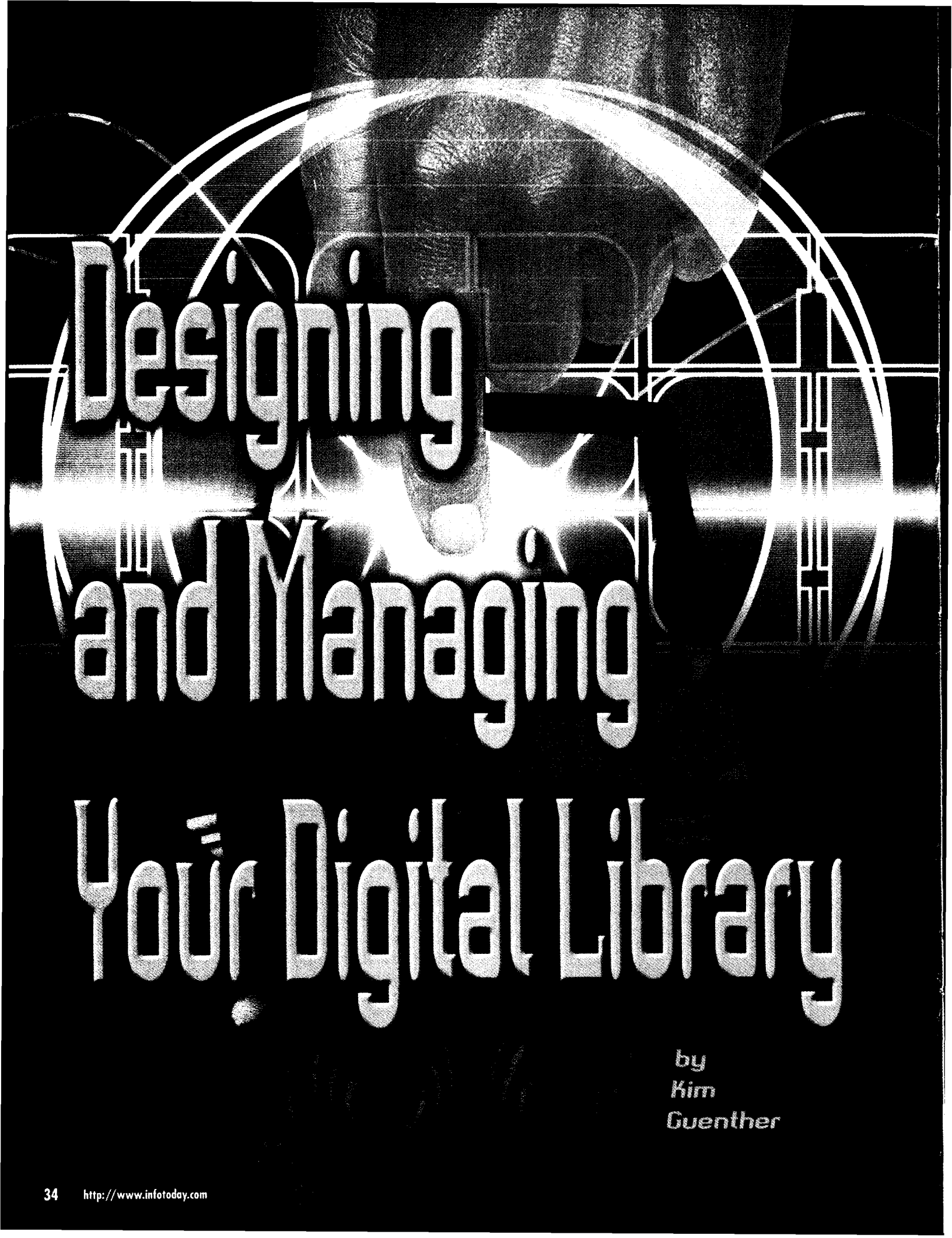

# Designing and Managing Your Digital Library

by Kim Guenther

**The digital library concept requires that librarians be information architects in order to build effective, scalable Web sites to serve the digital demands of patrons.**

Headline news pushed to the desktop; Web browsing from a palmtop; cyber bookstores, flower stands, and auction houses: All these are examples of how the Web has fundamentally changed where, how, and when we do business. "Just in time," a method introduced in the early '80s to manage inventory, is now an expectation of consumers for any business providing information online. "Just in time" applied to libraries is information delivered where you need it, when you need it, and in a format that is useful. This strategy requires libraries increasingly to adopt 24/365 service to keep up with the demands of their patrons for core services, and to offer services through what is often called the digital, electronic, or virtual library or reference desk.

## *Anytime, Anywhere, and Anything Libraries*

Over the past few years, the digital library has evolved from a Web-based information billboard to a vast collection of resources rivaling the library's own physical collection. First attempts to build the digital library integrated little more than the library's catalog, a cadre of subject-based lists reflecting electronic resources culled from the Web, and directional information regarding library hours and location. Core library services were slow to be integrated since electronic journals and books and even online forms were yet to be delivered effectively online. Today the digital library consists of collections and services offered in parallel to the library's own physical resources. The digital library concept, once difficult to define, is more tangible than ever, requiring librarians to be information architects in order to build effective, scalable Web sites to serve the digital demands of patrons.

But how do we build this virtual library from the ground up, replacing bricks and mortar with bytes and bandwidth? How do we re-create that sense of customized services, accessibility, convenience, community, and quality? These characteristics are expected of the traditional library, but are increasingly managed within components of the digital library through Web portals, virtual communities, and techniques to assess consumer needs and quality assurance.

Since digital librarianship and Web design are such large topics, I've tried to cover several subjects that will be important to you as you develop your Web presence. First, since electronic formats and standards are frequent obstacles to developing an accessible digital library, I'll discuss markup languages and metadata. Next I'll talk about developing a community for your users and what that entails in a virtual environment. Finally, I'll touch on ways to maintain the quality of your site, through data mining and patron profiling.

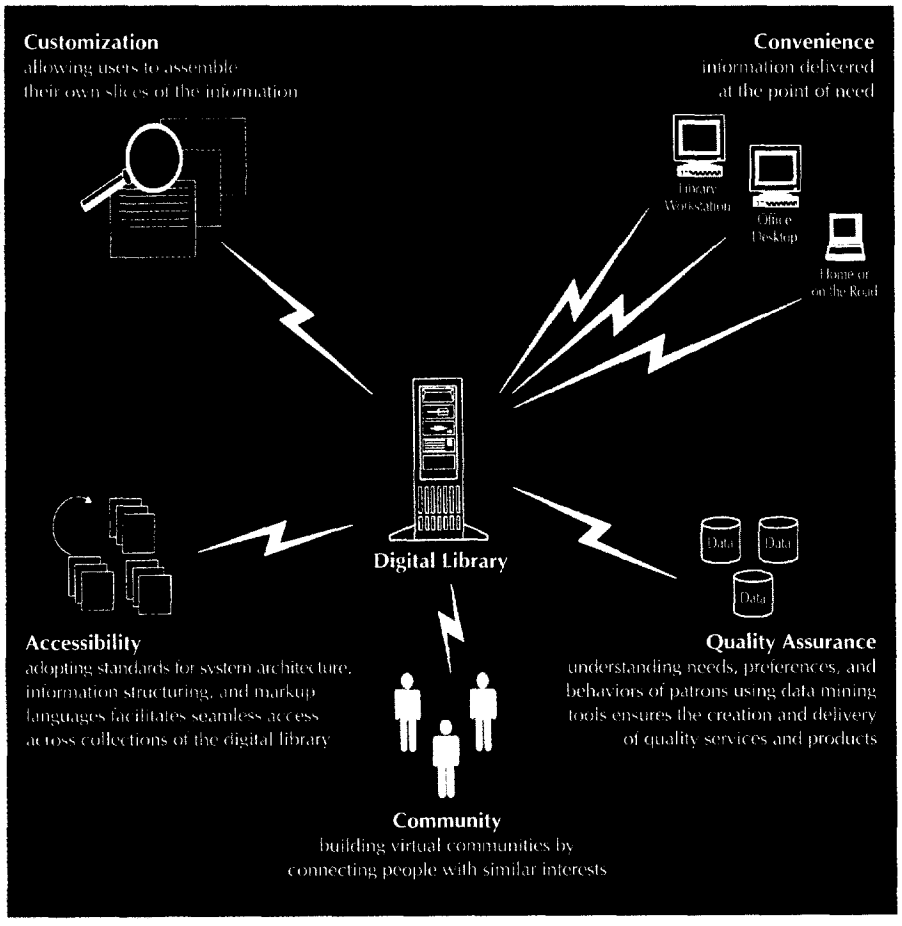


***Figure 1: Characteristics of the digital library***

## *Accessibility Issues: Developing the Standards*

One of the difficulties in developing the digital library is providing all electronic assets in a seamless and integrated format that supports resource discovery across collections. Not only must the system be flexible enough to facilitate sophisticated access by multiple user groups each with their own vocabularies, but it also needs to support identification and management of resources regardless of the resource's physical location. A large-scale research effort by the Library of Congress addresses these same issues as a part of its multiyear Global Digital Library project. The integration of standards

into digital library development is a major focus of this research—assessing standards for system architecture, information structuring, and markup languages, e.g., XML and SGML. On a smaller scale, libraries developing their own digital counterparts should adopt metadata element standards to describe their collections' digital holdings.

**"'Just in time' applied to libraries is information delivered where you need it, when you need it, and in a format that is useful."**

Metadata is data about data. Cataloging is an example of a metadata application providing description and aiding resource discovery within the library's catalog. On the Web, metadata provides description; supports searching through appropriate search engine indexing; facilitates management of Web pages; and also provides information on intellectual property, format, relationship to other resources, and other data elements describing the document. Unlike HTML used to present information, metadata is used to describe the details about the Web document. Metadata is contained with the http header between the HTML tags <head> and </head>. The use of a standard metadata element set, especially the keyword and descriptor metatags, substantially increases the odds of having Web pages indexed properly across search engines. OCLC's Dublin Core, the W3 Consortium's Resource Description Framework (RDF), and Platform for Internet Content Selection (PICS) all represent description standards under development to effectively describe digital objects on the Web.

Here at the University of Virginia's Health Sciences Library, we're adopting a subset of the Dublin Core metadata scheme to streamline both management and access to our Internet and intranet Web collections. A core data set is made available from the library's intranet site, KnowledgeWeb, to all staff members developing the library's Web pages. Dublin Core is also being considered by the library's Historical Collections department, a recent recipient of a $248,000 National Leadership grant from the Institute of Museum and Library Services (IMLS), to describe over 30,000 digital objects making up the Philip S. Hench Walter Reed Yellow Fever collection. The digitization project is part of a larger initiative to develop standards for describing and digitizing collections of historical medical documents, images, and artifacts to make them available to students and scholars throughout the world.

Incorporating a standard metadata scheme provides the necessary structure to support access to the collections from both the library's public access catalog through unique identifiers integrated into the collection level record, or "handles," and through search engines outside the local system. Increasingly, digital libraries incorporate unique identifiers, "handles," for long-term and reliable management of resources on the Internet, independent of physical location. Also called Uniform Resource Names, handles streamline access to local and remote digital collections, thereby providing seamless searching across collections.

## *Personalize, Customize, and Reach Your Community*

Consumers want to feel as though they are being taken care of, that someone is looking out for their interests and needs, and that they are someplace where they belong. When services were exclusive to the library's physical space, it was the library environment itself—the reading rooms, the study carrels—that provided both familiarity and a sense of community. Those with similar interests met and studied together in the library, and librarians provided customized services with explicit knowledge of the patrons' interests through direct contact. As contact with patrons occurs less over the reference desk and more through e-mail and Web-based forms of communication, recreating the tangible qualities of the physical library on the Web requires blending both product design (Web portals and virtual communities) and strategies that assess market viability and quality (customer profiling and data mining).

## *A Destination and a Launching Point All in One*

Initially launched by commercial search engines to entice Web browsing newbies with a "safe haven" or gateway to content on the Web, the Web portal has evolved. Portals are now used in cyber-partnership mega-sites that provide search engines, online shopping, news, reference tools, and communications services including e-mail, chat, and some level of personalization. Although the portal concept has been around for quite some time, providing users with a means to tailor the resources and layout of sites to meet their own needs has led to a recent surge in portal popularity. Libraries serving large and diverse audiences are increasingly adopting the portal mechanism to more effectively serve their users. In doing so, they are allowing users to create their own information pallets from among the library's electronic resources and services.

An example of a library portal is the University of Utah Spencer S. Eccles Health Sciences Library's Web site (http://medlib.med.utah.edu). Eccles Library launched its portal, Personalized Eccles, in 1999 to serve its users both inside and outside the university's Health Sciences Center. Personalized Eccles offers a full inventory of the library's electronic resources. It allows users to assemble their own "slices" of the information, supporting the patron's need for flexibility, individualization, and convenience. The end result is an individualized portal page reflecting the user's own needs and interests. (The Personalized Eccles CGI Perl script is available as shareware by contacting Wayne J. Peay, director of the Eccles Library at wayne@lib.med.utah.edu.)

Using a portal for customized delivery of information in the library is growing. The University of Virginia Health Sciences Library, Virginia Commonwealth University, and North Carolina State University are each developing Web sites supporting portal technology to meet the needs of their diverse user bases.

## *Building Virtual Communities*

Bringing those with similar interests together: It's what Amazon.com does so well, making it one of the most successful businesses on the Web. That success

has little to do with the idea of selling books online and everything to do with gathering information from its users and turning that information back to customers to help them make buying decisions. Where the company couldn't replicate the tangible experience of buying a book, it replaced that experience with other, more valuable services that a regular bookstore can't offer: online reviews, title recommendations based on shopping basket receipts, and customer profiles for future purchases. The net effect is like shopping in a physical bookstore with a few knowledgeable friends. In short, Amazon.com builds community around its services and products. This same idea of building community applies to the library environment serving groups of users with similar interests.

**"How do we re-create that sense of customized services, accessibility, convenience, community, and quality?"**

In 1996, the U.S. West Research & Information Group corporate library developed its own intranet Web portal to streamline and support the research efforts of its patrons/researchers. Through the use of a personal Web page created on-the-fly following a research request, researchers check the status of a request, look at past requests, and tap into other research occurring within the company. The library creates communities of researchers with similar interests by connecting knowledge domains through file sharing and by connecting researchers with interactive chat accessible from personal Web pages.

The customized and personalized approach to portals provides users with a familiar and highly useful Web interface to the library. Extending the individual's sense of comfort and easy access to a feeling of common interest is the key idea behind virtual communities.

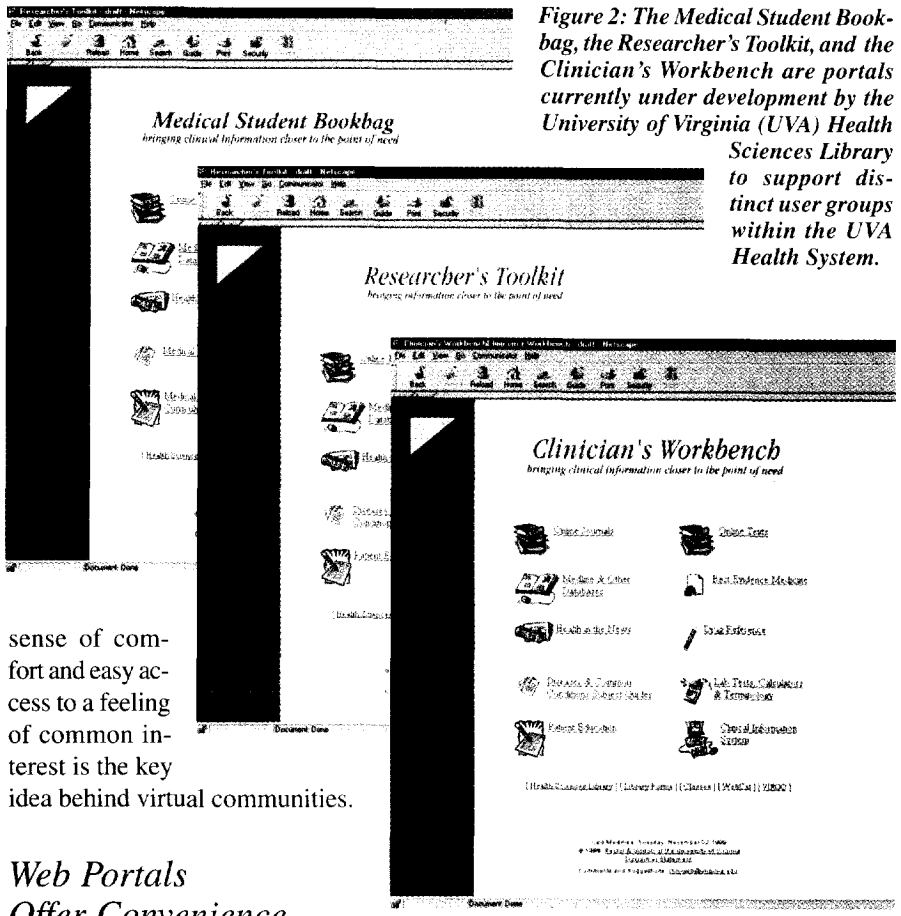


*Figure 2: The Medical Student Bookbag, the Researcher's Toolkit, and the Clinician's Workbench are portals currently under development by the University of Virginia (UVA) Health Sciences Library to support distinct user groups within the UVA Health System.*

## *Web Portals Offer Convenience*

Convenience is one of the most important ingredients of a digital library. In the virtual environment, convenience is measured with regard to accessibility, speed, and responsiveness. Today's networked environment allows libraries to do things unimaginable just a few years ago—namely desktop delivery of services and resources. Document delivery, literature searches, reference, and circulation—marketed and offered as standalone services from specific areas within the library—are now bundled and pushed out as products to the desktops of users. Marketing and delivering services as if they were products offers several benefits, such as streamlining the delivery of services and resources (one-stop shopping), aligning product development with specific user groups, promoting brand recognition among users, and marketing focused on tangible products.

The Clinician's Workbench, a portal page we're developing at the University of Virginia Health Sciences Library as part of our larger Library On Call portal project, aggregates useful services and resources needed by clinicians on the busy floors of the university's hospital and clinics. The goal in developing the Workbench is to make it accessible not only from the library's Web page, but more importantly from workstations on each floor within the Medical Center where clinicians can use its resources for decision making related to direct patient care. Development of the Clinician's Workbench supports clinicians' needs for "just-in-time" delivery of information packaged in an easy-to-use, identifiable library product, a k a digital branding. The Clinician's Workbench is one of several planned portal products geared toward distinct user groups within the UVA Health System. Library On Call and its associated portals are products under development by the UVA Health Sciences Library as part of its larger Digital Library Initiative.

## *Quality Assurance Tools*

Assessing quality is an ongoing process that's integrated at every stage of digital library development, from conducting thorough needs assessments with patron profiling and adopting Web development standards to extracting data from every service exchange after the site is launched. Quality assurance tools must extend across

all products and services, whether they're offered virtually or face-to-face.

Data mining is the process of collecting data to determine both patterns and trends to predict future outcomes. In larger industries, such as health care and banking, automated data mining is used to tap into large data warehouses and "data-marts" to extract knowledge in order to more effectively achieve business goals. Although the scope of data collected is somewhat smaller than in big business, the need to gather data in the library is just as important. Gate counts, internal and external checkout, document delivery/interlibrary loan, and reference requests all represent activities from which data is gathered in the physical library. Data- or statistics-gathering is a daily, weekly, and monthly operation in most libraries, and represents a tangible record of activities.

Before the digital library, statistics-gathering focused on services delivered in the library. With so much information now online, the information glut that librarians navigate on behalf of patrons is also a data glut, and requires us to extract knowledge from data produced in both our physical and virtual environments. Analyzing the server's log files is now an integral part of librarywide statistics-gathering. Data mining provides the tool to tie both physical and virtual libraries together to recognize trends and patterns among users. Understanding the needs, preferences, and behaviors of patrons ensures the creation and delivery of quality products and services.

## *Determining Needs: Ideas for Patron Profiling*

Patrons are the most important ingredient in a successful digital library. Partnering with library patrons is done through product-specific focus groups and patron profiling. Both are done to assess needs of specific user groups with the goal of tailoring services and products more directly. The Library On Call portal project we're developing will require significant amounts of patron involvement at all stages of product development. Patron profiling is an integral part of the product feedback loop and influences future releases of the product beyond version 1.0.

Although controversial, another way to gather profile information on Web site users is through cookie files. A cookie is a series of textual statements contained within the http request header passed back and forth between the server and the client's browser (if cookie-enabled). The Web server collects information about a site visitor and then passes this information back to the client's browser where the information is stored as a data file on the client's hard drive. This data or cookie file, usually less than 4 KB, is made up of information gleaned from the visitor's last visit to the site. The cookie is passed back to the server during subsequent visits, and information is customized accordingly.

On commercial Web sites, customization may include targeted advertising based on a user's interests or layout preferences based on the user's browser capabilities, and targeted product information pushed to the consumer during subsequent site visits. Gathering information on Web site users, also referred to as online profiling, has recently come under scrutiny by the Federal Trade Commission (FTC), which is assessing whether profiling by commercial Web sites is an invasion of privacy.

**"... providing users with a means to tailor the resources and layout of sites to meet their own needs has led to a recent surge in portal popularity."**

The public reaction to cookie files is mixed, especially in a library environment. Some believe it is an invasion of privacy; others believe the cookie is just another marketing tool that is gathering profile data in order to better target and tailor services to the consumer. Regardless, it may be a tool worthy of exploration as a way to expedite data mining across multiple systems within the library, and to get a better understanding of a user's current and future site visits. Data mining in the form of cookies is an effective way to market services and products one-to-one and one-to-many. This type of targeted marketing, also called push or permission marketing, is achieved with client-focused data mining occurring at all levels of digital library development.

## *The Viable Digital Library*

Making library services tangible by "productizing" clusters of services through the digital library provides convenience, customization, community, accessibility, and quality. Methods for making these products and services more tangible in a virtual environment will continue to evolve as tools to manage, identify, and structure content in digital form become more sophisticated and as library users become more Web-savvy. Once only discussed in the abstract among library and information science professors, the digital library now represents a viable alternative to the physical library's products and services, while still supporting its original mission to identify, organize, and disseminate information to users.

*Kim Guenther is the Internet/clinical information services coordinator for the University of Virginia Health Sciences Library, where she oversees the development of the University Health System Web site, leading an interdisciplinary team that provides support to over 150 Health System departmental Web subsites. She also manages the development of the library's Internet and intranet Web sites, and related clinically based Web projects serving users throughout the UVA Health System. Guenther holds an M.L.S. from the University of Maryland and has over 7 years of experience developing and managing large-scale Web sites for both nonprofit and for-profit organizations. Her e-mail address is guenther@virginia.edu.*

### Further Reading

Davis, Stan and Christopher Meyer. *Blur: The Speed of Change in the Connected Economy*, Ernst & Young Center for Business Innovation, Reading, MA: Addison-Wesley, 1998.

Dyson, Esther. *Release 2.0: A Design for Living in the Digital Age*, New York: Broadway Books, 1997.

Stout, Rick. *Web Site Stats: Tracking Hits and Analyzing Traffic*, Berkeley, CA: McGraw-Hill, 1997.

Tapscott, Don. *The Digital Economy: Promise and Peril in the Age of Networked Intelligence*, New York: McGraw-Hill, 1995.