

From: "RM Communications" <RM.Communications@nara.gov>
To: <BJS1@nrc.gov>, <JAH3@nrc.gov>
Date: 8/4/03 10:52AM
Subject: NWM 26.2003 RETRANSMISSION Request for comment on draft NARAGuidance on Managing Web Records

We are retransmitting this memo to Records Officers with a zipped file. Please note that for hyperlinks to work properly, all files must be downloaded to the same directory. We apologize for any inconvenience the original transmission may have caused.

August 4, 2003

NWM 26.2003

MEMORANDUM TO AGENCY RECORDS OFFICERS: Request for comment on draft NARA Guidance on Managing Web Records

Attached for your review are five HTML files and four associated image files that comprise the draft NARA Guidance on Managing Web Records. To access these files, and for the hyperlinks to work properly, please download the nine files found in the attached file to the same directory.

The guidance is organized into four sections and a separate overview/table of contents:
§ GENERAL BACKGROUND, RESPONSIBILITIES, AND REQUIREMENTS
§ MANAGING WEB RECORDS
§ SCHEDULING WEB RECORDS
§ APPENDIXES

As stated in the overview/table of contents, web site operations are an integral part of an agency's program. Managing web records properly is essential to effective web site operations, especially the mitigation of the risks an agency faces by using the web to carry out agency business. This guidance will assist agency officials in this regard, including agency program staff, webmasters, IT staff, and other agency officials who have a role in web site management and administration.

The draft is also being shared with the CIO Council with a request that agency CIO, web program, and records officer comments be coordinated. Please provide your comments or concurrence via email to Nancy Allard (nancy.allard@nara.gov) by September 3, 2003.

If you have any questions, please contact Ms. Allard on 301-837-1477 or via email.

HOWARD P. LOWELL
Acting Director
Modern Records Programs

Mail Envelope Properties (3F2E7315.B1E : 8 : 11038)

Subject: NWM 26.2003 RETRANSMISSION Request for comment on draft
NARAGuidance on Managing Web Records
Creation Date: 8/4/03 10:51AM
From: "RM Communications" <RM.Communications@nara.gov>
Created By: RM.Communications@nara.gov

Recipients

nrc.gov

twf3_po.TWFN_DO

JAH3 (John Harris)

BJS1 (Brenda Shelton)

Post Office

twf3_po.TWFN_DO

Route

nrc.gov

Files

MESSAGE

NWM 26 attach.zip

Mime.822

Size

1791

41054

59411

Date & Time

08/04/03 10:51AM

Options**Expiration Date:**

None

Priority:

Standard

Reply Requested:

No

Return Notification:

None

Concealed Subject:

No

Security:

Standard

<input checked="" type="checkbox"/> U.S. National Archives and Records Administration	
www.archives.gov	July 18, 2003
—	

NARA Guidance on Managing Web Records

July 2003

[<< NARA Guidance on Managing Web Records, Main Page](#)

SCHEDULING WEB RECORDS

Table of Contents

- Introduction
- 1. Are web-site related records Federal records?
- 2. What records should be covered in a web site schedule?
- 3. What purposes does a web schedule serve?
- 4. How is a web schedule developed?
 - 4.1 What are the records management risks associated with a web site?
 - 4.2 How does risk assessment help with developing a records schedule?
- 5. What is the structure of a web records schedule?
 - 5.1 Single Item Schedule for Web Content and Site Management and Operations Records
 - 5.2 Multiple Item Schedule for Web Content and Site Management and Operations Records
 - 5.3 Web Snapshots
 - 5.4 Examples of Scheduling Options
- 6. How are retention periods for web records determined?
- 7. Can the GRS or existing agency schedules be used for web records?

Introduction

This guide is intended to assist agency staff, especially records officers, in developing disposition schedules for the records relating to agency web sites. This guidance will discuss such matters as

- The types of records that must be scheduled,
- How risk to web records should be assessed,
- How web records schedules should be structured, and
- The factors records officers should consider in determining retention periods for web records.

[Up](#)

1. Are web-site related records Federal records?

The documentary materials agencies accumulate in connection with the transaction of official business are Federal records in accordance with the statutory definition included in the Federal Records Act (44 USC 3301). Since agency web sites are used to carry out agency business, the related records meet the statutory definition and, like all other records, can only be deleted in accordance with a NARA-approved disposition schedule. Web content, even if available in other forms, is a Federal record as is the documentation accumulated in connection with web site management and operations (e.g., logs documenting what was posted on a site, files relating to web site design, records that track usage, etc.).

Up

2. What records should be covered in a web site schedule?

A web schedule should cover web content records that document the information on the site itself. A web records schedule should also include web site management and operations records, which provide the site's context and structure.

Web content records include:

- The content pages that comprise the site, inclusive of the HTML markup;
- Images of web pages created to document web site content, if an agency chooses to document its site by creating such records;
- Listings of the URLs referenced by site's hyperlinks; and
- Records generated when a user interacts with a site.

Web management and operations records that provide context to the site include:

- Web site design records;
- Records that specify an agency's web's policies and procedures by addressing such matters as how records are selected for the site and when and how they may be removed;
- Records documenting the use of copyrighted material on a site;
- Records relating to the software applications used to operate the site; and
- Records that document user access and when pages are placed on the site, updated, and/or removed.

Web management and operations records that provide structure related to the site include:

- Site maps that show the directory structure into which content pages are organized and
- COTS software configuration files used to operate the site and establish its look and feel, including server environment configuration specifications.

More detailed information concerning the types of records agencies accumulate in connection with their web sites is available in GENERAL BACKGROUND, RESPONSIBILITIES, AND REQUIREMENTS, section 5.

Up

3. What purposes does a web schedule serve?

A web schedule is needed to fulfill an agency's statutory responsibilities as spelled out in the Federal Records Act. In addition, a web schedule mitigates the risks associated with the agency web site by ensuring that records needed to prove its trustworthiness are maintained for an appropriate period of time. A web schedule also provides you the legal authority to destroy web records when you destroy them in keeping with the legally approved retention periods found in the retention schedule. Finally, the scheduling process will result in the identification of any web-related records that warrant permanent retention and eventual transfer to the National Archives.

Up

4. How is a web schedule developed?

Developing web schedules involve several distinct steps. The most important of these are:

- Assessing the risks associated with the site and/or specific portions of the site;
- Determining the structure of the web schedule;
- Describing the specific series to be included; and
- Specifying retention periods for each series.

Developing a web schedule is ultimately the responsibility of the agency records officer who takes the lead in carrying out these steps. At all stages of schedule development, the records officer works closely with program staff who are responsible for the site's content, and with webmasters and IT staff who are responsible for web operations. It is these officials who are most familiar with the business needs and risks associated with the site. They must be involved at the beginning of the work when they provide the records officer with key data concerning the site and how it is used. They work with the records officer in carrying out risk assessments, and at the end of the process, they must review the final schedule to ensure that it adequately meets business needs and ensures that risks are mitigated.

Up

4.1 What are the records management risks associated with a web site?

The two basic records management risks associated with a web site are:

1. Failures to create records that are needed to ensure that the site is trustworthy; or
2. Failures to maintain records for an adequate period of time.

If these risks are not mitigated, an agency may be unable to document and/or validate transactions that occurred via a web front end interface, or it may not be able to show what was on a site at a given point in time and who put materials onto the site, and when they were modified or removed. As a result, program operations could be impaired, citizens' rights compromised, negative publicity might be generated among agency stakeholders, the media, and/or the public at large, and, in some instances, the agency could be exposed to costly litigation.

More detailed information concerning risk and risk assessment is included in MANAGING WEB RECORDS, section 2. Trustworthy records and their characteristics (i.e., reliability, authenticity, integrity, and usability) are discussed in MANAGING WEB RECORDS, section 1. Issues related to the scheduling of web records are addressed here in SCHEDULING WEB RECORDS.

Up

4.2 How does risk assessment help with developing a records schedule?

The analysis that is performed in a risk assessment helps you gather information about the web site, the agency programs that the web site supports, and the records that must be scheduled. Some of the information you will gather includes:

- How your agency uses its web site (see GENERAL BACKGROUND, RESPONSIBILITIES, AND REQUIREMENTS, section 1 for examples);
- How often the site and specific portions of the site are changed or updated;
- The degree to which the information on a web site or portions thereof is unique or is readily available in other agency records; and
- Whether the web site or portions of the site are considered high risk.

Up

5. What is the structure of a web records schedule?

Deciding on the level of analysis for risk assessment will resolve one aspect of schedule structure: whether or not the schedule describes records at the level of the entire web site or whether individual portions of the site are scheduled separately. If an agency chooses the latter approach, the items covering each portion of the site should describe pages or groupings of pages broadly in terms of their content or function. Describing portions of the site in too narrow a fashion increases the likelihood that the schedule will become out-of-date as the site changes over time. Drastic changes in site content and/or function would likely require revisions to the schedule, just as significant changes in the content or function of a traditional record series typically warrant changes to a previously approved schedule.

There is no hard and fast rule as to what number of items is appropriate in a web records schedule. Web management and operations records should be grouped, based on

business needs and level of risk, into no more than three or four series for the entire site, or for specific pages or groups of pages if portions of the site are scheduled separately.

The following are two approaches to structuring a web records schedule.

5.1 Single Item Schedule for Web Content and Site Management and Operations Records

One option would be to include in a single schedule item the web content records (either the entire site or portions thereof) along with the related records that pertain to site management and operations. This option would be appropriate if all of the records related to the site warrant the same retention period in order to meet business needs and mitigate risks. For the sake of simplicity and ease of management of the web site, an agency may also choose to use a single item and retention period for web records even if there are variations in risk. However, if this done, the retention period chosen must ensure that records needed to mitigate the highest level of risk associated with the site are retained and remain usable for the appropriate period of time. This approach would require that records with lower levels of risk be maintained for a period of time longer than is necessary. See APPENDIX D, option A for an example.

5.2 Multiple Item Schedule for Web Content and Site Management and Operations Records

Another option is to schedule web site management and operations records separately from web content. This approach would be warranted if business needs and the mitigation of risk mandate different retention periods for the site content records and the management and operations records. Follow these guidelines:

- If all management and operations records associated with a site (or individual portions) are needed for a uniform period of time in order to mitigate risk, then a single item for all such records might be appropriate. Web content would be covered by one or more separate items. APPENDIX D, option B provides an example of multiple items for web content and one item for all web management and operations records.
- If there is variation in the level of risk, then records having the same level of risk should be grouped together and each grouping assigned an appropriate retention period. Such variation is likely in higher risk situations. Therefore, in such situations, multiple items should be developed for the site management and operations records, regardless of whether web content is included in single item or in multiple items. See APPENDIX D, option C and APPENDIX D, option D for examples.

Up

5.3 Web Snapshots

One final point is in order concerning the structure of a web records schedule: whether

or not it should include an item for a web snapshot that captures the content pages and related site map as they existed at particular points in time. Business needs and the need to lessen risk determine whether or not such snapshots are warranted and their frequency. However, in determining when snapshots should be taken, an agency should also consider the frequency with which the information on a site changes. Other things being equal, the more frequent the site undergoes change, the more frequently should snapshots be taken.

Up

5.4 Examples of Scheduling Options

APPENDIX D contains examples of the different options that agencies may employ in scheduling their web site records.

Up

6. How are retention periods for web records determined?

When determining retention periods for web records, as with other records, the agency needs to assess how long the information is needed to satisfy business needs and mitigate risk, taking into account government accountability and the protection of legal rights. In the case of web content that is available in other places in addition to the web, agencies must also consider whether the existence of the information in other records has a bearing on the retention needs for the web records. In the case of information unique to the web site, the web version is the only recordkeeping copy.

- In many cases, particularly where the risk is low, the web content and the related site management and operation records should be assigned a retention period that allows disposal as soon as records are no longer needed in the conduct of agency business.
- In instances where risk levels are higher, web content and the related web management and operations records would probably warrant retention for a period of time that exceeds the time needed to satisfy all business requirements. The extra time needed in order to mitigate risk ought not to be more than 3-5 years beyond the retention period mandated by business needs alone. However, the mitigation of risk may require an even longer retention period in selected instances.

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(s) of a site have 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.

Up

7. Can the GRS or existing agency schedules be used for web records?

There are currently no items in the General Records Schedules that were developed to specifically cover web records. However, some items in the GRS may be applied to web site management and operations records. For example, GRS 14, items 1 and 2, which pertain, respectively, to routine information requests and acknowledgments/transmittals of requests forward elsewhere for reply, can be used to cover transactions of this sort that are generated via an agency web site. Another GRS item that may be used for web records is GRS 21, item 6, which covers routine artwork. Likewise, records relating to training staff regarding the agency web site might be covered by sub-items in GRS 1, item 29. Finally, a variety of items included in GRS 24 may be relevant to web management and operations records. Examples include items 1, oversight and compliance records; item 3, IT asset and configuration files; and item 5, files related to maintaining the security of systems and data; and item 6, user ID, profiles, authorizations, and password files.

As these examples demonstrate, a GRS item may be used for web management and operation records, if the basic content and function of the records is consistent with the GRS and the retention period is appropriate to meet business needs and mitigate risk. This principle also governs whether or not an existing agency-specific schedule item can be used for web management and operation records. For example, a schedule item covering graphics design can be used for records generated in connection with this function as it relates to the agency web site. Likewise, an agency schedule item for internal committee records could be applied to records accumulated by a committee established to advise the agency webmaster.

No items in the current GRS can be applied to web content. Unless an agency schedule item specifically covers records on the agency web site, the item cannot be used for the analogous web content. For example, a schedule covering paper versions of an agency's staff directory cannot be applied to the version posted to the agency web site.

Up

<input type="checkbox"/>

Page URL:

U.S. National Archives & Records Administration

700 Pennsylvania Avenue NW, Washington, DC 20408 • 1-86-NARA-NARA • 1-866-272-6272

NARA Guidance on Managing Web Records

July 2003

<< [NARA Guidance on Managing Web Records, Main Page](#)

APPENDIXES

Table of Contents

- [APPENDIX A](#) - Definitions
 - [APPENDIX B](#) - Web sites of possible interest
 - [APPENDIX C](#) - Web records management responsibility matrix
 - [APPENDIX D](#) - Web Schedule Options
-

APPENDIX A - Definitions

ActiveX—a set of "strategic" object-oriented programming technologies and tools developed by Microsoft allowing the creation of self-sufficient programs that can operate anywhere in an ActiveX network. A rival of the Sun's Java programming language. [from <http://whatis.techtarget.com/>]

ActiveX control—the rough equivalent of a Java applet. [from <http://whatis.techtarget.com/>]

Applet—an applet is a small program, or application module, that can be sent from a web site, along with a web page, to a user's workstation thereby allowing the client to perform a simple task without having to send a request back to the server. Applets are written in Java. [from <http://whatis.techtarget.com/>]

Back end—A term used to characterize program interfaces and services relative to the initial user of these interfaces and services. (The "user" may be a human being or a program.) A "back-end"

application or program serves indirectly in support of the front-end services, usually by being closer to the required resource or having the capability to communicate with the required resource. [from <http://whatis.techtarget.com/>]

CGI scripts—A set of instructions that use the common gateway interface [CGI] standard to pass a Web user's request from a Web server to an application program and to receive data back from the application to the Web server in order to forward it to the user. [from <http://whatis.techtarget.com/>]

Content management software—Software that facilitates web site administration by providing suites of web-related functionality that may include: template for web page construction, on-the-fly page creation from databases, versioning control, workflow, and import/export functionalities.

Content—The information that a document is meant to convey [Society of American Archivists Glossary]. Words, phrases, numbers, or symbols comprising the actual text of the record that were produced by the record creator.

Context—The organizational, functional, and operational circumstances in which documents are created and/or received and used [Society of American Archivists Glossary]. The placement of records within a larger records classification system providing cross-references to other related records.

Up

Cookie—Information that a Web site puts on your hard disk so that it can remember something about you at a later time. [from <http://whatis.techtarget.com/>]

DoD5015.2-STD (v. 2)—A standard, developed by the Department of Defense Records Management Program and endorsed by the National Archives and Records Administration, which sets forth mandatory baseline functional requirements for Records Management Application [RMA] software, defines required system interfaces and search criteria to be supported, and describes the minimum records management requirements that must be met based upon current NARA regulations.

FAQs—An increasingly common feature of the Internet (pronounced 'Fak' or 'Fax' in plural), it is usually provided on a web site as a list of 'frequently asked questions' (and answers). The FAQ seems to have

originated in many of the Usenet groups as a way to acquaint new users with the rules. Today, there are thousands of FAQs on the World Wide Web. [from <http://whatis.techtarget.com/>]

Front end—A term used to characterize program interfaces and services relative to the initial user of these interfaces and services. (The "user" may also be a program.) A "front-end" application is one that application users interact with directly. [from <http://whatis.techtarget.com/>]

HTML—Hypertext Markup Language, is the set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser page. [from <http://www.w3c.org>]

Hyperlink—(aka Hypertext Linking) the HTML syntax for expressing the ability for one Internet document to reference another document located either in another portion of the web site or at another web site on the World Wide Web. [from <http://www.w3c.org>]

Internal target—The HTML syntax for expressing the ability for an Internet document to reference another portion of itself. [from <http://www.w3c.org>]

Java—An object-oriented programming language developed by Sun Microsystems, expressly designed for use in the distributed environment of the Internet, that can also be used to build a small application module or applet for use as part of a web page. [from <http://whatis.techtarget.com/>]

JavaScript—An interpreted programming or script language from Netscape, which is used in web site development to do such things as: automatically change a formatted date on a web page; cause a linked-to page to appear in a popup window; or cause text or a graphic image to change during a mouse rollover. [from <http://whatis.techtarget.com/>]

Up

Legacy system—Applications and associated data that have been inherited from languages, platforms, and techniques earlier than current technology. [from <http://whatis.techtarget.com/>]

Portal—A web site or service that offers access to a broad array of resources and services, such as e-mail, forums, search engines, etc. It may simply be web site consisting solely of URLs to other web sites.

[from <http://whatis.techtarget.com/>]

Records schedule—A document describing records of an agency, organization, or administrative unit, establishing a timetable for their life cycle, and providing authorization for their disposition [Society of American Archivists Glossary], i.e., off-site storage followed by destruction or transfer to the National Archives.

Records series—File units or documents arranged in accordance with a filing system or maintained as a unit because they result from the same accumulation or filing process, the same function, or the same activity; have a particular form; or because of some other relationship arising out of their creation, receipt, or use. [Society of American Archivists Glossary]

RMA [Records Management Application]—Software used by an organization to manage its records. An RMA's primary management functions are categorizing and locating records and identifying records that are due for disposition. RMA software also stores, retrieves, and disposes of the electronic records that are stored in its repository.

Structure—The physical and logical format of a record and the relationships between the data elements. [from [http://www.archives.gov/records management/policy and guidance/electronic signature technology.html](http://www.archives.gov/records%20management/policy%20and%20guidance/electronic%20signature%20technology.html)]

Site map—The directory structure into which web content pages are organized. Uniform Resource Locator—(aka URL) The address of a file (resource) accessible on the Internet. [from <http://whatis.techtarget.com/>]

Up

APPENDIX B - Web Sites of Possible Interest

Agencies may wish to supplement this initial NARA guidance with these other applicable resources:

- NHPRC-funded research *Analysis and Development of Model Quality Guidelines for Electronic Records Management on State and Federal Websites* (http://slis-two.lis.fsu.edu/~cmcclore/nhprc/nhprc_toc.html)
- Policy and guidelines for keeping records of web-based activity in the Commonwealth Government: A policy for

keeping records of web-based activity in the Commonwealth Government

(http://www.naa.gov.au/recordkeeping/er/web_records/intro.html)

- Government of Canada *Internet Guide (3rd ed.)*: A general guide outlining and understanding and growing appreciation of the structural complexities of delivering information on the Internet.
(http://www.canada.gc.ca/programs/guide/main_e.html)
- National Library of Canada's *Preservation of Digital Information - Issues and Current Status* (<http://www.nlc-bnc.ca/publications/1/p1-259-e.html>)
- Government Paperwork Elimination Act
(http://www.cio.gov/documents/paperwork_elimination_act.html)
- Paperwork Reduction Act
(http://www.archives.gov/federal_register/public_laws/paperwork_reduction_act/3501.html)
- E-Government Act of 2002, December 17, 2002
(http://www.cio.gov/documents/e_gov_act_2002.pdf)
- Summary of Provisions; E-Government Act of 2002(http://www.whitehouse.gov/omb/egov/pres_state2.htm)
- Section 508 - Public Law: 105-220 (08/07/98) (b) Electronic and Information Technology Regulations.
(http://www.cio.gov/documents/section_508_august_1998.html)
- Children's Online Privacy Protection Act of 1998 (COPPA)
(http://www.cio.gov/documents/coppa_act_1998.html)
- OMB's *Privacy Policies and Data Collection on Federal Web Sites* (<http://www.whitehouse.gov/omb/memoranda/m99-18attach.html>)

Up

APPENDIX C - Web Records Management Responsibility Matrix

RM Responsibility	Who has responsibility?

	<ul style="list-style-type: none"> ○ Web Program Manager/Webmaster ○ Program Staff who create or review content ○ IT Personnel and Contractor Staff ○ Agency Office of General Counsel
Developing of schedules accounting for records documenting agency web site operations	<ul style="list-style-type: none"> • Working as a team: <ul style="list-style-type: none"> ○ Agency Records Officer ○ Web Staff ○ Program Staff ○ Agency office of General Counsel
Developing SOP's that accord with records schedules containing web site records	<ul style="list-style-type: none"> • Web Program Manager/Webmaster
Implementing RM-related SOPs	<ul style="list-style-type: none"> • IT Personnel and Contractor Staff
Creating and managing web site operations program records	<ul style="list-style-type: none"> • Web Program Manager/Webmaster
Creating web content	<ul style="list-style-type: none"> • Program Staff (content authors) • Public Affairs/Communications Staff (in many agencies)
Assisting agencies via guidance	<ul style="list-style-type: none"> • NARA
Approving records schedules documenting agency web site operations	<ul style="list-style-type: none"> • NARA

APPENDIX D - Web Schedule Options

NOTE: The examples in this appendix do not recommend a specific disposition. They are provided only for illustration of the ways you may structure your schedule items. You must determine the appropriate retention period based on your agency's business and risk mitigation needs.

A. Single schedule item for all web records

See SCHEDULING WEB RECORDS, section 5.1

Agency web site and all related web management and operations records - Destroy/delete when ____ old or when superseded, obsolete or no longer needed for the conduct of agency business, whichever is later.

B. Single schedule item for web content and single item for web management and operations records

See SCHEDULING WEB RECORDS, section 5.2

1. Web site content - Delete when ____ old or when superseded, obsolete, or no longer needed for the conduct of agency business, whichever is later.
2. Web management and operations records - Destroy/delete when ____ old or when related web content records are deleted, whichever is later.

C. Multiple schedule items for distinct portions of the web site but a single item for all management and operations records

See SCHEDULING WEB RECORDS, section 5.2

1. Web content:
 - a. Pages containing copies of agency issuances - Delete when superseded, obsolete, or no longer needed for the conduct of agency business.
 - b. All other pages - Delete when ____ old or when superseded, obsolete, or no longer needed for the conduct of agency business, whichever is later.
2. Web management and operations records -

Destroy/delete when ___ old or when related web content records are deleted, whichever is later.

D. Multiple schedule items for distinct portions of the web site and multiple items for web management and operations records

See SCHEDULING WEB RECORDS, section 5.2

1. Web content:

- a. Pages containing copies of agency directives and handbooks - Delete when superseded, obsolete, or no longer needed for the conduct of agency business.
- b. Pages containing annual reports, IG semi-annual reports to Congress, and Privacy Act reports - Delete when ___ old.
- c. Monthly snapshot of web pages - Delete when ___ old.
- d. All other web content - Delete when ___ old or when superseded, obsolete, or no longer needed for the conduct of agency business, whichever is later.

2. Web management and operations records

- a. Design records - Delete/destroy after ___ year[s].
- b. Program management records, including policies and procedures - Delete/destroy when 4 years old or when superseded, obsolete, or no longer needed for the conduct of agency business, whichever is later.
- c. Web site posting logs indicating when pages were posted, updated or removed - Delete/destroy when ___ old.
- d. All other records - Delete/destroy when superseded or obsolete.

Up



Page URL:

NARA Guidance on Managing Web Records

July 2003

<< [NARA Guidance on Managing Web Records, Main Page](#)

GENERAL BACKGROUND, RESPONSIBILITIES, AND REQUIREMENTS

Table of Contents

- [Introduction](#)
 - [1. How are agencies currently using the web?](#)
 - [2. Who is responsible for managing agency web sites?](#)
 - [3. What statutory and regulatory requirements apply to agency web operations?](#)
 - [4. Are web site-related records Federal Records?](#)
 - [5. What web site records must be managed and scheduled?](#)
-

1. Introduction:

Web site operations are an integral part of agency programs. This guide provides an initial, high level framework that you can use to manage both the content records on an agency web site and the records documenting web site operations. This guidance, which is based on statutory requirements, also provides principles that form a sound basis for agency web site management operations. It will help you reach agency program objectives by relating good records management practices (e.g., maintaining trustworthy web sites) and existing tools for meeting agency objectives (e.g., risk assessments).

It should be noted that while this guidance primarily applies to public Internet web sites, it is equally applicable to web sites that may be on agency intranets, virtual private networks, and security classified web sites. When web sites are jointly sponsored between agencies, one party must claim ownership of records management responsibilities. This guidance does not address unique issues relating to web portals or workgroup-based webs. Those unique types of web sites will be

dealt with in subsequent guidance. In addition, this guidance does not apply to internal agency applications (e.g., time and attendance or travel) with web-enabled interfaces because these are not web applications.

Up

1. How are agencies currently using the web?

Federal agencies are currently using the web in a variety of ways. Table 1 provides a presentation of these various activities, in a continuum of increasing complexity. This presentation is only an example and is not meant to be exhaustive.

Table 1. Types of Federal Agency Web Site Operations

Description of Agency Web Site Activity
Web site as a relatively static, repository for agency publications
Web site as a more dynamic, repository for agency publications
Web site as a limited communications tool <ul style="list-style-type: none">• How used: Web site allows minimal interaction with the

captured through the agency e-mail system and not as part of the web site.

Web site as repository for query-driven agency information (e.g., databases)

- **How used:** Web site as 'front end' interface for other agency application services, such as 'back end' services and agency databases
- **Change frequency:** Forms for displaying information may change infrequently, but databases that provide information posted to the forms updated continuously.

Web site as repository for a wide variety of information interactions

- **How used:** Web site as significant mechanism for other interactions with public, e.g.:
 - Providing form/application fill-in capabilities for initiation of other web- and non web-based agency services
 - Conducting public, electronic town meetings
 - Using the web site to solicit comments on proposed regulations
- **Change frequency:** Information both supplied by the end user and displayed on the web sites is varied in nature and changes frequently.

In reality, web sites serve multiple purposes and most agencies conduct more than one of these activities. As an analogy, web sites can be seen as an automated "customer service window" through which customers interact with a business on many issues. They could ask questions, get directions or information, make complaints, seek refunds, apply for jobs, order merchandise, or make payments. Each of these activities generate different records and must be analyzed to determine the appropriate level of management.

Records created from conducting program activities over web sites have a variety of characteristics, depending on how the web site is used. These characteristics influence the selection of an appropriate unit of analysis for a risk assessment (e.g., web site *in toto*, specific program portions of web site, etc.), as is discussed in MANAGING WEB RECORDS, section 2. Knowing the purpose, work processes,

and type of content is basic to understanding how to manage the web site.

Up

2. Who is responsible for managing agency web sites?

Responsibility for managing Federal agency web sites and related records various parties. To implement this guidance agencies need to assign appropriate responsibilities to each of the parties, perhaps using a team approach. Within an agency, organizational roles may vary, but would include:

- The ***Head of the Agency***, who is responsible for protecting the integrity of agency programs and trustworthiness of agency information (see *44 U.S.C. Chapter 31*);
- ***Federal Agency Records Officers***, who are responsible for ensuring adequate management and control over agency records, including agency web site-related records, and development of an overall records management program;
- ***Agency program staff***, who are responsible for contributing to the content pages that are posted on an agency's web site;
- ***Agency program staff*** who have gatekeeping responsibility for determining what program related content material gets posted;
- ***Agency webmasters***, who are responsible for ensuring both the documentation of web site operations and its compliance with agency recordkeeping procedures;
- ***Agency information technology personnel***, specifically system administrators, who work on the front line of both web recordkeeping and web operations, thereby bridging the needs of both the Records Officer and the Webmaster.
- ***Contractor staff*** procured by the agency to assist in various aspects of web site operations;
- ***Agency's Office of General Counsel staff***, who advise on potential legal issues (e.g., incorporation of copyrighted materials in web site design or e-FOIA); and
- ***Agency's Public Affairs or policy staff***, who may be responsible for coordinating the overall presentation of an agency web site.

Agencies are assisted in these responsibilities by:

- ***The National Archives and Records Administration (NARA)***, which is responsible for promulgation of policies and guidance

to assist Federal agencies to comply with statutory and regulatory recordkeeping requirements related to web site operations; and

- The *Office of Management and Budget (OMB)*, which has responsibilities under 44 U.S.C. Chapter 35 for coordination of Federal information policy.

These and other responsibilities are outlined in APPENDIX C.

Up

3. What statutory and regulatory requirements apply to agency web operations?

The Federal Records Act applies to all agency records, including web records (see section 4.1). There are a variety of other statutes and regulations that affect agency web site operations. These include, for example:

- *The Government Paperwork Elimination Act* (Public Law 105-277), which requires that, when practicable, Federal agencies use electronic forms, electronic filing, and electronic signatures to conduct official business with the public (some of which will be via agency web sites) by 2003;
- *The Paperwork Reduction Act*, which promotes the use of information technology (such as web sites) in the timely dissemination of public information while at the same time reducing information collection burdens on the public;
- *The E-Government Act of 2002* (Public Law 107-347), which establishes minimum acceptable security practices for Federal information systems, including agency web sites, and places other requirements relating to web sites on OMB, NARA, and agencies;
- *Section 508, The Electronic and Information Technology Regulations* (Public Law 105-220), a series of 1998 amendments to the *Rehabilitation Act* which includes information technology access requirements applicable to web sites for persons with disabilities;
- *The Children's Online Privacy Protection Act (COPPA)* of 1998, which requires verifiable parental consent regarding the collection of personal information from children under the age of thirteen via online forums; and
- OMB's *Privacy Policies and Data Collection on Federal Web Sites*, which describes requirements for Federal agency web

site privacy policies, including prohibitions regarding the use of 'cookies'.

Up

4. Are web site-related records Federal records?

Most Federal agencies use the Internet to assist in carrying out their mission. They may simply disseminate information also available in other forms or conduct business (e.g., e-government initiatives). Agencies must document all of their agency programs (44 U.S.C. 3101) including web sites which are part of its overall public message.

Federal personnel should understand that agency web pages meet the definition of a Federal record¹ and therefore must be managed as such. Records relating to managing web sites, tracking use (hits), documenting what was posted and when, and documenting decisions about web content and design, are also Federal records.

In brief, much if not all documentation related to agency web site operations are Federal records. As such, they must be scheduled and can only be deleted with a NARA-approved disposition authority.

Up

5. What web site records must be managed and scheduled?

You must identify those web site-related records that are needed to ensure trustworthiness for each web site. Use the concepts of reliability, authenticity, integrity, and usability, as discussed in MANAGING WEB RECORDS, section 1.1, to establish criteria for identifying the web site-related records needed to document agency programs. The following list includes many of the records that might result from agency web site operations. Your agency may not produce all of these types of records. These records may exist in electronic or nonelectronic format.

- *HTML-encoded pages*: The content pages comprising an agency web site, inclusive of the HTML markup.
- *Images of web content pages*: Screen prints or graphic files produced by other means, such as images of pages created on-

the-fly by content management software. (This would only apply if the agency chose this approach to document web site content pages.)

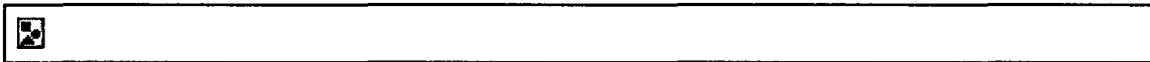
- ***Comprehensive listing of URLs referenced by a web site's hyperlinks:*** In effect, a bibliography of all uniform resource locators referenced via hyperlinks embedded within a web site's content pages, along with a textual reference describing the destination of the hyperlink.
- ***Records generated interactively on the web site:*** Records that are created dynamically in real time when a user interacts with an agency web site (e.g., on-the-fly, text-based page creation, forms filled out online, etc.).
- ***Additional content files referenced within HTML-encoded pages:*** Files having the ability to 'self-execute' (e.g., CGI scripts, Java/ActiveX applets, customized programs that generate on-line sound or moving images) as well as files that are static (e.g., these include graphic files, multi-national character sets, etc.), both external to the HTML-encoded content pages but referenced in the HTML syntax.
- ***Web site design records:*** Records such as graphic design layouts for pages or embedded image maps, and/or minutes of meetings documenting the production of such.
- ***Copyrighted web page content and records documenting the use of such content materials:*** Many Federal web sites incorporate copyrighted content (e.g., works for hire such as custom produced graphics files) in the content of web pages. These copyrighted materials, as well as additional records are necessary for documenting the appropriate use of such (as determined by your Office of General Counsel).
- ***Web site program operations records:*** Program management files including records that document policies and procedures for agency web site operations, including those that:
 - Provide detailed procedures for documenting how records are selected, created and approved for web posting, and how they will be revised or removed.
 - Specify what records will be created and how they will be created for interactive sections of web sites.
 - Document procedures used in the operation of the site.
 - Specify the relationship of the webmaster and other staff involved in preparing and posting web documents to program officials and the agency records officer.
 - Demonstrate the development of policies and procedures to ensure Section 508 compliance.
 - Otherwise explain or illustrate site development and management procedures.
- ***Web site system software-related records:*** Records related to

the application software used in conjunction with operation of the web site. These include:

- Records produced in the analysis and selection of any commercial off-the-shelf [COTS] software.
- Records describing customization of COTS web-related software.
- Documentation relating to COTS web-related software.
- *Web site logs and statistical compilations:* Records including logs of user access (frequency of "hits"), site posting logs (indicating when pages were posted, updated, or removed from the site), system load and search result statistics, and ad hoc reports containing such.
- *Web site map:* The directory structure into which web content pages are organized.
- *COTS software configuration files:* Files used to operate the web site and establish its look and feel including server environment configuration specifications.

Up

¹ I.e., they constitute "...machine readable materials, ...made... by an agency of the United States Government... in connection with the transaction of public business... (44 U.S.C. 3301)."



Page URL:

NARA Guidance on Managing Web Records

July 2003

Web site operations are an integral part of an agency's program. Managing web records properly is essential to effective web site operations, especially the mitigation of the risks an agency faces by using the web to carry out agency business. This guidance will assist agency officials in this regard, including agency program staff, webmasters, IT staff, and other agency officials who have a role in web site management and administration.

This guidance is comprised of four sections:

- **GENERAL BACKGROUND, RESPONSIBILITIES, AND REQUIREMENTS** - This section outlines the various ways in which agencies use web sites, the roles played by different agency staffs in web operations, and the basic statutory requirements that govern web sites, especially the Federal Records Act. This section also identifies the types of records agencies typically accumulate in connection with their web sites.
- **MANAGING WEB RECORDS** - This section is intended to assist agency staff in managing their web records. Aimed particularly at webmasters, IT staff, and those program officials responsible for web content, this section outlines the steps an agency must take to ensure trustworthy web records and mitigate the risks associated with web site operations.
- **SCHEDULING WEB RECORDS** - This section is intended to assist agency staff, especially records officers, in developing disposition schedules for web records. It addresses such matters as the types of records that should be covered in web schedules, how these schedules might be structured, and the factors an agency should consider in determining how long records should be retained.
- **APPENDIXES**



Page URL:

U.S. National Archives & Records Administration
700 Pennsylvania Avenue NW, Washington, DC 20408 • 1-86-NARA-NARA • 1-866-272-6272

NARA Guidance on Managing Web Records

July 2003

[<< NARA Guidance on Managing Web Records, Main Page](#)

MANAGING WEB RECORDS

Table of Contents

- [Introduction](#)
 - [1. What are trustworthy records?](#)
 - [1.1 What are the characteristics of trustworthy records?](#)
 - [1.2 How can I maintain a trustworthy web site?](#)
 - [2. Risk and risk assessment](#)
 - [2.1 What are the records management risks associated with web sites?](#)
 - [2.2 How can I conduct a records management risk assessment?](#)
 - [2.3 How do I determine the unit of analysis for doing a risk assessment?](#)
 - [2.4 Who is involved in assessing risk?](#)
 - [3. Mitigating risk](#)
 - [3.1 What steps should I follow to help mitigate risk in managing web records?](#)
 - [3.2 How should an agency manage web site content in order to mitigate risk?](#)
 - [3.3 How frequently should I capture a snapshot of my web site's content pages?](#)
 - [3.4 How do I track changes to web site content pages between snapshots?](#)
 - [3.5 When preserving long-term web content pages, how could I treat hyperlinks?](#)
 - [3.6 Who is responsible for managing web content pages?](#)
 - [3.7 Who is responsible for managing web management and operations records?](#)
-

Introduction

This guide is intended to assist agency staff in managing their web records. It is particularly geared to the needs of program officials, who provide the information posted on web sites, and those staff who manage agency web sites, including webmasters and IT staff.

1. What are trustworthy records?

Trustworthy records are essential for an agency to meet its legal and internal business needs. Reliability, authenticity, integrity, and usability are the characteristics used to describe trustworthy records from a records management perspective. How these terms apply to web sites and web records is discussed more fully in [section 1.1](#).

Creating and maintaining trustworthy records requires resources. Agencies need to conduct a risk analysis to balance the level of trustworthiness of records against costs and risks. The level of resources used to ensure these characteristics depends on the agency's business needs and perception of risk. (See [section 2](#) for a discussion of risk assessment.) Web site operations that are critical to the agency business need a greater assurance level that they are reliable and authentic, maintain integrity, and are usable over a longer period of time than less critical operations.¹

Up

1.1 What are the characteristics of trustworthy records?

- **Reliability.** A reliable web site is one whose content can be trusted as a full and accurate representation of the transactions, activities, or facts to which it attests and therefore can be depended upon in the course of subsequent transactions or activities.
- **Authenticity.** An authentic web site is one that is proven to be what it purports to be and to have been created by the agency with which it is identified.

Web site-related records should be created by individuals who have direct knowledge of the facts or by instruments routinely used within the business to conduct the transaction.

To demonstrate the authenticity of a web site, agencies should

implement and document policies and procedures which control the creation, transmission, receipt, and maintenance of web site records to ensure that records creators are authorized and identified and that records are protected against unauthorized addition, unauthorized deletion, and alteration (e.g., via hacking).

- ***Integrity.*** The integrity of a web site refers to it being complete and unaltered.

Agencies need to have policies and procedures for routinely updating and modifying their web sites. The policies and procedures also support records management needs. A way of ensuring integrity is to develop a clear-cut policy for audit trails. As stated in the ISO Technical Report 15489-1², sec. 7.2.4, "records systems should maintain audit trails or other elements sufficient to demonstrate that records were effectively protected from unauthorized alternation or destruction."

Another aspect of integrity is the structural integrity of a web site's content-related records. The structure of a web site, that is, its physical and logical format and the relationships between the pages and content elements comprising the site, should remain physically or logically intact. Failure to maintain the web site's structural integrity may impair its reliability and authenticity.

- ***Usability.*** A usable web site is one that can be located, retrieved, presented, and interpreted. In retrieval and use, you should be able to directly connect the web site to the business activity or transaction that produced it. You should be able to identify both the site and its content within the context of broader business activities and functions. The links between content, contextual, and structural web site-related records that document agency web site activities should be maintained. These contextual linkages should provide an understanding of the transactions that created and used them.

Up

1.2 How can I maintain a trustworthy web site?

For web site records to remain reliable, authentic, with integrity, and useable for as long as they are needed, you must maintain the content, context, and sometimes structure of the site. A trustworthy web site includes not only the content pages, but also information about the web site that relates to the context in which it was created and used. Specific

contextual information varies depending upon the business, legal, and regulatory requirements of the business activity. Structural information on the organization of the web site supports its long-term integrity.

For Federal agency web-based operations, NARA interprets web-related records that provide content, context, and structure as follows:

- **Content:** the actual html-encoded pages themselves and additional content files referenced therein (or graphic files produced by other means, such as images of pages created on-the-fly by content management software). Maintenance of these web site content records is necessary to support all of the characteristics of trustworthiness: reliability, authenticity, integrity, and usability.
- **Context:** administrative and technical records necessary for or produced during the management of an agency web site. Maintenance of these records provide a context for web operations, which attests to the reliability, authenticity, and integrity of an agency's web site.
- **Structure:** a site map indicating the arrangement of a web site's content pages and software configuration files of content management systems. Maintenance of this record provides a structure for content records and thereby enables the integrity and usability of both current and preserved versions of an agency web site.

Records in all of these categories contribute to the adequate documentation of agency web site operations. An agency's risk assessment of its web site operations advises which agency records are necessary to ensure the trustworthiness of its web site operations, how the records should be maintained appropriately, and how long those records are to be retained (see [section 3](#)).

Up

2. Risk and risk assessment

Typically, agencies conduct risk assessments in order to establish appropriate levels of management controls prior to undertaking new program initiatives. NARA assumes that such risk assessments have been conducted for development of agency web site operations. These risk assessments can also be used to establish records management controls.

Agency records management practices are based on operational needs and perceptions of risks. Operational needs (e.g., providing public information,

documenting transactions with the public) determine the way agencies address trustworthiness of web site operations (see [section 1.1](#)). Risk assessment and risk mitigation, along with other techniques, are used to establish both management controls for and documentation requirements of agency activities. The emphasis in this guidance on risk assessment relates to *Clinger-Cohen* requirements for incorporation of risk management into program activities, particularly for those that are dependent upon information technology (e.g., web site operations).

Up

2.1 What are the records management risks associated with web sites?

From a records management perspective, risk relates to (1) challenge to the trustworthiness of the records (e.g., legal challenge) that can be expected over the life of the record; and, (2) unauthorized loss or destruction of records. Consequences are measured by the degree of loss which the agency or citizens would suffer if the trustworthiness of the web site-related records could not be verified, or if there would be unauthorized loss or destruction.

Examples of records management-related risks associated with agency web sites are mainly technical risks. Loss of information could result from:

- An inability to document and/or validate transactions that occur via an agency web site front end;
- An inability to reconstruct views of web pages that were created dynamically and existed only virtually for the time that they were viewed;
- Compromise of e-government transactions; and
- An inability to track web-assisted policy development or document agency decisions relating to agency web operations.

Each of these technical risks engender a variety of basically programmatic consequences, including:

- Litigation or liability if an agency is unable to verify what was on its site at a given point in time;
- Impairment of program operations;
- An inability to produce records that document accountability and stewardship of materials posted to the agency web site;
- Unwanted media attention;

- Dissemination of misinformation;
- Financial losses due to compromising the citizens' or government's rights;
- Negative reactions of agency stakeholders (e.g., the Executive or Legislative branch); and
- Compromise of the agency's mission.

Up

2.2 How can I conduct a records management risk assessment?

A risk assessment should address the possible consequences of untrustworthy, lost, or unrecoverable records, including the legal risk and financial costs of losses, the likelihood that a damaging event will occur, and the costs of taking corrective actions. Agencies may have formal risk assessment procedures that may be applied to agency web site operations.

The assessment factors may include *records management threats*, *visibility*, *consequences*, and *sensitivity*.

- *Records management threats* relate to the likelihood of experiencing technical risks discussed in [section 2.1](#) (e.g., risks of unauthorized destruction of web site-related records, litigation risks associated with inability to reconstruct views of web sites at specific points in time, risks associated with inability to document web site policy decisions, etc.).
- *Visibility* is the level of active public awareness of an agency's web site operations.
- *Consequences* describes the level of negative organizational, economic, or programmatic impact if web records are untrustworthy, lost, or unrecoverable.
- *Sensitivity* characterizes the agency's assessment of the importance of web site operations.

The results of this assessment support agency programs by providing a basis for determining what types of web site records should be created, how they should be maintained, and how long they should be maintained. The assessment will assist agencies in ensuring that the level of risk is tolerable and that resources are properly allocated. The results of this assessment can also aid in the development of web site records schedules.

Up

2.3 How do I determine the unit of analysis for doing a risk assessment?

One key aspect of conducting a risk assessment is determining the appropriate unit of analysis, *i.e.*, the entire web site or portions of the site. This is important because it impacts on the choice of management controls (see section 3.1) and scheduling (see SCHEDULING WEB RECORDS, section 5). This concept is flexible to allow you to adapt it for your particular site and management needs. Possible units of analysis you could use when applying the risk assessment factors include the entire site, portions of the site related to specific functions or organizations, clusters of pages on a specific subject, etc. The basic options for analysis are:

- Evaluate the web site *in toto*;
- Evaluate the web site basically as a whole, minus one or two portions that exhibit substantially different characteristics; and
- Substantially break out web site pages or groups of pages based on function or other characteristics.

First consider whether the site has a single level of risk or varying levels of risk. Use the risk assessment factors. If the level you have chosen for analysis has more than one answer to any of the factors, you may need to consider breaking out those portions. Note that changes in any of the four factors could affect the risk level.

Example:

One means of portioning the NARA web site for risk assessment is dividing it by program areas; for example, the **Records Management** portion of the NARA web site (see http://www.archives.gov/records_management/index.html).

Another example of portioning is to separate out specific content pages; for example, the **FOIA** portion of the NARA web site (see http://www.archives.gov/research_room/foia_reading_room/foia_reading_room.html).

A final example of portioning the NARA web site would be on the basis of the nature of the content pages; for example, those pages comprising the **Archival Research Catalog database** (see http://www.archives.gov/research_room/arc/) or unique, one-time exhibits such as those on the **Charters of Freedom** (see http://www.archives.gov/exhibit_hall/charters_of_freedom/charters_of_freedom.html).

Determine the unit of risk assessment in consultation with other agency staff associated with the web site (see APPENDIX C).

If you decide, for operational reasons, to manage the web site as a single unit, all components will be treated the same in terms of risk. You will need to manage all parts of the site in accordance with the highest level of risk determined for any portion of the site.

Once you select the unit(s) of analysis for conducting the risk assessment of your web site, determine a level of risk using an accepted methodology. Once you have determined the level of records management risks for the site or portions of the site, you will need to protect the records appropriately. Develop a plan to address records issues (e.g., types of records needed to document the web-based activity, length of time they are needed to support the business purposes), as well as IT issues (e.g., security of the site and information exchanged over the site) and management/internal controls on the processes. The agency's program, web, and IT staff, and Agency Records Officer should contribute to developing the plan.

Up

2.4 Who is involved in assessing risk?

A number of players are involved in conducting a risk assessment, including program staff responsible for the content of the material placed on the site, the agency webmaster, records managers, IT professionals and legal staff.

- The program staff brings expertise concerning the nature of the information on the site and who uses it; perhaps even more important, they are in the best position to know what problems might arise if information on a site is incorrect, out-of-date, or lost.
- Webmasters best understand how information is placed on the site, revised, and/or removed and, in addition, they know what records are created or should be created when these actions take place.
- The agency records manager plays a valuable facilitative role in relating business processes to the records that result from those processes. Records managers also provide support in developing records management processes for ensuring trustworthy records. They also have the knowledge and experience to help relate the risk assessment results to deciding what records should be created

and how long they need to be retained for business purposes. Records managers can then incorporate these retention decisions into web records retention schedules.

- IT professionals have the expertise and know how to use computer technology to help mitigate the risks identified in the risk assessment.
- Legal staff contribute the knowledge of the legal requirements that the agency must follow.

Up

3. Mitigating risk

This section addresses risk mitigation issues that are of particular relevance to program staff responsible for web content and to webmasters. These issues include how to mitigate risk by producing a web snapshot and other means of documenting web site content, how changes to sites between snapshots can be tracked, and how hyperlinks may be treated when preserving long-term web content pages. This section also addresses the roles of the webmaster and program staff in schedule implementation.

Up

3.1 What steps should I follow to help mitigate risk in managing web records?

Some of the steps outlined here are the same as for other kinds of records. You should address each of these steps. However, your risk assessment will inform the level of effort you will need to commit to mitigate your risks. These steps include:

- Create and maintain documentation of the systems used to create and maintain your web records.
- Ensure that your web records are created and maintained in a secure environment that protects the records from unauthorized alteration or destruction.
- Implement standard operating procedures for the creation, use, and management of your web records and maintain adequate written documentation of those procedures.
- Create and maintain your web records according to these documented standard operating procedures.
- Train agency staff in the standard operating procedures.

- Develop a retention schedule for your web records and obtain official NARA approval of that retention schedule. (See SCHEDULING WEB RECORDS.) You will need to cite the official disposition authorities found in your schedule if your agency is faced with legal challenges to produce records that have been destroyed.

Up

3.2 How should an agency manage web site content in order to mitigate risk?

You must preserve the records as long as they are needed for business operations. While traditional records management techniques apply fairly easily to relatively static contextual and structural web site records, managing web content pages is more complex. Web content pages may be frequently changed or updated, and when updates or redesign of web site maps changes the relation/organization of web pages. Agencies may preserve web site content pages by: (1) producing a stand-alone copy or snapshot,³ of all content pages on the site at a particular time, and (2) accompanying this snapshot with a site map⁴ that shows the relationship (i.e., directory structure) of those pages to each other. If your agency decides to take snapshots, you must decide:

- How frequently should a new snapshot be captured;
- If it is necessary to track changes in both the content pages and the site map that occur between snapshots; and
- If it is, how to track these changes (see section 3.4). The answers to these questions depend on your risk assessment of web site operations.

Content Management Systems (CMS) are used to manage the content of a web site. The system consists of a content management application (CMA) and a content delivery application (CDA). The CMA can relieve the webmaster of many of the decisions and actions required to manage the creation, modification, and removal of content from a web site. A CDA uses and compiles that information to update the web site. CMSs can be used to create audit trails associated with content that is created on-the-fly.

To ensure availability of current web content you may use web server back-up software or an Internet-based service to preserve copies of files or databases to restore the content in case of equipment failure or other catastrophe.

Instead of snapshots or preservation of web content in a records management application (RMA), agencies may decide to manage the live versions of web site content pages **while up on the web site**. For low risk web sites, the current posted version of a site plus the standard operating procedures in place used to manage the site and a log of changes may be sufficient for business purposes. Please note that this option may not be appropriate for medium and high risk sites because recordkeeping copies are not ever set aside.

Up

3.3 How frequently should I capture a snapshot of my web site's content pages?

Determine the frequency of snapshots of a web site's content pages and site map by using the risk profiling factors described in section 2.2 . The unit(s) of analysis for the risk assessment would correspond to the unit considered for the snapshot. For portions of a web site considered of higher records management risk, snapshots would be more frequent. The stakeholders discussed in APPENDIX C should cooperate in deciding how frequently snapshots should be taken.

Up

3.4 How do I track changes to web site content pages between snapshots?

Four types of changes can occur to a web site's content between snapshots:

1. Changes to the content of an individual page without changing its placement in the overall organization of the web site,
2. Wholesale replacement of an individual page (or sections of pages) without changing its placement in the overall organization of the web site,
3. Changes in location of a page (or groups of pages), or
4. Combinations of changes of these first three types.

Changes of the first two types (i.e., changes to content without changing the page's placement in the overall organization of the web site) can be treated as a version control issue. You must decide how to best keep track of the versions of content pages.

One non-automated approach to version control is to annotate changes of content pages as a comment in the HTML coding. The comment, which will not appear when the page is displayed as a browser, could indicate when the page was changed (e.g., `<!--Updated by MDG on 03/02/03-->`) or could reference the page which it wholesale replaced (e.g., `<!--This page replaced content page Introduction_1.html on 09/10/02-->`). Another manual approach would be to maintain a log file of content changes of the first two types of changes. (Keep in mind that neither of these approaches would allow you to actually reconstruct views presented at a particular time. This may be found acceptable per your risk assessment).

Alternatively, you may use content management software (CMS) to track versions of web pages in the first two cases. CMS would also offer limited page view reconstruction capabilities—default settings for the databases that support most CMS software would retain only recent changes.

You can handle major changes to the site's directory structure by producing a new site map at the time of major revision. This could be accomplished in a manual or automated manner.

One automated way to track changes is to manage the web content records with a DoD 5015.2-certified records management application (RMA). A DoD 5015.2-certified RMA allows you to impose version control over changed copies of documents. If you use the RMA to store iterative copies of web pages as they are changed, you will be able to see how many times and when each page was changed. Web pages may be added to an RMA's repository manually or via any of the automated tools discussed below. Please note that DoD-certified RMAs have been endorsed by NARA for civilian agency use because they comply with records management regulations. However, none of the other tools described below were designed for records management.

Another tool is a type of search engine called a "spider" or "crawler." A spider is a program that visits web sites and reads their pages and other information in order to create entries for a search engine index. Spider/crawler software can be used to identify changes to web site content and/or gathering of content related to specific site (sub)units.

When justified by risk assessment, you may want to be able to closely reconstruct the content and structure of a site by combining records of updates to web content pages with snapshots of web sites. The degree of exactness to which a web site may be reproduced depends on whether changes to all static and dynamic files referenced within HTML-encoded content pages were also tracked between snapshots.

3.5 When preserving long-term web content pages, how could I treat hyperlinks?

Web content pages frequently use hyperlinks to: (1) jump to another location within the page, (2) jump to another page within the web site, or (3) jump to a page on another web site. Depending on the preservation strategy chosen, it is possible, and in many cases likely, that these hyperlinks will not continue to function in the preservation copy of the web content pages. For sites with high records management risk, agencies might want to implement the following approach, in order to enhance the usability of preservation copies of long-term web content pages.

Hints for Managing Content of High RM Risk Web Sites

Internal target hyperlinks	For hyperlinks that simply transfer control to a different location within the same page (aka internal target), no additional work is required, as the link will continue to function when the content page is interpreted by a browser application.
Hyperlinks not under local RM Control	For hyperlinks that transfer control to either a different page or another web site (aka hyperlinks) that is not under local records management control, NARA suggests that agencies consider requiring web site content developers to modify the HTML syntax of web content pages containing such hyperlinks on a day-forward basis. This modification would include the insertion of an HTML comment after the hyperlink that described, in

	<p>following the hyperlink that describes it, as follows:</p> <pre>DoD Standard 5015.2 <!-- Hyperlink to Joint Interoperability Test Command's 5015.2 Records Management Design Criteria Standard, link active on 01/03/02 --></pre>
<p>Hyperlink to new page within same web site</p>	<p>When a page includes a hyperlink that transfers control to another page <i>in the same web site</i>, it would be necessary to insert comments describing the hyperlink only when the site was not being scheduled <i>in toto</i> for the same retention (and those comments could reference the series containing the destination of the hyperlink).</p>

Another alternative would be to produce what is in effect a bibliography for all of the hyperlinks referenced within the content pages comprising a site. This would consist of a list of all URLs referenced by hyperlinks, along with a description of what each URL refers to (much as in the comment used in the previously suggested method).

Up

3.6 Who is responsible for managing web content pages?

Agency program personnel create and manage records as evidence of their business operations. Some of these records may be posted on an agency's web site. In addition to the original records, *the web version of content pages must also be managed*, as they are used for a different purpose from the original records.

Content pages on the web site may originate in any program areas within the agency. The agency should establish clear guidelines for managing records on web sites. The guidelines should specify whether the program office or the webmaster's office (or other office responsible for the web site) is responsible for implementing the records management policies of the agency for these records. The program office and the personnel responsible for agency web operations may each have specific responsibilities in this area. Ideally, agencies should have a team of individuals, including program staff, web management staff, and records management staff who develop records management plans for the web site. Responsibilities include both the development of records schedules

and their implementation.

Web content pages may be scheduled as records of the (program office) content owner or the web services providers. The decision should parallel that of other responsibilities—e.g., the responsibility for keeping content current, setting security levels, and identifying access requirements. (See SCHEDULING WEB RECORDS for additional information on scheduling web records.)

Each agency decides which office will be responsible for *implementing* records records schedules. If the web content pages are scheduled as records of the relevant program office, that office also implements the schedule. The program office will need to establish procedures to ensure that the schedule is properly implemented, including notification to the web operations staff when web content records need to be destroyed.

Up

3.7 Who is responsible for managing web management and operations records?

Agency personnel who manage the web site are responsible for managing the contextual and structural records necessary to adequately document agency web site operations. Web management records are managed the same as other program records in the agency. Web management records provide context and structure for web sites and do not present the same complexities as content records, which are frequently revised or replaced. Hence, standard records management techniques should be sufficient.

Up

¹ For guidance on whether records are trustworthy for legal purposes, consult your Office of General Counsel.

² ISO/TR 15489-2:2001, Information and documentation - Records management - Part 2: Guidelines. See <http://webstore.ansi.org/ansidocstore/find.asp>.

³ NOTE: Depending on the mechanism for producing a web site snapshot (e.g., simple device backup), this *may not* meet NARA transfer requirements for those web sites appraised as permanent. NARA will subsequently develop more detailed guidance regarding transfer requirements for web sites appraised as permanent.

⁴ Site map could be graphical or an index listing of directories.



Page URL:

U.S. National Archives & Records Administration
700 Pennsylvania Avenue NW, Washington, DC 20408 • 1-86-NARA-NARA • 1-866-272-6272