June 2004

Innovations Affecting Us -- XML in Action

Norman Desmarais

Providence College, normd@providence.edu

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

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

Recommended Citation

Desmarais, Norman (2004) "Innovations Affecting Us -- XML in Action," *Against the Grain*: Vol. 16: Iss. 3, Article 43.

DOI: <https://doi.org/10.7771/2380-176X.4373>

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.

The URL Clearinghouse Offers Vendor URLs

by **Laura B. Cohen** (Network Services Librarian/Webmaster, LI 140 University at Albany, Albany, NY 12222; Phone: 518-442-3492; Fax: 518-442-3567) <LCohen@uamail.albany.edu>

The **URL Clearinghouse** is a set of instructions for creating URLs for licensed digital resources organized by vendor. It focuses on direct URLs for databases and e-journals on the title level. The **Clearinghouse** is freely available on the Web and located at <http://library.albany.edu/clearinghouse/>.

Why the need for a **URL Clearinghouse**? The availability of URL services in the form of URL lists such as **Serial Solutions**, or link resolvers such as **SFX**, hasn't done away with the need for librarians to create vendor URLs. Online catalogs, custom databases, subject pages, research guides, course Web pages, electronic reserves and other Web-based tools all employ librarian-maintained URLs that connect to licensed resources. Anyone who works with these URLs knows that figuring out how to create them poses multiple challenges. Vendors employ different types of URLs that have distinct requirements in their construction and behave differently when clicked on by a user's mouse. For example, some vendor URLs must be constructed according to a specific formula. Others may become transformed by scripts after a mouse click, rendering destination URLs that cannot be used to make a connection. These are just two of the issues that face librarians who do this work.

To be sure, there are vendor Websites that offer clear instructions on URL creation that are easy to find and to use. But this is not always

the case. Instructions are sometimes difficult to locate, inaccurate, or unclear. Some vendor sites offer multiple hyperlinked title lists organized in various ways, but use different URL structures on these different lists. This leaves the librarian with an understandable confusion about which URLs to use. Even on the same list, URLs may be structured inconsistently. Those vendors that do use the recommended URLs on their sites rarely indicate this, and librarians are left to guess if these are the authorized ones or if some other structure is recommended. These and many

other problems have created a state of uncertainty about doing this work. Contacting the vendor can help, but support staff isn't always fully in the know.

I had an interesting experience at this year's **Charleston Conference** that illustrates this point. I stopped by the **Project Muse** booth and asked the sales rep why all the URLs on the **Muse** site were incorrect. **Muse** uses directory URLs, yet all were missing the final trailing slash, for example http://muse.jhu.edu/journals/social_politics rather than http://muse.jhu.edu/journals/social_politics/. If this type of URL is missing the slash, it takes the browser and server four "conversations" to complete the transaction rather than the normal two. This causes unnecessary hits to the **Muse** servers and slows down retrieval for **Muse** customers. The rep listened with interest, and within a day she called the

Webmaster to convey my question. She reported that the Webmaster had no idea why the slash would make anything better but she was willing to add it. All the URLs on the **Muse** site were corrected.

It is difficult for librarians to establish best practices in URL management under these conditions. This challenge is exacerbated by the fact that librarians tend to do this work in isolation. The library community has not pooled its knowledge about vendor URLs or provided a central repository where this knowledge can be easily obtained.

I have created such a repository. It has come out of my work of providing proxy access to licensed resources at the **University at Albany**. Determining URLs for proxy access has been enormously time-consuming and has involved a range of problem-solving techniques. Over the years I have been keeping records of how I found individual URLs and what were the rules of their construction. As my records became more extensive, I began to realize that colleagues in other libraries were working on the same issues. A centralized repository made much more sense. Out of this idea was born **The URL Clearinghouse**.

Each vendor record in **The URL Clearinghouse** provides step-by-step instructions for creating URLs. The site also includes information about the varieties of URLs employed by vendors. Librarians are invited to add vendors to the **Clearinghouse** to help make it a resource that has the potential to be widely useful.

Innovations Affecting Us — XML in Action

by **Norman Desmarais** (Acquisitions Librarian, Phillips Memorial Library, Providence College, Providence, RI 02918; Phone: 401-865-2241; Fax: 401-865-2823) <normd@providence.edu>

The software developers at **xrefer** envisioned the potential of **XML** (eXtensible Markup Language) as early as 1999. When they began to create **xreferplus**, they tagged the content with an abundance of metadata in **XML** compatible format. This metadata is fully searchable and interactive with the various components of the product. Because **xreferplus** uses **XML** compatible metadata, it is also compliant with **SFX** and similar link resolvers (see "For SFX See Librarian." *Against the Grain* 15:3 (June, 2003) pp. 102-103).

Xreferplus contains the full text of 147 books from twenty-seven publishers. Subject areas cover art, biography, business, geography, history, language, literature, law, medicine, music, philosophy & psychology, religion, science, social sciences, and technology. Technology and art, health, and science are the areas that saw the most growth in 2003. The library includes general encyclopedias like the *Crystal Reference Encyclopedia*, *The Columbia Encyclopedia*, *The Hutchinson Encyclopedia*, and *The Macmillan Encyclopedia*. The *Philip's Encyclopedia 2004* was added in the last week of January. The library also includes ten language dictionaries. There are about 76,000 audio pronunciations in MP3 format for the *American Heritage Dictionary*, the *Academic Press Dictionary of Science and Technology*, and *Dorland's Medical Dictionary*. Any images from the published sources are also included.

There is also a conversion engine that will quickly calculate English or metric equivalents for area, distance, energy, speed, temperature, fuel consumption, power, volume, and weight. This engine is fully integrated within **xreferplus**.

Many of the titles are from British publishers, especially **Peter Collin Publishing**, **Blackwell**, **Bloomsbury**, **Routledge**, and **Thames and Hudson**, showing the product's origins; but American titles have been added from publishers like **Houghton Mifflin**, **HarperCollins**, and **Wiley**. Content from American publishers will be the fastest growth area for 2004. **Harvard University Press** has recently signed up for a couple of music dictionaries. The individual titles are all richly indexed with **XML** compatible metadata that allows them to be fully cross-referenced. Unlike traditional cross-references which are tied to the framework of a single book, **xreferplus's** cross-references span the whole digital library to connect relevant information.

One can browse each book in the digital library, as in the print equivalent, to explore its content; but **xreferplus's** real power lies in the search engine. This engine, which resembles that for an Advanced **Google** search, permits **Boolean** searching, thematic searching, stem-

continued on page 92

ming, and concept searching. It supports the usual **Boolean** operators (AND, OR, NOT) but not proximity operators like NEAR or WITHIN X WORDS. One can search by subject, exact phrase, quotations, or words that are spelled or sound like the keyed entry. One can search the full text or only the headings, all the words or any of them, or exclude certain words. One can also search by exact phrase or limit the search to a particular topic or group of topics.

This is impressive enough for a search engine; but it just hints at **xreferplus's** power. When the results of a search display on the screen, a column of "**xreferences**" appears on the left. These are related cross references generated by the metadata. They go beyond traditional cross-references, which connect related terms within a single source, by cutting across topics, titles, and publishers as well as text, audio, and image files.

Clicking on an **xreference** triggers the retrieval of additional relevant material. In other words, it operates like a hyperlink that generates additional **xreferences**. Whereas traditional cross references are one dimensional, confined to the framework of a single book, **xreferences** add another dimension to reference works by providing links to navigate a collection of books or sources.

This feature is called the **Research Mapper** and uses information visualization tools that characterize hyperbolic browsers (see "Innovations Affecting Us — Hyperbolic Browsers: From GUI to KUI." *Against the Grain* 15:5 (November, 2003) pp. 95-97). This can be fascinating and intriguing to watch because the researcher can observe the retrieval engine at work as it builds the nodes and changes the distribution layout in real time, depending on what it locates. If the search engine detects a relationship between two terms, it uses **xrefer's** algorithm to test that relationship. It then groups the results by subject in nodes.

The presentation looks like a complex molecule that grows like an amoeba (see Fig. 1). The **Mapper** shows not only the entries that match the search term but also how they relate to each other and to other entries as well. The closer the results appear to another, the more they have in common. This guides the researcher to the most relevant information related to his or her research and leads to areas that warrant further exploration.

Fig. 1: Xreferplus's hyperbolic browser display for the term HOLOCAUST.

Moving the cursor over a node displays the title or subject content of that node. Pressing the shift key and clicking on a term expands the concept of that term and builds a bigger cluster, generating new and unforeseen results/relationships. This is a great tool to find desired information quickly, to expand one's knowledge of a given area, or to locate information when one doesn't know exactly what to look for. This is particularly good for someone who doesn't know what they don't know and wants to learn. By representing the search results in a visual form, **xrefer** lets researchers put more context into their research and see more results on a single page.

Fig. 2: Moving the cursor over a node highlights the content represented by that node in a yellow box and may display the content of other nodes as well. Clicking the entry in the yellow box will display its entry in a new window.

The **Research Mapper** provides a visual map of how search terms and topics in **xreferplus** are interconnected. One can access it either from the set of search results or from the home page. One can also analyze the structure of a map by holding down the Ctrl key while moving the mouse around the map. The relationships change as one moves the cursor from one node to another. The effect is illustrated below.

Fig. 3: Researchers can analyze the structure of a map by holding down the Ctrl key while moving the mouse around the map.

The bottom right corner of the **Mapper** window displays the number of results being mapped. Researchers can opt to have a maximum of 100, 200 (default), 300 or 400 results mapped. The higher the number, the longer it will take to complete the visualization. **Research Mapper** requires Java.

While **xreferplus** is SFX compliant because of the XML metadata, subscribers will probably opt not to set it as a target, as most of the entries in the reference works contained in the library have brief entries. Rather, this product would best serve as a starting point for research. Subscribing institutions might choose, instead, to have their link resolver use **xreferplus** as a starting point to link to other library resources like the public access catalog or specialized databases or more detailed reference sources. Institutions might link to individual titles in **xreferplus** from a library web page or a departmental portal.

Another nice feature of **xreferplus** is that it lets the researcher format the results in APA, MLA, or **Chicago Manual of Style** format when sending a reference to a friend or to oneself. There does not seem to be any option for downloading to disk other than email, copy and

continued on page 93

paste, or printing. **xreferplus** also automatically adds a bibliographic citation to identify the source of the reference. **MARC** records are also available for download free of charge. Catalogers can select all the records, created by **DGI (Donohue Group Inc.)**, or a subset which they can select by title.

One can subscribe to the complete **xreferplus** library or choose a "pick and mix" option to subscribe to 100 or 125 titles of their

choice. The price of **xreferplus** is negotiated by site and a 30-day free trial is available. **Xrefer** does not license **xreferplus** by simultaneous users; so the product is available to all users of a subscribing institution, including remote users.

Database administrators will appreciate the usage reports that will allow them to monitor title use and change their subscription mix if they opt for the "pick and mix" option. Statistics are collocated monthly and provide the following information: number of sessions; the number of searches performed; the number of entries displayed; the top twenty entries

viewed; the top twenty search terms performed; totals for pages including advanced search, help, home, tour, and **xreferences**; and the number of entries viewed in each of the titles in the subscription. The product can also be integrated with the "Ask-a-librarian" link which appears beneath the institutional name at the top-right of the **xreferplus** pages. **Xrefer's** American office is located at 31 St. James Ave. Suite 370, Boston, MA 02116 tel: 866-426-2654. Public views of some pre-determined interactive maps can be viewed at <http://www.xreferplus.com/>.

Wandering the Web — Portals & Politics: Federal Government Gateways

by **Rosemary L. Meszaros** (Coordinator, Government Information & Law, Western Kentucky University)

Column Editor: **Jack G. Montgomery** (Western Kentucky University) <jack.montgomery@wku.edu>

What's this? Has the federal government become **Google-ized**, **YAHOO!-like**? In a word, yes! The government megasearch engine of them all is **FirstGov.gov**. Launched in 2000, **FirstGov.gov** works with federal agencies to encourage them to create portals organized around customer groups and topics, instead of agency names. It is much better organized and complete than a later White House initiative, the disappointing, **E-gov.gov** <http://www.egov.gov>. With **FirstGov.gov's** nudge, federal government portals have proliferated. You can see some of your federal tax dollars at work by visiting these gateways to government information. The best point about the "official government Websites" is that they are official. That is, the information comes right from the spigot. Here are a few of the best.

Matthew Lesko, watch out!

FirstGov.gov: <http://firstgov.gov/> — The megasearch engine of them all, **FirstGov.gov** seems to fly in the face of the **USA Patriot Act** by styling itself as "the official U.S. gateway to all government information. . . transcend[ing] the traditional boundaries of government. . . vision is global — connecting the world to all U.S. government information and services." Searches federal, local, tribal, state, the District of Columbia, and U. S. territorial governments' Websites.

FedStats: <http://www.fedstats.gov/> — Gateway to statistics from over 100 U.S. Federal agencies plus all of the states. Search by topic, keyword, or state. **FedStats** also includes links to statistical collections like Statistical Abstract of the United States, State and Metropolitan Area Data Book, National Transportation Statistics, and Health, United States.

FedWorld: <http://www.fedworld.gov/> — Find a federal job — updated daily. Search by geography, adjacency search a position by keywords; you can exclude military jobs.

GovBenefits.gov: <http://govbenefits.gov/index.jsp> — This is a free, confidential tool that

searches for government benefits. It does not require name, phone number, Social Security number, or any other identifying information. A series of questions narrows the retrievals to pertinent federal or state government benefit programs along with information about how to apply.

Grants.gov: <http://www.grants.gov> — Unearth and apply for competitive grants in 900 programs from 26 major Federal grant-making agencies.

Medicare.Gov: <http://medicare.gov/> — In English, Spanish and Chinese, this Website features drug pricing comparisons and other up to the minute information on Medicare.

Regulations.gov: <http://www.regulations.gov/> — Locate, review, and submit comments on federal regulations that are open for comment and published in the *Federal Register*, the Government's legal newspaper. As a member of the public, you can submit comments about these regulations, and have the Government take your views into account.

Recreation.gov: <http://www.recreation.gov/> — Auto touring, biking, hiking, boating, camping, fishing, lodging, water sports, winter sports, and more for every state, the District of Columbia, and Puerto Rico. For those of us who, like **Daniel Boone**, never really get lost but may become confused, there are maps and maps and maps, from simple outlines to ones with complicated overlays.

And, finally, a megaportal to search:

Cross-Agency Portals: http://firstgov.gov/Topics/Cross_Agency_Portals.shtml — Gateways that bring together federal information and services from multiple agencies about a particular topic or for a particular customer group. An alpha index navigates the page. A click on College Students brings you to <http://www.students.gov>. There are links to testing sites (GED, ACT, GRE, TOEFL, CLEP and others), preparing for college, choosing a school, details on financing an education, a parents' guide, and much more. Click on Seniors <http://firstgov.gov/Topics/Seniors.shtml> and an array of services

for Seniors appears, everything from Consumer Protection to Education, Volunteerism to Retirement & Money to Travel & Leisure. Interested in After School Programs? Click on it and connect to **Afterschool.gov** <http://www.afterschool.gov> for sites for kids, planning activities, and how to get money.

Cross-Agency Portals is an innovative site which draws together government resources and reveals a new way to research the treasury of government Websites.

Backtalk from page 94

eral Motors to implant tags during manufacturing process to help them recover stolen cars — that data could be misused.

- Governments in search of terrorists could search for everyone.
- Insurance companies could place sensors near cancer treatment, AIDS clinics, hospitals, pharmacies, etc. to gather data and run it against the records of those seeking new policies.
- Libraries/publishers place tags in books and all sorts of bad people and organizations collect the information on the readers and use it for their purposes. (I'd like to know what books **GWB** reads — if any).

Give yourself a few minutes; I'm sure you, like me, have sufficient paranoia/malicious intent in you to think of a few more fearsome scenarios.

But as with everything, one person's problem is another's pot of gold. **Karen Dearn** in the 6 April 2004 edition of *The Australian* reports on a new product being developed by **RSA Security**. The product is seen as useful to both individuals and companies who don't want to be spied upon. Presumably people will one day be able to carry blocker devices around with them to prevent close by or remote scanners from reporting on them.

continued on page 86