

ONLINE SEARCHER

MARCH | APRIL 2016 VOL 40 | NO 2

INFORMATION DISCOVERY, TECHNOLOGY, STRATEGIES

Meeting User Needs With Cataloger's Desktop

Advanced Twitter
Search Commands

ACRL's New Information
Literacy Framework

Vendor Videos for
Training and Sales

contents

MAR/APR 2016



features

Meeting User Needs With Cataloger's Desktop 10

Bruce Johnson, Derek Rodriguez, and Susanne Ross

The Library of Congress, assisted by Search Technologies, Inc., redesigned Cataloger's Desktop to help librarians around the world create metadata to bibliographically control library resources. This custom search application required paying close attention to user needs and deploying ongoing, iterative changes.

Synthesizing, Extracting, and Customizing Web Data..... 16

Ernest R. Perez

Tools such as the free import.io facilitate customized data capture from websites and storage of information as formatted, easily transferable data files. Long time data librarian Ernest Perez walks us through the process. Data extraction services, using browser technology, offer opportunities to expand opportunities for information professionals.

Survey Research: Polling and Beyond.....22

Barbie E. Keiser

The internet provides a fantastic platform for pollsters, whether they're in the political sphere or not. Barbie Keiser explores some online survey tools and examines those that provide survey results. She adds tips about understanding and constructing survey instruments. **SUPPLEMENTAL CONTENT AND LIVE LINKS**

New STN: Remaking a Classic While Keeping the Classic.....28

Robert E. Buntrock

STN has long been the gold standard for chemistry searches. Its newest iteration, called New STN,

adds important new features while retaining the best of its classic version, writes experienced chemical searcher Bob Buntrock.

Advanced Twitter Search Commands.....50

Tracy Z. Maleeff

Billions of people use social media. But for information professionals, it's the advanced search capabilities that attract their attention, Tracy Maleeff goes deep into the advanced search features of Twitter, detailing

not only what they are, but also how best to use them.

Vendor Videos for Training and Sales..... 56

Marydee Ojala

Increasingly, our vendors are using videos to sell their products and effectively train both information professionals and end users on how to use those products. *Online Searcher* editor-in-chief Marydee Ojala describes the advantages and disadvantages of video and looks at individual vendors' video channels.

searcher's voice...

THE SEARCHER'S VOICE...33

Resurrecting the Reference Interview

Barbara Quint

As bq reminds readers—in part by sharing one shining moment from her own search history—when you help your client more precisely define what the question is, the more precise the results will be.

INTERNET EXPRESS.....35

How Public Libraries Can Help Syrian Refugees

Irene E. McDermott

Irene McDermott shares this sobering statistic to kick off her column: As of November 2015, more than 800,000 Syrians had sought refugee status through the United Nations. She goes on to share a multitude of ways information pros can help these refugees begin their new lives. And it all starts with the public library. **LIVE LINKS**

Technological Singularity and the Future of Information and Libraries.....38

By Nancy K. Herther

In writing about our quickly changing technological world, Nancy Herther uses the backdrop of astronomical and physics concepts such as singularity to theorize how these technologies could impact information professionals as well as humankind as a whole. **LIVE LINKS**

ACRL's New Information Framework: Why Now and What Did It Discover?..... 46

By Robert Berkman

Bob Berkman interviews Sharon Mader, visiting program officer for information literacy at ACRL, to find out the impetus behind the association's release of its Information Literacy Framework for Higher Education and what the Framework's biggest challenges are to date.

columns

On the Net.....	62
Searching for Definitions	
Greg R. Notess	
Online dictionaries and definition sources provide valuable alternatives to print dictionaries. It's not only web search engines and specific websites that help with definitions, it's also commercial databases.	
The Dollar Sign.....	65
Branding Irons at the Ready: Researching Brand Names	
Marydee Ojala	
Look around, and you're bound to notice at least one brand name, if not more. Researching brand names involves reference works, general business databases, intellectual property resources for trademark identification, and news stories.	
The Open Road.....	68
Open Knowledge Versus Knowledge Management	
Abby Clobridge	
What do open access, open knowledge, and knowledge management have in common? Quite a lot, says Abby Clobridge, who delves into the similarities and differences. Consider who is the intended consumer, security and privacy issues, metrics, and culture.	
InfoLit Land.....	71
Evidence and the Doubter	
William Badke	
Conspiracy theories, fake science, distorted views of events, and incorrect interpretation of facts are the province of the doubters, which creates real dilemmas for those teaching information literacy. Outliers may be proved correct or dead wrong. Bill Badke outlines five lessons for the information-literate when striving for credibility in research.	
Control-Shift.....	74
Do This, Not That: A Holistic View of the Library's Web Presence	
Darlene Fichter and Jeff Wisniewski	
New developments in library website design include multiple site versions, artisan design when coding, and encouraging contributions of content that support responsive design. Educating staff about content creation for websites is integral to excellent user interfaces.	
Hard Copy.....	77
Recommended Reading on Transformations, Opting Out, Expert Information, and Causal Relationships	
Deborah Lynn Wiley	
Online Spotlight.....	80
Questioning and Streamlining Decisions	
Mary Ellen Bates	
Decisions, decisions—we all have to make them. Mary Ellen Bates suggests a triage method for the reference interview and information delivery. She recommends taking advantage of the value-added features of resources, both commercial and on the free web.	
departments	
Front Lines.....	4
Future, Thy Name Is Information	
Marydee Ojala	
Industry News.....	6
Search Engine Update.....	8
Greg R. Notess	
Conference Corral.....	59
Technology Transforms Libraries: ALA's Midwinter Conference	

Vice President, Content	Dick Kaser
Editor-in-Chief	Marydee Ojala
Senior Editor	Barbara Quint
Editorial Services Manager	Lauree Padgett
Assistant Production Editor	Alison A. Trotta
Production Manager	Tiffany Chamenko
Senior Graphic Designer	Jackie Crawford
Advertising Director	David T. Panara
Ad Trafficking Coordinator	Jackie Crawford
V.P., Marketing & Business Devel.	Thomas Hogan, Jr.
Circulation Manager	Janeen Welsh
V.P., Information Technology	Bill Spence
President & CEO	Thomas H. Hogan
Chairman of the Board	Roger Bilboul
V.P. & CAO	John Yersak

Online Searcher: Information Discovery, Technology, Strategies (ISSN: 2324-9684; USPS: 955-110) is published six times per year (January/February, March/April, May/June, July/August, September/October, November/December) by Information Today, Inc., 143 Old Marlton Pike, Medford, NJ 08055-8750 USA; Phone: (609) 654-6266; Fax: (609) 654-4309; Internet: www.infotoday.com. Registered in U.S. Patent & Trademark Office.

Periodicals postage paid at Vincennes, N.J., and additional mailing offices. © Copyright, 2016, Information Today, Inc. All rights reserved.

No part of this publication may be reproduced in whole or in part in any medium without the express permission of the publisher.

Printed in USA

POSTMASTER: Send address changes to *Online Searcher*, 143 Old Marlton Pike, Medford, NJ 08055-8750.

RIGHTS AND PERMISSIONS

Permission to photocopy items is granted by Information Today, Inc. provided that a base fee of \$3.50 plus \$0.50 per page is paid directly to Copyright Clearance Center (CCC), or provided that your organization maintains an appropriate license with CCC.

Use CCC's online service to obtain permission to use these materials in academic coursepacks or for library reserves, interlibrary loans, document delivery services, or as classroom handouts; for permission to send copies via email or post copies on a corporate intranet or extranet; or for permission to republish materials in books, textbooks, and newsletters.

Contact CCC at 222 Rosewood Drive, Danvers, MA 01923; (978) 750-8400; Fax: (978) 646-8600; www.copyright.com. If you live outside the USA, request permission from your local Reproduction Rights Organization. (For a list of international agencies, consult www.ifro.org.)

For all other requests, including making copies for use as commercial reprints or for other sales, marketing, promotional and publicity uses, contact the publisher in advance of using the material. For a copy of our Rights and Permissions Request form, contact Lauree Padgett, lpadgett@infotoday.com.

ONLINE ACCESS

Visit our website at www.infotoday.com/OnlineSearcher.

Searchable archive of all articles with digital document delivery: www.ifi-infocentral.com

Contents also available online under direct licensing arrangements with EBSCO, NewsBank, ProQuest, and Gale and through redistribution arrangements with information service providers including, Dow Jones Factiva, LexisNexis, CCLC, STN International, and Westlaw.

SUBSCRIPTION INFORMATION

Subscriptions to *Online Searcher* are available at the following rates (per year): Subscribers in the U.S.—\$143; subscribers in Canada or Mexico—\$160; subscribers outside North America—\$187. All amounts payable in U.S. funds. Subscribe online or email custserv@infotoday.com.

Back issues: \$27 per copy, prepaid only. Mailed issues within the U.S. must be claimed within 45 days of publication date.

Change of Address: Mail requests, including a copy of the current address label from a recent issue, and indicating the new address, to *Online Searcher*, 143 Old Marlton Pike, Medford, NJ 08055-8750 USA.

Reprints: For quality reprints of 500 copies or more, contact David T. Panara, Advertising Director, Information Today, Inc., 143 Old Marlton Pike, Medford, NJ 08055-8750; Phone: (609) 654-6266 x 149; email: dpanara@infotoday.com.

DISCLAIMERS

Acceptance of an advertisement does not imply an endorsement by the publisher.

The views in this publication are those of the authors and do not necessarily reflect the views of Information Today, Inc. (ITI) or the editors.

While best efforts to ensure editorial accuracy of the content are exercised, publisher assumes no liability for any information contained in this publication.

The publisher can accept no responsibility for the return of unsolicited manuscripts or the loss of photos.

EDITORIAL OFFICES

If you are interested in any of the following please contact the editor Marydee Ojala, P.O. Box 76225, Indianapolis, IN 46278-0225; (317) 676-8100; Fax: (317) 676-8900; marydee@ivmision.com.

• **Letters to the Editor:** We welcome comments and suggestions regarding the content or appearance of *Online Searcher*. Letters must include your name, address, and email address. We reserve the right to edit all submissions.

• **New Products and Updates:** Direct all general product press releases, announcements, and news to: Marydee Ojala at the address above. Please note that our News editorial lead time is 6 weeks prior to each issue.

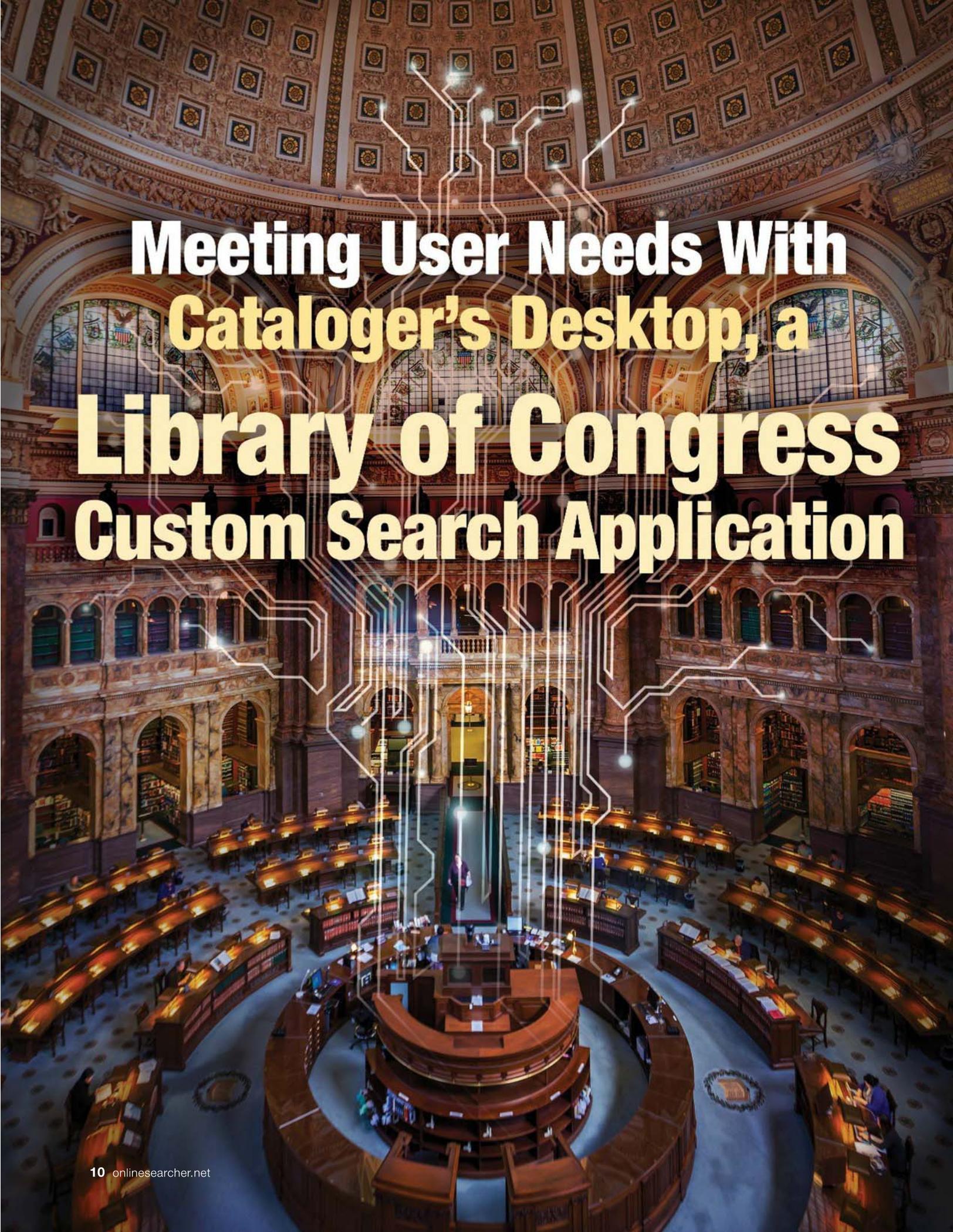
• **Writing:** Please contact the editor if you're interested in writing for *Online Searcher*.

To contact Senior Editor Barbara Quint: 932 11th Street, Suite 9, Santa Monica, CA 90403; (310) 451-0252; bquint@gmail.com

LIST SERVICES AND ADVERTISING

David T. Panara • (609) 654-6266, ext. 132 • dpanara@infotoday.com





Meeting User Needs With Cataloger's Desktop, a Library of Congress Custom Search Application

By Bruce Johnson, Derek Rodriguez, and Susanne Ross

As the United States' oldest federal cultural institution and the largest library in the world, the Library of Congress (LC; loc.gov) has millions of books, recordings, photographs, maps, and manuscripts in its collections. LC also provides leadership to libraries worldwide. As part of its mission to provide services to the global library community, LC information professionals developed Cataloger's Desktop ("Desktop"), a searchable cataloging, metadata, and library automation documentation system consisting of 300-plus resources (loc.gov/cds/desktop).

Librarians use Desktop in their daily work to find the instructions they need to create metadata to bibliographically control library resources. More than 10,000 librarians at approximately 1,000 subscribing institutions use it. Since its initial release in 1994, Desktop has evolved into a widely used and authoritative web-based service that allows professional catalogers to work more efficiently with the most up-to-date, authoritative cataloging information at their fingertips.

Search Technologies (searchtechnologies.com) provides software development and managed services to support Cataloger's Desktop. LC and Search Technologies recently redesigned Desktop as part of a system migration project. Critical success factors for this project included the following:

- Engagement with subscribers during design and implementation. Collecting user insights and acting on that feedback was essential to ensuring the updated service met end-user needs.
- Iterative software development allowing for continuous review and testing throughout the project

- Selecting current open source and licensed technologies appropriate to meet LC's needs for search, content preparation, and the user experience

This article highlights the work involved in implementing the best-in-class Cataloger's Desktop 4.0, a custom search application.

OVERVIEW OF DEVELOPMENT HISTORY

Like many documentation systems, Cataloger's Desktop began as a collection of print publications from a variety of sources. Each publication used its own editorial practice, and little thought was given to making them easy to use with one another.

Pre-1994: In the early 1990s, online documentation was in its infancy. The Library of Congress began considering how to make it easier for librarians to find answers to technical cataloging questions. Print publications of the time were published in a variety of forms and frequencies. Some publications appeared in interfiled loose-leaf form, while others had multiple supplements that had to be searched independently to answer basic questions. Much time was lost searching and manually updating resources that were limited by the print nature of their distribution. This led LC to explore technologies to allow users to search and navigate among several electronic resources on their desktop computers.

1994–2005: A goal for the new application was to allow librarians to use online documentation resources while working concurrently in an online cataloging application. A Windows-based application appeared to be the best solution. Accordingly, the Library of Congress settled on FolioViews software for distributing about a dozen resources on CD-ROM. Librarians used Cataloger's Desktop to search and browse content in the electronic product, which was a radical technology leap forward at the time. (Remember that Windows was new in the mid-1990s and had not achieved the level of acceptance that it enjoys today.)

Viewing a resource: Subscribers can view documents in the application. Search terms are highlighted to provide context. Tables of contents, inter-document links, and next/previous links allow users to easily and intuitively navigate in context. (See Figure 2 below.)

- Harvesting website content
- Generating metadata to support Desktop's search relevancy model
- Creating hierarchical and navigation metadata to support browse and document linking
- Providing file repository services to support content rendering

Query Processing Language (QPL): QPL (searchtechnologies.com/query-parsing-language) parses user queries, implements the relevancy model, and supports custom features such as multi-term synonyms.

Search service layer: A Java-based servlet application provides interfaces for querying Solr using SolrJ, the Java API for Solr, and interacting with the personalization data store (MySQL).

User experience: The user experience supports search, browse, view, and personalization and is built on HTML 5 and Bootstrap (getbootstrap.com) technologies. The application is responsive, supporting adaptive display through web browsers and tablet computers.

Hosted in the cloud: Desktop is deployed on Amazon EC2 (aws.amazon.com/ec2), leveraging cloud services for high availability, backup, and extremely fast system response.

LAUNCH OF CATALOGER'S DESKTOP 4.0

Desktop was ready by late summer 2014 for its last rounds of testing and a launch in concert with a new content release. LC worked with Desktop customers to ensure a smooth release.

Expert users were recruited from the customer base and given early access to Desktop 4.0. These users provided valuable feedback about adjusting to the revised application and were available as go-to resources for other subscribers.

To further facilitate the release, LC hosted webinars attended by hundreds of subscribers around the world to introduce the new system and address questions. These webinars were recorded and subsequently posted online for subscribers who were unable to attend the webinars due to time-zone challenges. Emails sent to the Desktop discussion list kept customers informed.

Desktop 4.0 launched on Sept. 10, 2014. Very few support questions were received, as customers found the new updated service intuitive, easy-to-use, and requiring little to no training.

BENEFITS TO USERS AND LC AS A PUBLISHER

Additionally, the application supports work tasks and provides multiple, intuitive pathways for information discovery. Desktop meets the needs and information-seeking styles of a wide range of subscribers. It is responsive to multiple screen dimensions; it can be used side-by-side with other applications on a desktop computer or on tablet computers. The improved Desktop allows users to find needed information efficiently, saving time in their work.

Anna M. Ferris, associate professor, University of Colorado–Boulder, a longtime user of Cataloger's Desktop, told LC, "As an original cataloger, I have witnessed Cataloger's

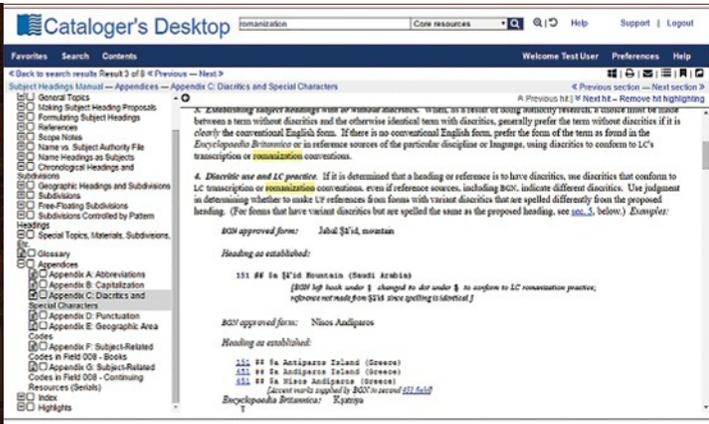


Figure 2: Viewing a document in Cataloger's Desktop

Subscribers can personalize Desktop:

My resources: Users can define a personalized resource set that they use most often.

Internationalization: Users can select one of four languages for their user-interface: English, French, German, or Spanish.

Bookmarks: Subscribers can save bookmarks for specific documents and share them with colleagues.

Shortcuts: Users can create shortcuts to be saved in a browser or inserted into an external web document.

Embedded help: Desktop features context-sensitive help and tool tips that provide instructions and guides to using the application.

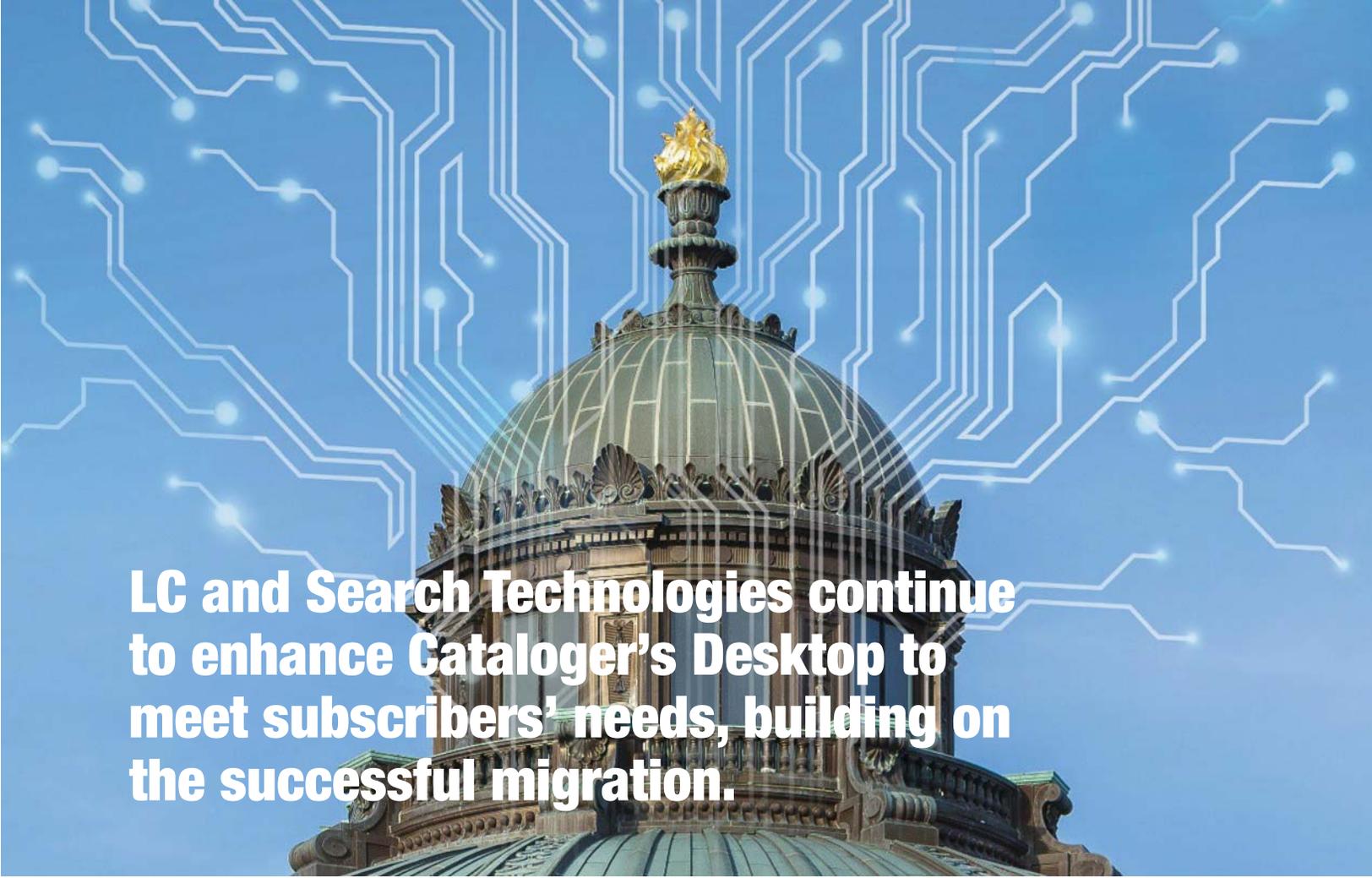
ITERATIVE SOFTWARE DEVELOPMENT AND TECHNOLOGIES FOR DESKTOP 4.0

Search Technologies architects and engineers worked with LC throughout 2014 to develop the new Desktop. Using a "code a little, test a little" approach, Search Technologies provided LC with new versions of the service for evaluation every couple of weeks. Frequent testing and feedback cycles allowed the team to adjust Desktop's design and development schedule.

Desktop employs the following technologies:

Solr search engine: The open source Solr search engine (lucene.apache.org/solr) is the core of the new system. Solr indexes content and supports all search, browse, and navigation actions.

Content processing using Aspire: Desktop includes content in a variety of formats including Folio infobases, SGML, XML, HTML, PDF, MS Office documents, and remote websites. The Aspire Content Processing Framework (searchtechnologies.com/aspire) prepares Desktop content by doing the following:



LC and Search Technologies continue to enhance Cataloger's Desktop to meet subscribers' needs, building on the successful migration.

Desktop's evolution as the working cataloger's most indispensable tool. Each development phase has resulted in a state-of-the-art resource that has evolved with the proliferation of resources we must routinely consult. What's more, the new and improved Desktop interface provides end-users with a remarkably intuitive and user-friendly system." She added, "As my institution's Desktop 'expert user,' my role as trainer for new catalogers has nearly been superseded by the excellent enhancements made to the training and support modules within the resource. In short, the Cataloger's Desktop team should be commended for being so forward-looking and for providing us with the tools that help us be more efficient catalogers."

As a publisher, LC is no longer dependent on unsupported, legacy technologies. The application uses open source and licensed components designed to support services such as Desktop. Using modern technologies allows the team to continue to improve the service to meet user needs. Deploying Desktop on Amazon EC2 results in high availability, reliable system management tools, and an extremely fast response time. Subscribers made the switch to the new user interface with little difficulty and have not looked back.

According to Beacher J. Wiggins, director for acquisitions and bibliographic access, Library of Congress: "To recap the evolution of Cataloger's Desktop is to realize the enormous strides that have been made to enable greater cataloging

efficiency. It is very gratifying to acknowledge this progress and service to LC and the larger library community. Where was Cataloger's Desktop when I was still cataloging!?"

GOING FORWARD

LC and Search Technologies continue to enhance Cataloger's Desktop to meet subscribers' needs, building on the successful migration. Engaging with Desktop users throughout the migration project and incorporating feedback into the design were essential to its success. Iterative software development practices allowed LC to review the application as it was developed. Finally, the choice of current technologies for Desktop 4.0 provides LC with a platform to build on for years to come.

[This article is based upon a presentation given by Bruce Johnson and Derek Rodriguez at the Enterprise Search & Discovery conference held in Washington, D.C., in November 2015 (enterprise searchanddiscovery.com). -Ed.]

Bruce Johnson (bjoh@loc.gov) is senior library information systems specialist, Policy and Standards Division, Library of Congress. **Derek Rodriguez** (droduiguez@searchtechnologies.com) is managing architect, Search Technologies. **Susanne Ross** (sross@searchtechnologies.com) is product marketing manager, Search Technologies.

Comments? Contact the editor-in-chief (marydee@xmission.com).