NHPRC ELECTRONIC RECORDS RESEARCH FELLOWSHIP PROGRAM
2005-2006
Project Proposal

Content Management Systems: Who Makes the Rules?

David Mitchell and Jill Katte
Duke University Libraries
University Archives

Executive Summary
Confronted with the task of managing ever-increasing amounts of web-based information, educational institutions like Duke University are adopting enterprise content management systems (CMS) to help manage these electronic records on a large scale. The use of CMS in higher education presents institutional archivists and records managers with both opportunities and challenges in ensuring that CMS-based web records are scheduled and preserved correctly. If funded, this project will examine the use of CMS at educational institutions around the country, and will gather data from information technology professionals (specifically, CMS administrators) and archivists on CMS usage/planned usage, CMS applications that are being employed to control versioning and web record life cycles, rules that are in place to guide the life cycle of web records, and how those rules are established. In addition, the project will examine in depth the CMS recently put to use at Duke University (ContentXML) and develop best practices for transferring certain web content to an archival environment.

This project will contribute to the management and preservation of electronic records, specifically web records, by:

- Enhancing understanding of the information technology community about the importance of preserving web content for regulatory, administrative, or archival purposes;
- Enhancing archivists' and records managers' knowledge about technological tools and processes available to them for identifying, managing, and preserving web records;
- Increasing the dialogue between web record creators, archivists and records managers, and CMS administrators so that the gap in the archival record gradually shrinks;

- Identifying applications that are part of a CMS that can be utilized to tag or harvest archival web content;
- Identifying rules that should be in place to control web content versioning and life cycle management; and,
- Determining whether those rules are arbitrarily devised by users or CMS administrators or by records professionals via retention and disposition schedules/guidelines.

Background

The widespread use of the World Wide Web and Internet has changed the way society conducts business, amuses itself, buys goods and services, and carries out many of the routine, day-to-day tasks once reserved for paper forms and telephone calls. Countless websites cater to the varied tastes and preferences of almost every citizen. To that end, one can go online and visit sites focusing on antiques, bands and music, every sport imaginable, all types of humor, intellectual endeavors, auctions, vacation destinations, and workplace safety. The sky (and bandwidth) is the limit.

At educational institutions across the country, websites have become the de facto method of publishing course catalogs and other information, registering for classes and checking grades, and carrying out tasks that only recently were paper-based. Websites represent a technology that students, staff, and faculty assume will create a seamless interface with the world as they know it. The technology is not always perfect and not all web applications are destined to succeed. However, the answer to most information-related dilemmas seems to be web accessibility. How many times have we visited a site only to encounter obsolete information or broken links?

But the prevailing web presence of university records creators has created a new challenge for archivists and records managers at these institutions: how should web records that were solely paper records less than ten years ago be managed, scheduled, preserved as archival content, and made available to interested parties? Web records cannot be placed neatly into acid neutral folders and placed in a box on a shelf. To worsen the problem, web records are fragile and can be changed effortlessly by a webmaster. Archivists and records managers know that detailed metadata must exist if web records are to be captured and preserved in digital repositories, but will that capture occur if it must be done manually? The preservation of web records is really the tail end of the problem. At the front end is the issue of understanding the content of web records, how that content is shared

and altered by users, and how random editing is done without consideration of versioning, the integrity of the original content, or the implications of losing what was once an official site. Often the permanent eradication of unique content is a fait accompli, but this dangerous action impacts the overall structure of a site and weakens its context for preserving a trustworthy web record.

One potential solution for better managing the content of websites is to employ a content management system (CMS). An enterprise CMS can have several benefits. It can allow users with different skill levels to create and update content. It separates the content and the style of the website so that one can be changed independent of the other. Content can be shared more readily in a CMS, and the content's life cycle can be controlled with workflow and versioning tools. Other applications within a CMS make it a functional, flexible, and deployable method for securing intellectual and physical control over web records.

In some regards, however, a CMS can heighten the obstacles for archivists and records managers who are responsible for preserving web records with long-term value and promoting the efficient and compliant management of university records in all formats. Certain questions that must be asked by those groups, and issues that must be addressed by CMS administrators, often are brought to the table after the fact. That is, the content has changed multiple times over, or has been purged from the CMS based on users' preferences, without regard for preserving that which is archival. As the Internet and websites become the preferred environment for distributing annual reports, handbooks and other publications, or policies and procedures, significant gaps in the archival record can form as content is altered without appropriate capture.

Methodology

With this project, we strive to answer the following questions:
- How will the implementation of a CMS affect electronic records management at Duke University and similar institutions?
- Have other institutions implemented an enterprise CMS to control and manage information within their websites?
- Which CMS solutions are they using or planning to use?
- Do these various systems provide means for content creators to produce metadata determining the life cycle of electronic records? If so, are these metadata tools being utilized?

- Are content creators basing versioning or disposition decisions on established institutional records schedules? If not, how are the decisions being made?
- Are CMS administrators collaborating with archivists and records managers to create schedules for CMS-based electronic records?

To find the answers to these questions, we plan to conduct a study of institutions with the "Doctoral/Research Universities -- Extensive" and the "Doctoral/Research Universities -- Intensive" Carnegie classifications, which include about 250 large and medium-sized colleges and universities. Each of these institutions potentially has significant amounts of electronic records content within their public web portals.

We intend to develop two parallel web surveys, one for each institution's web content managers or CMS administrators, and the other for each institution's archivists and/or records managers. By conducting parallel surveys, we hope not only to determine current CMS practice at Duke's peer institutions, but also to gauge the level of communication and collaboration between institutions' website development units and their archives and records management programs. Duke University provides access to ViewsFlash by Cogix, a web-based survey authoring and data collection tool, which we will use for this project.

A graduate student intern will assist us in gathering background and contact information for CMS or website administrators, as well as archivists or records managers, at each institution in our sample. In addition, the student will assist us in analyzing the data that is collected through the survey to help us draw conclusions about the use of CMS in our sample. Finally, the student will help us create a web site to summarize our project, disseminate our results and recommendations, and encourage communication between the archives and records management community and the IT community.

Outcomes
Ultimately, we propose to evaluate current institutional CMS-based electronic records practices in order to recommend best practices for the ongoing and proper management and ultimate preservation of website records. Those best practices will be developed not only for the archives and records management field, but also for the higher education information technology field.

After the project has concluded, we plan to share our findings with a wide range of researchers and practitioners by submitting articles to professional and scholarly journals; presenting our results at conferences such as those sponsored by the Association of Records Managers and Administrators, the Society of North Carolina Archivists, and the Society of American Archivists; and by creating a project website. We strive to disseminate our findings not only to the archives and records management field, but also to reach out to the higher education information technology field. We hope that by conducting this study and disseminating our best practice guidelines while many of these institutions (including Duke University) are still in the planning and early-adoption stages of CMS, the loss of valuable web records can be slowed.

On a broader scale, this project represents an opportunity to increase communication and collaboration between two groups who desperately need one another's help if the fragile content of websites is to be viewed by our successors. We hope this study will open the door for increased understanding, knowledge, and dialogue about the electronic record keeping objectives of two groups of practitioners.

Project Timeline and Benchmarks
July -- August 2005
Co-investigators obtain Duke University Institutional Review Board certification.

Obtain contact information at "Doctoral/Research Universities" for CMS administrators/website content managers and archivists or records managers. If an institution does not have an established archives or records management program, contact information for manuscripts and special collections librarians will be collected. This contact information will form the "invite list" for the project's online survey.

August – September 2005
Develop questions to include in parallel web surveys to be sent to approximately 250 "Doctoral/Research Universities." The short series of questions will be developed in parallel to gather responses both from CMS administrators or website content managers and archivists/records managers. Both groups will receive the survey so that conclusions can be drawn about the level of communication the two groups undertake in order to manage and/or preserve web content. Survey questions will focus on whether a CMS is in use or being planned; whether the CMS contains

a workflow tool that allows for setting content expiration or archiving dates and, if so, whether it is being used; and what bases form the date setting – user discretion or formalized retention policies.

Build surveys using ViewsFlash survey software tool available through the Duke University Office of Web Services. The surveys will be designed to gather predominantly qualitative data. It is feasible that quantitative data will also be collected that answer questions associated with the amount of web content (number of files, size of files, etc.) both being managed and preserved.

Interview and hire a graduate student for a one-semester (Fall 2005) appointment to collect incoming response data and assist co-investigators with data analysis and presentation. Student should be familiar with basic archives and records-related principles and web and electronic records in general. Some knowledge of statistics and statistical software packages like SPSS or SAS preferred.

September – October 2005
Send out e-mail requests to "invite list" and solicit responses to survey. Respondents will be asked to complete the survey prior to October 31, 2005.

November – December 2005
With assistance from graduate student, analyze data from respondents and identify several that are candidates for on-site follow-up interviews. Proposed criteria for on-site interviews include use or planned use of a CMS, use or planned use of workflow or similar CMS tool, an active archives and records management program (s), and willingness on the part of the institution to participate in follow-up discussions.

January – March 2006
Travel to selected institutions to conduct on-site follow-up interviews.

Begin discussions with Duke University Office of Information Technology personnel on its use of CMS and methods employed to control content.

March – May 2006
From those discussions, and with input from relevant parties at Duke (University Archives, Medical Center Archives, library web services, Office of Information Technology, etc.), develop best practices for the management and preservation of web content that, when employed, will ensure that appropriate rules are followed

for setting web record expiration and archiving dates, and that specific web records with persistent value are tagged for preservation in an archival repository, e.g. digital repository or archives server. The best practices will be recommendations distributed as a project product to the larger archives and records management community, as well as the institutions participating in the survey.

Conduct small pilot program to capture web content identified as archival using ContentXML workflow tools at Duke University. This pilot also will test the best practices developed following the survey and data gathering portion of the project.

May – June 2006
Complete final report of project findings, conclusions, inferences, and recommendations. Prepare presentation to give at 2006 NHPRC Electronic Records Research Fellowships Symposium.