Web Sites as Recordkeeping & “Recordmaking” Systems

Web sites are important sources of organizational records; not properly capturing such records in trustworthy recordkeeping systems is risky

Rick Barry
On March 2, 2004, the Washington Post broke a story concerning lead contamination in the District of Columbia’s drinking water. Neighboring Arlington County, Virginia shares the same source, and the article noted discrepancies in the county’s public-facing Web site.

A follow-up front-page story the next day stated:

“Arlington County officials began recommending yesterday that pregnant women and young children drink only tap water that has been flushed or filtered, after preliminary tests of water in eight homes showed that five had elevated levels of lead … As late as mid-afternoon yesterday, the county’s Web site carried the headline ‘Lead Not a Concern in County Water.’ The Web site did not mention that, on February 23, the county’s Public Works Department quietly began sampling water in Arlington homes built before 1988, the last year lead solder was used.”

Feeling sure that the contamination problem did not affect Arlington, officials had decided to leave the “Lead Not a Concern in County Water” announcement on the county Web site until they received results from a special testing program.

The story illustrates the importance of Web sites as sources — possibly the only sources — of many organizational records and the risks of not properly capturing such records in trustworthy recordkeeping systems. The story was picked up in local TV news coverage and received so much publicity that the U.S. Congress held hearings on the subject. It became the source of daily reporting by Post investigative reporters for the entire month, exposing accountability issues in agencies at the federal, regional, and local government levels, including Web site representations and e-mail communications.

This is one of several recent news stories in which, to the embarrassment of the organizations involved, journalists have reported on the sudden and controversial alteration or deletion of Web content in apparent attempts to “change history.”

The fact is that Web sites produce official representations to the public. Plainly stated, Web sites make records, but they do not keep records in ways that match up to sound recordkeeping requirements. Chief executive officers (CEOs), attorneys, chief information officers (CIOs), auditors, and content, records, and other information managers: Beware.

Web Sites as “Recordmaking” Systems

The use of Web-based e-business applications on Web platforms is almost ubiquitous in the private sector. E-government applications (including Web-based) have become increasingly prevalent in the public sector as well, with mandates at various government levels to implement citizen access to e-gov services in the 2003-2005 timeframe.

Moreover, citizens are demanding such access. A 2004 Pew Internet & American Life Project survey report “How Americans Get in Touch with Government” found that 97 million adult Americans (77 percent of Internet users) participated in e-gov in 2003 by visiting Web sites or e-mailing government officials to conduct transactions (e.g., paying bills, obtaining licenses), obtaining information, or solving problems. This reflected a growth of 50 percent from 2002.

“E-Gov Alliance” is a collaborative effort among several communities of Redmond, Washington — Bellevue, Bothell, Issaquah, Kenmore, Kirkland, Mercer Island, Sammamish, Snoqualmie, and Woodinville — to provide a unified approach to automated building processes and services. MyBuildingPermit.com is a model example of Web-based e-gov at the local government level. This multi-jurisdiction system allows local citizens to apply for, pay for, and receive electrical, mechanical, plumbing, and other permits — all Web-based public records — for each of the participating cities. (Editor’s note: See “Why Records Cooperatives?” pg. 49.) The customer-friendly system distributes system costs across participating governments, significantly reducing their individual total cost of ownership (TCO), a primary concern of CIOs. Bellevue is currently spearheading another project to provide content management services, including trustworthy recordkeeping capabilities for interested alliance members.

Just as organizational enterprise resource planning (ERP) systems, call centers, e-mail, and instant messaging systems are important producers of electronic records, so are organizational Web sites, intranets, extranets, and other emerging information and communications technologies (e.g., instant messaging, Web logs or “blogs,” agent and virtual reality technologies) when used for business purposes. Blogs are viewed by some organizations as being more effective than Web sites for conducting public information and crisis management activities.

Observant archivists and records managers have been aware of the mounting Web-records issue for a few years through such sources as the National Archives and Records Administration (NARA), research funded by National Historical Publications and Records Commission (NHPRC), an independent research funding arm of NARA, and related research and implementation work in several other organizations. A NHPRC
study of several federal and state government organizations by Charles R. McClure and J. Timothy Sprehe found many disparities where dynamic Web site contents (records) were more up-to-date than the “official” records. For example:

In Michigan, the State Administration Board is putting official minutes of meetings up on a Web site, knowing that no print version of the minutes exists … Although such “Web-only” records are being created by many states, the prevailing opinion is that most information on state Web sites is … unimportant from a recordkeeping standpoint … In contrast … federal agencies exhibited consensus that informational materials were appearing on Web sites that qualified as official records. The materials in question were “original” … not copies of materials available in some other medium such as paper … [and] not being transferred into existing agency recordkeeping systems.

Most recordmaking systems are not sufficiently robust to preserve the principal characteristics of records. Moreover, they are not necessarily recordkeeping systems.

Web Sites as Recordkeeping Systems

Although recordkeeping laws and standards do not always explicitly address electronic records, virtually all recognized definitions of the word “record” embrace or do not exclude electronic records such as Web content. ISO 15489 Information and Documentation – Records Management does address electronic records. It “applies to the management of records, in all formats or media, created or received by any public or private organization in the conduct of its activities.” It further states, “records created in the public domain, such as the World Wide Web, require a broad range of contextual information.”

As with other digital records, some Web-based records will be of long-term evidentiary, secondary information, research, historical, or collective-memory value to the organization. Those will require a “continuing” or “permanent” (i.e., indefinite) retention period and rigorous architectural and technological platforms to survive multiple software version updates and new system migrations.

ISO 15489 defines records system as an “information system which captures, manages, and provides access to records through time.” A trustworthy recordkeeping system captures, protects, preserves, and provides ready access to records, possibly for many decades or indefinitely, and serves as the primary source of business documentation. In addition to a record’s actual content, it preserves its structure, business context, and association with other like records. It preserves a record’s authenticity (it is what it purports to be), reliability (accurate representation by a knowledgedgeable source), integrity (complete and unaltered) and usability (can be located, retrieved, presented, interpreted, and understood over time). Achieving this level of trustworthiness requires more rigorous functionality than most automated systems possess.

The main practice in recent years to address electronic records (beyond printing them out) has been integration of a DoD 5015.2-certified records management application (RMA) with an existing enterprise document management system (EDMS). This has not

It is essential that policies for Web publishing be formulated by a group representing key stakeholders that address Web mastering, content management, and records management requirements.
always turned out to be as effective as advertised. Most RMA/EDMS integrations were unable to take account of Web records without adding still another layer of tricky software integration.

By contrast, Bellevue and Arlington County governments recently opted to procure enterprise content management (ECM) systems that were also certified as 5015.2-compliant. Bellevue’s city manager took the further important step of officially endorsing ISO 15489 and DoD 5015.2 as regime and software-level enterprise city standards. At present, there is no certifying authority for 15489 as there is for 5015.2, although Standards Australia is developing a compliance suite against which organizations adopting 15489 may be assessed.

While it is still early, the Bellevue and Arlington approach of implementing a recordkeeping ECM has the potential for significantly reducing TCO, making for better capture, access, and management of records and other documents in digital, paper, and other analog forms while being more attractive to IT, archives, records management, and finance.

The timeline for seeing more than a few examples of this kind of implementation approach will depend in large part on how quickly the CEO and IT communities pick up on two principles:

- Legacy records and increasing volumes of current and future electronic records are major elements of the organization’s intellectual capital.
- Web sites are among the key organizational recordmaking systems that are not recordkeeping systems and that place organizations at risk for what, in Information Nation, Randolph Kahn and Barclay T. Blair call TCF (total cost of failure) or the cost of compliance failure.

Web Site Records Management Issues

Web Content Management

The term “content management” was initially limited to the management of Web publishing. This has changed as the understanding of ECM has matured to include all enterprise content and with advances in ECM technology that make it possible to do that. This approach is exemplified in the Bellevue and Arlington examples. Because a high percentage of enterprise content is records, it is essential that the management of content/records be integrated at one or more levels – organization, policy, systems, standards, procedures, and training.

To illustrate using the Arlington County example, Web site style, content standards, and publishing were being managed under the county library director while content creation responsibility was distributed at the department level. The county’s CIO understood the relationships between enterprise content and records management and thus saw the importance of integration at the ECM system level. But because there had not yet been adequate integration at the other levels, there were no policies or procedures requiring preservation of Web records in a trustworthy recordkeeping system. Consequently, when the lead-contamination story broke in the Post and the contentious Web content (“Lead Not a Concern in County Water”) was removed and replaced, no official copy of the original announcement was retained in any form.

Organizational Management Rules

The NHPRC study recommends that organizations provide three separate but closely coordinated roles in the management of their Web sites:

- **Webmasters** – manage the information technology aspects of Web sites
- **Content managers** – create and manage the informational content of Web site postings
- **Records officers** – ensure that official records management and archival responsibilities are carried out

Recognizing these roles for Web sites and establishing responsibilities for each are essential steps toward risk reduction through coordination of content and records management. Depending on the size of the organization and other staffing considerations, content creation may be centralized or decentralized to operating units. Moreover, content creation responsibilities may change under certain circumstances. Where content creation might normally be decentralized, in crisis-management circumstances it may be elevated to a higher, centralized, multidisciplinary authority and revert back to the routine when the crisis is over.

In the Arlington County example, the “Lead Not a Concern in County Water” announcement that the Post cited had been created by the content manager in the Department of Public Works (DPW). When the Post noted that Arlington had undertaken special drinking water tests while that announcement was still on its Web site, the page was immediately removed from the DPW home page. Responsibility for information releases on this subject shifted to the public information office under a multi-disciplinary team that included managers from DPW and the health and legal departments. The team removed the DPW page, replaced it with content on the county home page advising citizens of testing results it had received the same day that showed elevated lead levels in five of eight residences, and posted precautionary measures. The case illustrates the risky nature of withholding information that is contradictory to Web-published information, especially in government organizations where such information is easily leaked and can become the source of embarrassment and citizen cynicism when revealed.

Whether content creation is centralized or decentralized, Web publishing, standards, and the look and feel of pages throughout the site should be centralized to maintain the organization’s “branding” so that public users will know that they are still browsing the same organization’s Web pages.
McClure and Sprehe noted numerous cases of multiple domain names within the same agency, complicating difficulties in coordinating both Web site content and style and leaving public users uncertain about relationships (if any) of one site to another.

Where the organizational culture values its records as prime intellectual assets, it may place Web publishing, standards, and recordkeeping functions under the CIO. If the organization values its records only as a means of reducing risk, it may place the archivist and records manager function under the chief counsel. However, these should not be seen as mutually exclusive value sets.

The CIO model is widely used in the federal government and elsewhere with varying degrees of success. In some cases, this approach has been seen as a way to better integrate records and compliance management and to “hardwire” them into the organization’s information and technology architecture. In other cases, the CIO has used the integration to cherry-pick positions out of the records group to further build the IT group.

Web Policy Creation

However Web content is organized and managed, it is essential that policies for Web publishing be formulated by a group representing key stakeholders that address Web mastering, content management, and records management requirements. Stakeholders may include those responsible for content management, archives and records management, libraries, IT, legal, auditing, and public relations.

Where Web content is decentralized, Web policymakers should also consider the desirability of procedures for elevating topic-specific content creation to a centralized multi-disciplinary management team during crisis situations. Like all coordination, this may result in slower response times during rapidly changing events. What it loses in time, however, it likely gains in more accurate information that takes into account the expertise of key stakeholders.

The rapid uptake of e-business and e-gov applications using Web publishing systems has outpaced the ability of many organizations to properly manage the records produced in these systems.

Web Content Dating, Removal, and Destruction

Web content dating, removal, and destruction are among several Web site standards that must be addressed. They are open to considerably different treatment by different content managers in ways that can have serious recordkeeping consequences.

Some Web sites do not date their content. Some carry the current date on the home page only. Others use different conventions on different pages. Individual content managers may use different conventions for similar announcements. To illustrate, again using the Arlington County example, its Department of Environmental Services (DES) Web page FAQ on “Drinking Water Information” (www.arlingtonva.us/Departments/EnvironmentalServices/wepd/waterops/EnvironmentalServicesWaterops.aspx) is undated as this publication went to press. As the FAQ was revised several times during the lead-contamination incident, it probably should have been clearly marked with correct “update” or “revised” dates for concerned citizens visiting it daily.

Another DES page regarding Hurricane Isabel (September 2003) (www.arlingtonva.us/departments/EmergencyManagement/emergency/EmergencyManagementIsabelWater.aspx) showed the date and time the viewer opened or refreshed the page. However, that date is labeled as “updated” even though it is an unchanged, year-old announcement.

Other pages follow the same practice but label the dates “revised.” Perhaps it
simply reflects a lack of standards or it is to give the appearance of being updated on a daily basis, but the practice both misrepresents reality and creates higher public expectations than can be met. It is also inconsistent that a year-old emergency announcement would remain on the Web site while the controversial “Lead Not a Concern in County Water” announcement would be removed and destroyed without retaining a copy. Policy should require appropriate, consistent standards for such matters as content/page dating, removal, and destruction. These considerations are essential to proper Web site recordkeeping, as are the appraisal and designation of Web-site disposition management schedules, preferably through a hands-off archival authority.

Final Analysis: Web Content Is a Record

The rapid uptake of e-business and e-gov applications using Web publishing systems has outpaced the ability of many organizations to properly manage the records produced in these systems. Often this is coupled with a lack of appreciation in the executive corridors that Web sites even produce records. So long as this technology is used for business, customer-facing, and public-facing purposes, the content and transactions on such sites constitute organizational records and therefore must be captured, preserved, and managed into paper-based or electronic records systems. Most such applications are adopted to reduce paperwork, and some include multimedia content not amenable to recording on paper.

For most organizations, this means that integration of Web content and electronic records management is essential. Failure to do so puts the organization at considerable legal, regulatory, and even ethical risk and opens it to alienation of its client and public base. Moreover, it robs the organization of one of its most precious assets — hard-earned and well paid-for institutional memory.

Rick Barry is a management consultant and Principal of Barry Associates, a consulting firm that specializes in information management, information technology, and electronic archives and records management. Barry is content manager for www.mybestdocs.com and a co-founder of OpenReader™, a cooperative project to create an open, next-generation software built on XML and related open standards for reading multimedia and digital publications and for long-term preservation of, and continued easy access to, multimedia digital documents. He may be contacted at RICKBARRY@aol.com.

References


