How Do You Get Started Building a University Web Portal?

As the meeting room began to fill with EDUCAUSE 2000 attendees, you could almost feel the hearts of the more technically inclined sink as the announcement reverberated across the room: "The slides for this presentation are available online at the URL on the screen, but ... *it only works with Internet Explorer*". This was probably not the best way to begin a presentation on developing Web portals.

All you had to do was look at the conference program to see that Web portals were the hot topic at EDUCAUSE this year. There were breakout sessions, corporate presentations, and vendor exhibits all willing to provide information as to why and how your institution should go about building a portal. Howard Strauss, Manager of Academic Applications at Princeton University, presented a fast-paced and often, humorous session on what a university Web portal should be.

Strauss began by citing several examples of how the term "portal" has been misused and then proceeded to share with his audience his image of a true Web portal. While often noted as good examples of portal technologies, such Web sites as Excite and Yahoo! are horizontal in nature (users have access to the same public data, but from one starting point). A university portal, if implemented correctly, should be vertical in nature as well (having access to public, institutional, and even private information from a single starting point). This means that, if an institution were to connect its portal to student information, registration, and other internal systems, the resulting portal could be a highly customizable, information-rich environment that is relevant to the institution's community.

Strauss defines a portal as a "Customized Personalized Adaptive Desktop" or CPAD. A true CPAD is "a knowledge portal" that mines enterprise and global information sources and enables collaboration within the community. Strauss went on to further explain the components of the term CPAD:

- **Customization.** Customization occurs when you authenticate to the portal. You should only have to authenticate once to gain access to all available tools and resources. The more the portal knows about the user, the better its customization will be. This customization dimension of a portal should determine what you see the first time you log in and then change what you see as your role, responsibilities, cohorts, and functions change. The portal should also customize depending upon the device that is used to access it.

- **Personalization.** A user should be able to personalize a Web portal. Ideally, one should be able to subscribe to various channels (news, weather, sports, and other specific types of information), set application parameters (stocks to track, e-mail and chat defaults, reports and other individual-specific settings), and format the portal page itself (by changing colors, style, font, and other layout settings).

- **Adaptive.** A portal should adjust to the user's schedule and workflow. For example, budget development applications should only be visible to a person having budgetary responsibilities during the appropriate timeframe. The portal should also recognize the links one uses most often, displaying them in a prominent location while eventually hiding links that have not been used in a while.

- **Desktop.** A portal should resemble the contemporary PC desktop. It should hide the operating system from the user while providing access to files, applications, the Web, e-mails, LANs, and anything else that the user commonly accesses via a computer.

So, what makes an efficient vertical portal? According to Strauss, most of what you need should be on the first page and everything should be available having logged in only once. Do not bookmark other Web
While there is no doubt that Strauss did an excellent job explaining exactly what a true portal is, what its components should be, and how those components can interrelate with one another; he may have missed his original mark a bit. The title (and abstract) hinted that this presentation would be more technical in nature. Someone attending this session after examining the conference program may have expected advice or case studies regarding development platforms, application environments, standards, institutional and administrative support, and maybe even design and development processes.

In the end, this was an informative session and quite valuable to attendees just beginning to think about developing their own university portals. By the way, the URL on the screen was: http://www.princeton.edu/~howard/slides/portals_files/frame.htm

Digitizing Research Collections for Access

Education is evolving. One of the most fundamental changes that academia is facing is the "digital evolution" of our educational resources  the library being the most important resource affected. While there are academics that want to hold onto pulp and paper exclusively, the race is on to digitize library collections. EDUCAUSE 2000 was ripe with sessions covering issues such as whether to digitize or not, methods of digitization and cataloguing, what to digitize, pitfalls to avoid when digitizing, and even the impact of digitizing collections on the existing physical university library.

Elizabeth Shaw (Technical Projects Manager, Digital Research Library and Visiting Lecturer, Department of Library and Information Sciences) of the University of Pittsburgh presented what may be considered the most insightful session pertaining to collection digitization. Her full-day, pre-conference seminar covered a lot of ground and was relevant for administrators, project managers, and programmers alike. With her MLIS from the School of Information at the University of Michigan and depth of experiences with digital library projects, Shaw conveyed a fresh and most unique perspective on the subject of digitizing collections.

Shaw began the discussion with a very basic, but key, question: Why create a digital collection? The answers may vary greatly but, in any case, it is important to be sure that the justification be sound. For example, an institution may want to keep a collection accessible while limiting physical contact. However, by digitizing the collection and placing it online, the institution may inadvertently create a new demand for the originals on the part of individuals who did not previously know that such items were in existence. While there are many valid reasons to digitize a collection, the commitment in terms of time and money can be substantial and, therefore, must be considered carefully. Digitization of text can cost from $1 to $7 per page, while images may cost as much as $20 per page. These figures do not include the cost of indexing or describing the materials being digitized. Cataloguing can account for as much as two-thirds of the entire cost of a conversion project. Also, according to Shaw, conversion usually equals approximately one-sixth the total cost of maintaining a digital collection over the first ten years. She also highly recommends that anyone embarking on a digitization project understand the implications of the institution's existing and planned infrastructure and support, the preservation needs of the original materials, the research needs of the anticipated audience and the institution itself, as well as all relevant intellectual and property rights and privacy issues.

Extending use or availability is one of the most common justifications for digitizing a collection. Digitization alone will definitely not increase accessibility. Shaw notes that the amount and quality of the metadata (or "data about data") collected on an item or collection directly correlates to its usefulness. The more metadata collected, the more sophisticated user access can be. There are three types of metadata; all three are important to the final digital collection. Descriptive metadata is perhaps the most obvious. This metadata; contains contextual information and finding aids that are useful to eventual users of the digital
collection as resource discovery tools. Administrative metadata is information that allows a repository to manage and/or maintain the collection. Some examples of administrative metadata are scanning resolution; date scanned; version/edition information; file name, location, and format; and even ownership, rights, and reproduction information. The third type of metadata is structural in nature. Structural metadata is relevant to an item's presentation to a user; examples include navigational clues (how an object relates to other objects); available formats; and information like native pagination, document structure, and sequencing. Shaw reminded her audience (or, perhaps, it was a warning) that one librarian's metadata is another's data. With this in mind, anyone embarking on a digitization project should examine closely the standards (such as EAD and Dublin Core) for capturing and storing data that are available and evolving. Shaw gave a final list of things to consider when deciding upon a metadata scheme: the uniqueness of identifiers, the consistency of the data, and the trade-offs that will have to be made. These trade-offs usually revolve around the fact that the more metadata that you collect the more costly the project will be. One must also remember, however, that returning to collect more metadata later may be more costly than collecting the information during the original digitization process.

Once the decision has been made to digitize a collection and a data collection scheme is chosen, one must then decide upon what digitization method is most appropriate. The method of digitization will quite often depend upon the item being converted. While many texts may be digitized using an OCR (Optical Character Recognition) technique, others will require transcription. If transcription is necessary, it is often cost effective to double- or triple-key the text for accuracy rather than review a single transcript. With images, one may want to examine the item's anticipated use. A bi-tonal representation of an image may work well for some applications, while grayscale or full-color scans may be appropriate for others.

In terms of both digitization and data collection, Shaw's rule is simple: Aim for the highest standard that you are capable of achieving and maintaining when considering your financial, system, support, and other resources.

Elizabeth Shaw has organized an impressive list of resources for anyone involved in digitizing collections or who may be working with a digital library project. This extensive document can be accessed on the World Wide Web at: http://www.pitt.edu/~ejshaw/

**Brian T. Watson** is a Management Information Specialist with The University of Georgia's College of Agricultural and Environmental Sciences in Athens, Georgia. E-mail: bwatson@arches.uga.edu