

National Archives and Records Administration Web Harvest Background Information

Tuesday, April 15, 2008

The National Archives and Records Administration (NARA) issued a memorandum to agency records officers on March 27, 2008, stating that NARA would not conduct an end of administration web snapshot or harvest of Executive Branch websites nor require agencies to do so. This memorandum did not apply to Presidential records or to records of the Congress.

NARA made this decision for the following reasons:

First, NARA collected web records in 2000 and 2004 in an abundance of caution because we had not yet issued records management guidance to federal agencies on managing their web records in accordance with the Federal Records Act (FRA). In January 2005, however, NARA issued "Guidance on Managing Web Records," which addresses agencies' responsibilities for identifying, managing and scheduling web materials they identify as Federal records.

Accordingly, each agency is now responsible, in coordination with NARA, for determining how to manage its web records, including whether to preserve a periodic snapshot of its entire web page. NARA has repeatedly and systematically informed agencies of their statutory responsibilities regarding web records in records management training, briefings to senior agency IT, counsel, and program officials. With the support of OMB, NARA promulgated Bulletin 2006-02 to clarify agency responsibilities for scheduling records as required by Section 207(e) of the E-Government Act of 2002. NARA more recently clarified, in Bulletin 2008-03, that web records are included under this mandate.

NARA will continue to make agencies aware of their legal responsibilities regarding web records, follow up to ensure the agencies are taking actions to meet their requirements in law and regulation, and ensure permanently valuable web records are transferred to the National Archives.

A full list of the web records management guidance products is provided at the conclusion of this memo.

Second, NARA is concerned that a government-wide web harvest could have the unintended consequence of providing some Federal agencies with the false sense that they do not need to manage their web records. They might ask, "If NARA is taking a web snapshot, why does my agency need to manage our records? Won't all of the permanent, historically valuable material be captured and preserved?"

Third, it is not at all clear to NARA, in light of the web and ERM guidance cited herein, that there is continuing permanent archival value of a Federal agency web snapshot taken on one random day near the end of a Presidential term. As the 2005 guidance stated (on

National Archives and Records Administration Web Harvest Background Information

Tuesday, April 15, 2008

page 23, Scheduling Web Records, para. 6) [<http://www.archives.gov/records-mgmt/policy/managing-web-records-index.html>]:

As with other agency records, most web records do not warrant permanent retention and should be scheduled for disposal in accordance with the guidance provided above. In instances where NARA determines that a site or portions of a site has long-term historical value, NARA will work with the creating agency to develop procedures to preserve the records and provide for their transfer to the National Archives.

While a snapshot may provide some indication of “look and feel” of a particular department’s or agency’s web presence on one particular day out of over 1,400 days of a Presidential term, the web snapshot does not systematically or completely document agency actions or functions in a meaningful way. Such records are found in other ongoing, systematic records series that agencies must identify, and NARA approve, for retention and disposition, including in some cases transfer to the National Archives for permanent preservation.

The web snapshots themselves are not complete for even the one day that they are taken. They do not, for example, include intranet sites, the internal department and agency web sites that include more complete indexes of agency materials that potentially have value for systematically documenting the actions and functions of an agency. Moreover, because of the manner in which web snapshots must be undertaken, they are technically incomplete. Compilation of the seed lists that indicate which government websites to harvest is imprecise given the lack of a definitive source. In addition, because of cost constraints and other limitations, the snapshot is limited to a 4-level hierarchy harvest. This effort results in only a high level and very uneven view of an agency’s web presence, with broken links and no connection to more in-depth agency material, including “deep web” databases that are not captured in this process but are an increasingly important feature of the web.

#####

Because Congress is not covered by the Federal Records Act, NARA will continue to conduct a web harvest of Congressional web sites for the same reason we did so for Federal agencies before formal guidance was issued. The harvest will, like previous Congressional snapshots, only document the websites as they appeared to the public on the specific day they were harvested, with all of the shortcomings identified above.

At the end of the current administration, NARA will also receive a snapshot of the White House website. Unlike Federal agencies governed by the Federal Records Act, the White House is governed by the Presidential Records Act, under which all Presidential records

National Archives and Records Administration Web Harvest Background Information

Tuesday, April 15, 2008

are treated as permanent and transferred to NARA for preservation at a Presidential Library.

To view the guidance issued by the National Archives in chronological order, see:

- “Transfer Instructions for Permanent Electronic Records: Web Content Records,” issued in September 2004 and found at <http://archives.gov/records-mgmt/initiatives/web-content-records.html>. This guidance covers the requirements for the transfer of permanent web records to the National Archives.
- “NARA Guidance on Managing Web Records” issued in January 2005 and found at <http://www.archives.gov/records-mgmt/policy/managing-web-records-index.html>. This guidance outlines Federal agency responsibilities for managing and scheduling web records.
- NARA Bulletin 2006-02 issued in December 2005 found at <http://www.archives.gov/records-mgmt/bulletins/2006/2006-02.html>. This guidance provides agencies guidance on implementing Section 207(e) of the E-Government Act of 2002.
- “Implications of Recent Web Technologies for NARA Web Guidance” issued in September 2006 and found at <http://www.archives.gov/records-mgmt/initiatives/web-tech.html>. This guidance discusses the Federal records management requirements for dealing with Web 2.0 technologies like wikis, blogs, and RSS feeds, as well as web portals.
- “Tips for Scheduling Potentially Permanent Web Content Records” issued in May 2007 found at <http://www.archives.gov/records-mgmt/publications/web-tips.pdf>. This guidance provides tips for scheduling potentially permanent web content records.
- NARA Bulletin 2008-03 issued in March 2008 found at <http://www.archives.gov/records-mgmt/bulletins/2008/2008-03.html>. This guidance further clarifies that Federal web records fall under the NARA Bulletin 2006-02.