



Smart Libraries™

Formerly Library Systems Newsletter™

50 East Huron Street, Chicago, Illinois 60611-2795, USA

Number 12
Volume XXIII
December 2003

Configuration key to filter effectiveness

Now that the U.S. Supreme Court has found the filtering provisions of the Children's Internet Protection Act (CIPA) to be constitutional, many public and school libraries are faced with the tasks of selecting and configuring filtering software. The choice of filtering software may be less significant in terms of accuracy than the configuration setting, according to a study by the Kaiser Family Foundation and the University of Michigan's School of Information and Health System.

Key to evaluating filtering software is the effectiveness of the filter, measured by the rates of overblocking and underblocking. Overblocking occurs

when a filter erroneously blocks a site. Underblocking occurs when the filter doesn't block an inappropriate site. A tradeoff between over and underblocking usually occurs—if one measure improves, the other degrades.

Filtering software employ two blocking methods. Keyword blocking finds taboo words on Web pages and blocks those pages. This method is relatively inexpensive for a vendor to implement but is prone to overblocking. The second method, called site blocking, relies on precompiled lists of sites. The two methods are not mutually exclusive.

See *Filter* on page 3

What does usage research tell us?

Understanding users is crucial to reshaping library services for the Internet age. User expectations are shaped by commercial and consumer services that challenge the library's traditional approach to information services.

Extensive research conducted by Carol Tenopir and Don King reveal trends in the changing use of electronic resources provided by the library. (Tenopir is a professor at the School of Information Science at the University of Tennessee in Knoxville and King is a research professor at the University of Pittsburgh.) Understanding these changes allows librarians to better manage collections and resources.

Academic collections have three types of users:

- Search experts (librarians), who want control of the process with Boolean, fielded searches, and drop-down menus
- Subject experts (faculty), who want easy access to more content
- Novices (students), who want speed and convenience

How faculty members find articles

Faculty users, in particular, reflected behavior trends that librarians could use to evaluate collections.

See *Users* on page 4

IN THIS ISSUE

Configuration Key to Filter Effectiveness
PAGE 1

What Does Usage Research Tell Us?
PAGE 1

Standards in the News
NISO Adopts New Policies
ISSN Standard to be Revised
VRA Releases Cataloging Guide
New Standard for Library Identifiers
PAGE 2

New E-book Models
PAGE 4

RedLightGreen Streamlines Search Results
PAGE 5

OCLC Releases FRBR Algorithm
PAGE 5

OCLC Googles the World's Libraries
PAGE 6

PRONOM and JHOVE aid digital preservation
PAGE 6

More New Faces at VTLS
PAGE 7

Access *Smart Libraries Newsletter*, *Library Technology Reports*, and Searchable Archives at ALA TechSource Online
PAGE 7

Correction

U. of Tennessee selects Aleph 500

In Marshall Breeding's November 2003 article, "Keeping Score with the ARL Libraries," we incorrectly stated which system the University of Tennessee at Knoxville selected to replace its Dynix Horizon system.

The University selected Ex Libris' Aleph 500 to replace its former system.

Contact: www.aleph.co.il/aleph

ISSN 1541-8820



Standards in the news

NISO adopts new policies

The National Information Standards Organization (NISO) has adopted policies on intellectual property rights, conflict of interest, and the inclusion of patented technologies in NISO standards. Given the controversy related to the OpenURL standard, the new patent policy is particularly interesting.

The patent policy states that NISO seeks to develop and promote standards that avoid embedded patents, though NISO acknowledges this avoidance may not always be possible. If a standard must include patented technology, NISO will seek assurance from the patent holder that a free license would be offered, that the patent would not be enforced, or that a license would be made available under reasonable and nondiscriminatory terms.

The three new policies are available in the Resources section of the NISO website.— *PLC*

Contact: www.niso.org

ISSN standard to be revised

As a result of the five-year review of the ISSN standard (ISO 3297), the International Organization for Standardization (ISO) has convened a working group, ISO/TC 46/SC 9/Working Group 5, to revise it. Objectives for the revised standard include:

- To clarify the types of resources to which ISSN may and may not be assigned, including those in electronic formats
- To provide definitions to support the introduction of any new concepts into the revised standard (such as continuing resource)
- To clarify policies for assigning ISSN to different editions or versions of a resource

The group is charged to complete its work by October 2006.— *PLC*

Contact: www.nlc-bnc.ca/iso/tc46sc9/wg5.htm

VRA releases cataloging guide

The Visual Resources Association (VRA) has released the first seven chapters of *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images* (CCO). Until now, little published documentation existed on data content standards applicable to cultural objects. Building on existing standards, CCO provides guidelines for selecting, ordering, and formatting data used to populate elements in a metadata record.

The guide can be applied to various metadata structures in use for describing works of art and their visual surrogates, including *Categories for the Description of Works of Art* (CDWA) and *VRA Core Categories, Version 3.0* (VRA Core). The CCO project team is seeking feedback on these preliminary chapters.— *PLC*

Contact: www.vraweb.org

New standard for library identifiers

Approved for publication in October, the *International Standard Identifier for Libraries and Related Organizations* (ISIL) standard, ISO 15511:2003, defines and promotes the use of a set of standard identifiers to uniquely identify libraries and related organizations.

ISIL uses a flexible structure that can accommodate existing identifiers and be used to create new identifiers. The identifier can be assigned by a national bibliographic agency or a non-state entity. It consists of a prefix that identifies a library identifier string and the country or nonstate agency responsible for assigning the identifier string.

Because the ISIL accommodates overlapping national and transnational coding systems (for example, the British Library and OCLC), an organization could have more than one ISIL. The Danish National Library is the International ISIL Registration Authority. It will maintain a central source of information about ISIL-authorized national assignment agencies and coordinate compliance with the standard.— *PLC*

Filter *from page 1*

Lists of sites to block are generally compiled by programs that spider the Web looking for particular keywords, with or without human review. And most products have the ability to perform a second level of keyword blocking on pages allowed through the site-blocking filter.

Filtering programs generally define multiple categories for blocking such as pornography, nudity, drugs, hate speech, and job search (attractive to corporate customers). Categories can be enabled or disabled when configuring the filtering software.

The litmus test

The Kaiser study sought to determine the extent to which young people's access to legitimate health information was blocked by Internet filters. The study tested seven commonly used products at three configurations. The least restrictive level blocked categories related to pornography only. The intermediate level blocked additional categories such as drugs, nudity, weapons, and hate. The most restrictive level blocked any category that a school might plausibly block, such as tobacco, jokes, and games.

Researchers simulated searches on 24 health topics without blocking and at each of the three levels of blocking. They also simulated a search for pornography. In the end, 3,053 health sites and 516 pornography sites were tested against the filters.

A key finding of the study was that setting filters to more restrictive levels “does not significantly increase their effectiveness at blocking pornography, but does significantly increase overblocking of non-pornographic health content.”

At the least restrictive level, from 0.6% to 2.3% of legitimate health sites were blocked. At the most restrictive level, 15% to 35.5% were blocked. This variance indicates that although the different products do vary in effectiveness, how a product is configured has a more significant impact on effectiveness than the choice between different products.

Considering these findings in relation to CIPA, libraries should remember that no vendor offers a category that directly reflects CIPA requirements. Librarians will have to choose the categories that map the closest. Also, the methodology used may overstate underblocking because CIPA applies only to images, not to text.— *Priscilla L. Caplan*

Contact: The report of Kaiser study, “See No Evil: How Internet Filters Affect the Search for Online Health Information”

www.kff.org/content/2002/20021210a

CIPA sections of WebJunction

www.webjunction.org

E-rate Central

www.e-ratecentral.com

ALA's CIPA site

www.ala.org/Content/NavigationMenu/Our_Association/Offices/ALA_Washington/Issues2/Civil_Liberties,_Intellectual_Freedom,_Privacy/CIPA1/CIPA.htm

Consider filters' other features

Effectiveness is not the only way to evaluate filtering software. Other features libraries should evaluate include:

- Whether the software runs as a client on each workstation, on a local network server, or as a remote proxy application
- If the software runs as a client and whether it is compatible with hardware and other software on library machines
- How easy the software is to install and update, and who does it—the library or the vendor
- The ease of adding and removing computers or configuring the software differently for different machines
- The ease of adding local entries to the blocklist and to the whitelist (sites that should be permitted)
- What logs are kept and what statistical reports are provided
- What procedures are required to override a block on a particular machine. (The Supreme Court judgment indicated that a library could be liable if it fails to unblock access to constitutionally protected sites on the request of a user.)— *PLC*

Users *from page 1*

- 49% browse print and electronic journals to conduct background research or to keep up with the literature. The number of faculty members who browse in print is declining due to a decrease in print subscriptions.
- 24% search to find older articles in databases and to determine whether time should be spent to retrieve them from print collections. Faculty increasingly is searching databases as a way to find articles.
- 11% discover articles through citations and 13% from references by colleagues. Linking, which didn't exist five years ago, is increasingly important as a way to find and read articles.

Trends in faculty access:

- The average number of articles read by scientists has increased 44 percent over the past 25 years. This increase is in part attributed to the number of articles conveniently available online and the time the researcher saves in retrieving pertinent articles.
- 85% of articles read were published in the last five years (journals are of interest because of their currency). More journals are

extending their online backfiles, though, so predicting future usage is difficult.

- Reduced use of print collections raises the cost per use and causes librarians to question the viability of those collections.

Use of databases provides more content for faculty use and emphasizes searching over browsing as discovery tools. Although these trends are stronger in the sciences, other fields are likely to follow as electronic journals become more widely available.— *Judy Luther*

Contact: [Patterns of Journal Use by Scientists through Three Evolutionary Phases](#)

www.dlib.org/dlib/may03/king/05king.html

[Patterns of Journal Use by Faculty at Three Diverse Universities](#)

www.dlib.org/dlib/october03/king/10king.html

[Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies](#) by Carol Tenopir

www.clir.org/pubs/reports/pub120/pub120.pdf

NEW E-BOOK MODELS

In early 2004, eBooks Corp., based in Perth, Australia, will introduce its eBook Library (EBL) to the U.S. market. It will allow users to download individual titles to their PCs for designated periods using Adobe's Content Server 3.0. This approach lets libraries manage loans and reserve transactions.

EBL's new model is distinguished by the concept of Non-linear Lending™, which allows multiple users to access the same book simultaneously. The total number of pages accessed per book per year is restricted, however, to satisfy the publishers' interest in selling multiple copies of the book. Limited copy-paste and printing are allowed.

Developed with a board of advisors (academic and research librarians and publishers from the United States, Australia, Europe, and United Kingdom), EBL offers new capabilities that support library applications. These capabilities include the ability to order, receive, and catalog an e-book with a single click. Usage reporting will enable librarians to manage their collections more frequently.

Individual chapters can be designated for electronic reserve or be combined with chapters from other e-books in library-designed e-packs. An innovative short-term rental model is designed to reduce interlibrary loan (ILL) costs.

For three years eBooks Corp. has been building a commercial database of 25,000 e-books, including significant academic content from Oxford, Cambridge, and Taylor & Francis. eBooks Corp. has just implemented the capability for full-text searching across all titles, which will allow users to identify characters in novels and to select books they would not otherwise discover.

Certain subject areas—travel, religion, self-help, business, fiction, and computing—are popular in electronic form as evidenced on the company's commercial home page, where users are able to download books in Adobe, Microsoft, and Palm formats.— *JL*

Contact: www.ebooks.com

RedLightGreen streamlines search results

With funding from the Mellon Foundation, Research Libraries Group (RLG) has created RedLightGreen, a new interface for its database of 42 million unique items and 126 million records. The pilot for RedLightGreen, tested this fall at Columbia University, New York University, Swarthmore College, and the University of Minnesota, should make RLG's database more accessible to undergraduate researchers and Internet users. RedLightGreen features include:

- **Search using keywords, not Library of Congress subject headings.** Using Recommend Inc's MindServer technology, RedLightGreen employs data mining tools to correlate subject searches and uncover information that has previously been hard to find. The software creates a context for terms and then groups related items. As a result, the user can perform keyword searches and identify books and resources without needing to understand how a thesaurus works. The search results page offers clickable options for refining a search by specific subjects, authors, and languages.
- **Relevant results ranked by subject and holdings.** By evaluating the relationships among different subject headings assigned to the same work, the system interface can determine relevance. Knowing how widely titles are held among RLG's member libraries also indicates the popularity of specific works. These two tools are combined in sorting search results to present a useful list to the user.
- **Shorter results lists.** Using Functional Requirements of Bibliographic Records (FRBR), RLG has restructured its data-

base of bibliographic records based on the work rather than an expression (performance) or manifestation (format) of it. This method effectively groups the 395 entries by author Michael Crichton into 20 titles. One of those, *The Andromeda Strain*, represents 36 bibliographic records.

- **Ability to export customized results.** Users can format citations to create bibliographies in selected styles: American Psychological Association (APA), University of Chicago (Turabian), or Modern Language Association (MLA) and then save or e-mail them.

Early RLG user studies defined five significant service requirements:

- **Discovery:** Offer students relevant, current, and authoritative works.
- **Filter:** Let students easily narrow and sort their results.
- **Availability:** Provide local holdings and other options for obtaining materials.
- **Personalization:** Support functional integration with the student's research.
- **Usability:** Deliver an intuitive interface.

RLG's research echoes the findings in 2002 studies by the Digital Library Federation and Outsell, Inc., and a separate study by OCLC, showing that Internet users valued accurate information but felt the Web fell short of their expectations half the time.

Users have a chance to take the interface for a test drive and provide feedback at www.redlightgreen.org. — *JL*

Contact: www.rlg.org/redlightgreen

OCLC releases FRBR algorithm

In August 2003, OCLC made its algorithm for FRBR publicly available for ILS vendors and national libraries to experiment with applications. Use of FRBR effectively shortens the list of search results. Thom Hickey, chief scientist at OCLC Research, expects to see FRBR incorporated into First Search on the reference side in late 2004.

FRBR (pronounced *fur-ber*) will increase the ease of use of any catalog or database that has been "frbr-ized" as it consolidates all the editions and formats of a work for the user before presenting them with detailed records on each entry. For example, FRBR would group the 400 records of Arthur Conan Doyle's *The Adventures of Sherlock Homes* under one work on a search results list. The user could then click on that entry to examine the expression (performed by a different artist), the manifestation (video, CD, DVD), and the item (copy) of that work.

The FRBR algorithm describes a process that compares information from MARC 21 records against a standard name authority file and groups the records based primarily on the author and title.

The concept for FRBR came as a recommendation from a 1998 IFLA meeting, and RedLightGreen by RLG is one of the first implementations.—*JL*

Contact: www.oclc.org/research/software/frbr



OCLC GOOGLES the WORLD'S LIBRARIES

To increase awareness of library resources, OCLC is making 2 million records available for indexing by Google. Although these records represent only 3.8% of the 53 million WorldCat records, each title is held by more than 100 libraries.

About 12,000 libraries are participating in this pilot, which runs June 2003 to June 2004. Internet users searching for a book will see a Google listing for the

item that appears in WorldCat (whose byline is "Window to the world's libraries"). When users select the WorldCat link, they can enter their ZIP code, and they will then be presented with a list of nearby libraries that have the item they searched. Internet users will not be able to search WorldCat for other items.

At the end of the pilot, OCLC will evaluate feedback from participants, focus

groups, click-through statistics, and other metrics to understand the pilot's strengths and weaknesses. The primary criterion is whether this effort helped libraries and their users. Libraries are included by default if they meet certain criteria, or they can petition to be included or to be excluded.—*JL*

Contact: www.oclc.org/worldcat/pilot/default.htm



PRONOM and JHOVE aid digital preservation

The preservation community has two new tools for obtaining digital file format information.

The Digital Preservation Department of the U.K. National Archives is offering a database system called PRONOM with the aim of storing information on all software products ever used to generate electronic records. The application became freely available on the Web in October 2003.

To preserve electronic records, PRONOM identifies software products that are no longer supported and aids in planning a migration path for the records generated by these products. It manages information about the file formats used to store electronic records and the software applications needed to render these formats.

The database is growing rapidly and contains information on more than 250 software products, 550 file formats, and 100 manufacturers. Users can request or submit new information for inclusion in PRONOM using an online form.

The JSTOR/Harvard Object Validation Environment (JHOVE) is a tool for identifying, validating, and characteriz-

ing digital objects. Developed by the Harvard University Library and JSTOR, JHOVE is an extensible software application that can answer questions such as What format is this digital object? Does it conform to the specifications for its format? What are its significant properties?

Librarians frequently condition digital preservation policy and processing decisions on the digital file format. To achieve necessary operational efficiencies, repositories need to be able to automate decisions about file formats to the fullest extent possible. JHOVE (pronounced *jove*) expedites this process.

JHOVE is a stand-alone, command-line-oriented Java application with an extensible plug-in architecture. In its initial release, JHOVE includes modules for recognizing and validating ASCII and UTF-8 encoded text, TIFF, and PDF. The JHOVE website offers documentation, a tutorial, examples of use, and instructions for downloading.—*PLC*

Contact: <http://hul.harvard.edu/jhove/jhove.html>
www.pro.gov.uk/about/preservation/digital/pronom

More new faces at VTLS

The flow of personnel from Ex Libris (USA), Inc. continues: Debra Novosad has joined VTLS, Inc. as Virtua product manager and Jan Sheppard will serve as a sales executive. Before her tenure at Ex Libris, Novosad was employed by Dynix Corp. (then epixtech, Inc.) for 11 years.

The migration of two executives from one company to another might not be noteworthy, except that this particular path is well worn. Carl Grant, president of VTLS; Martha Gettys, vice president, sales; and Ron Passmore, director of marketing each moved from Ex Libris to VTLS, as noted in the September 2003 issue of *Smart Libraries Newsletter*.

Despite the infusion of top talent, sales of Virtua remain sluggish domestically. VTLS has made some sales of Virtua in the last few months, including Edward Via Virginia College of Osteopathic Medicine at Virginia Tech (in VTLS' hometown of Blacksburg, Va.), the Scugog Memorial Public Library in Port Perry, Ontario, Canada, and the American Public University System, based in Charlestown, W.V., specializing in distance education.

None of these sales can be considered major wins. The real mark of success for VTLS will be if or when a major North American library selects Virtua in a head-to-head competitive process with the other top-tier automation systems.

With some of the most experienced sales executives in the industry at the helm, the stakes are high to demonstrate that Virtua has a place in the top tier of library automation systems.— *MB*

Contact: www.vtls.com

Access *Smart Libraries Newsletter*, *Library Technology Reports*, and searchable archives at ALA TechSource Online

Do you prefer immediate access to periodicals? Are you actively converting your serial subscriptions online? Do your facilities have limited physical storage space? Subscribe to ALA TechSource Online to answer all these needs.

With ALA TechSource Online, you have instant access to all *Library Technology Reports* since 1999 and the complete archives for *Smart Libraries Newsletter* (formerly *Library Systems Newsletter*)—July 1981 through the present.

Subscribers can search both publications and the archives by keyword, author, title, date, or issue. Issues are presented in .pdf format. Subscriptions are \$370 and are accessible for one year from the date of purchase.

Contact: www.techsource.ala.org



Library Technology Reports Expert Guides to Library Systems and Services

Jan/Feb 2004 Marshall Breeding

Integrated Library Software: A Guide to
Multiuser, Multifunction Systems

Mar/Apr 2004 Lori Bowen Ayre

Filtering and Filter Software

May/Jun 2004 Barbara T. Mates

Computer Technologies to Aid Special Audiences

Jul/Aug 2004 Susan Gibbons

Establishing an Institutional Repository

Sep/Oct 2004 Stephen Chapman

Techniques for Creating Sustainable Digital Collections

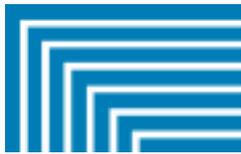
Nov/Dec 2004 Paula D. Watson

E-Publishing Impact on Acquisition and Interlibrary Loan

Coming soon—

Jan/Feb 2005 Brad Eden

3-D Visualization Technologies



ALA TechSource
www.techsource.ala.org

Smart Libraries Newsletter
American Library Association
50 East Huron Street
Chicago, IL 60611-2795 USA

NON PROFIT
US POSTAGE
PAID
PERMIT 1479
ROCHESTER, NY

December 2003 Filtering software selection and configuration

Smart Libraries Newsletter

Smart Libraries Newsletter delivers hard data and innovative insights about the world of library technology, every month.

Contributing Editors
Marshall Breeding
615-343-6094
marshall@breeding.com

Priscilla L. Caplan
352-392-9020, ext. 324
pcaplan@ufl.edu

Judy Luther
610-645-7546
judy.luther@informedstrategies.com

Andrew K. Pace
919-515-3087
apace@unity.ncsu.edu

Publisher
Miriam Wuensch
800-545-2433, ext. 3252
312-280-3252
mwuensch@ala.org

Editor
Chris Santilli
630-495-9863
chris@wordcrafting.com

Administrative Assistant
Judy Foley
800-545-2433, ext. 4272
312-280-4272
jfoley@ala.org

TO SUBSCRIBE

To reserve your subscription, contact the Customer Service Center at **800-545-2433, press 5 for assistance**, or visit www.techsource.ala.org.

The 2003 subscription price is just \$85 US.

Subscribers who would like an e-mailed version of the newsletter each month should forward their e-mail address and ALA identifier (the 7-digit number printed on the top line of the address label that appears on page 8 of your newsletter) to mwuensch@ala.org. Type "e-mail my Smart Libraries" into the subject line. Issues will be e-mailed in addition to your print subscription and at no additional charge.