make the user experience more convenient and fruitful. It may be that this is easier to do with reference eBooks than other titles. All of **ABC-CLIO's** eBooks are accessible both on the entry and on the index term level through **Paratext's Reference Universe**, which is custom built for the purpose of deep linking. We're happy to provide the entry and index term linking information to any library system vendor who desires it. We also provide a

stable and simple URL consisting of the site, the ISBN, and, where desired, the page number. This allows a user to go directly to a particular page, say a professor's extended reading selection in an encyclopedia. We are in the process of providing direct-to-entry linking.

I'd also posit that **Chuck's** argument should be extended back to the print collection. Students are already voting their choices by their strong preference for electronic resources over print resources (at least in journals and reference). The library's print collections are far less accessible than even an eBook with only title level linking, as a quick full text search will give a user all of the occurrences of a word within the title in a few seconds. That's not even possible in a print book. So, I think the e-resources are much more accessible, probably 80% of the way there. Concerted efforts by all players mentioned by **Chuck** can get us most of the rest of the way within a few years.

Response from Stephen Rhind-Tutt, Alexander Street Press

I agree wholeheartedly with **Chuck's** comments. In the past few years **Alexander Street** has worked hard to create letter, scene, diary, chapter, interview, track and speech level indexing because it is much more powerful to do so. Many works are collections of smaller items such as essays. They are written by different people, at different times on different subjects and should be indexed accordingly. If we don't do this we're limiting our ability to ask the most basic questions such as what has been written, about which subjects, by whom and when. We hear much about the demise of the monograph. Surely source level indexing contributes substantially to this? 600 pages of a journal typically has some 50 bibliographic records, each of which has an average of ten highly specific fields. That's 500 entry points to the handful of entry points that identify a 600 page monograph. Moreover, the journal article is supported by hundreds of direct citation level links from an increasing number of full-text journal providers. Deeper indexing is critical to saving end-user time because it reduces the amount of work an end-user has to do to digest what they've found and because, by being more specific, it allows the scholar to sort through fewer records for a successful search. Above all it is a pre-requisite to providing a higher functionality search system that can answer high level queries — e.g., "What comments by company presidents were made in response to visionary articles written by **Chuck Hamaker**?"

NISO Metasearch Initiative Targets Next Generation of Standards and Best Practices

by **Cynthia Hodgson** (NISO) <chodgson@niso.org>

and **Andrew Pace** (North Carolina State University) <andrew_pace@ncsu.edu>

and **Jenny Walker** (Ex Libris, Inc.) <jenny@exlibris-usa.com>

This is the first of a two-part article on the Metasearch Initiative of the **National Information Standards Organization (NISO)**. Part 1 focuses on the issues that prompted the creation of the **Metasearch Initiative** and reviews **NISO's** plan of action. Part 2, which will appear in a future issue, will review findings and recommendations of the **NISO** metasearch committees.

Metasearch — also called parallel search, federated search, broadcast search, and cross-database search — has become commonplace in the information community's vocabulary. All speak to a common theme of allowing search and retrieval to span multiple databases, sources, platforms, protocols, and vendors at once.

The **Z39.50** protocol has been the primary mechanism for providing metasearch since the first version of the standard was issued in 1988. This standard, which was initially designed to search across disparate library catalogs, has several drawbacks in today's metasearch environment. It was not designed for operation in a Web environment; it was not intended for article-level citations; and for many providers it is overly complex to implement, thus creating a high barrier of entry for many content providers.

Metasearch software providers have implemented a variety of protocols to ensure access to content including: **Z39.50**, IT search standards such as **SQL**, newer Web standards such

as **XQuery**, customized proprietary point-to-point connections, metadata harvesting, and HTML parsing or screen scraping. This multiplicity of protocols that must be supported and the lack of commonly implemented standards, best practices, and tools make the metasearch environment less efficient for the system provider, the content provider, and ultimately the end-user.

Metasearch Challenges

At the **2003 ALA Midwinter** meeting in Philadelphia, a group of resource providers met to discuss their concerns on the loads their systems were encountering from metasearch en-

continued on page 28