

U.S. Department of Energy Office of Legacy Management



Procedure: 243.2

Effective: 8/2/06

SUBJECT: PRESERVATION OF WEB SITE-RELATED RECORDS

1. PURPOSE.

To provide U.S. Department of Energy (DOE) Office of Legacy Management (LM) personnel with a framework to manage web site-related records.

2. CANCELLATION. None

3. REFERENCES.

- a. 44 United States Code (U.S.C.) Chapter 29, Records Management
- b. 36 Code of Federal Regulations (CFR) Chapter XII, Subchapter B, Records Management
- c. DOE Order 243.1, Records Management Program
- d. LM Procedure 200.4, Records Management
- e. LM Procedure 200.2, Electronic Records
- f. LM Procedure 200.2-1, Electronic Recordkeeping Systems and Electronic Information Systems
- g. DOE Standard DOE-STD-4001-2000, Design Criteria Standard for Electronic Records Management Software Applications
- h. Office of Management and Budget, Policies for Federal Agency Public Websites, December 17, 2004

INITIATED BY: Business and Resource Management
NO. OF PAGES/ATTACHMENTS: 7 pages, 3 attachments

- i. National Archives and Records Administration (NARA) Guidance on Managing Web Records, January 2005

4. DEFINITIONS.

- a. Content -- Any material that is available on a Federal public web site. For agency intranet web sites, the term refers to any material that is available on the intranet site.
- b. Content Management System (CMS) -- Software that facilitates web site administration by providing web-related functionality that may include templates for web content construction, on-the-fly page creation from databases, version control, workflow, and import/export functionalities.
- c. Disposition -- A broad term which may refer to any of the following:
 - Destroying records;
 - Offering and transferring those records to NARA;
 - Retiring or transferring records to a records storage facility;
 - Transferring records from one office or agency to another; and
 - Donating records to a Government or non-government entity.
- d. Domain name -- Defines the administrative boundaries and content of an agency's web site. LM's assigned domain is: lm.doe.gov.
- e. Electronic Record -- Any information that is recorded in a form that only a computer can process and that satisfies the definition of a record.
- f. File Plan -- A comprehensive outline specific to a project, organization, or function that defines its records series, file organization, records custodians, active file locations, file transfer instructions, file retention and disposition instructions, and other project- or organization-specific instructions that provide guidance for effective records management.
- g. Freedom of Information Act (FOIA) -- Law enacted in 1966 and codified at 5 U.S.C. 552. The basic purpose of the law is to ensure an informed citizenry in order to hold the Government accountable.
- h. Hypertext Markup Language (HTML) -- The set of markup symbols or codes inserted in a file intended for display on an Internet browser page.
- i. LM Personnel -- Federal and contractor personnel associated with LM.

- j. Privacy Act -- Provides safeguards against unwarranted invasions of privacy through the misuse of records by Federal agencies by restricting disclosure of personally identifiable records maintained by agencies; granting individuals increased rights of access to records maintained about them; granting individuals the right to seek amendment of records maintained about them upon a showing that the records are not accurate, relevant, timely or complete; and establishing a code of “fair information practices” that requires agencies to comply with statutory norms for record collection, maintenance, use, and dissemination.
- k. Program Records Official (PRO) -- Individual who ensures that all LM records management practices are properly executed.
- l. Records -- Include all books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the U.S. Government under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations or other activities of the Government, or because of the informational value of data in them.
- m. Records Liaison Officer (RLO) -- Individual(s) designated by the PRO to oversee the LM records management program in cooperation with the DOE Records Officer.
- n. Records Management Application (RMA) -- 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.
- o. Risk Assessment -- An examination of the potential harm that may result from exposure to certain hazards.
- p. Site Map -- A display of a web site’s hierarchy, similar to an organization chart. Typically, site maps break down a web site’s content into increasingly specific subject areas to help the visitor understand its structure, from the main navigation pages to their subordinate pages.
- q. Snapshot -- A copy of web site files at a certain point in time. A snapshot includes all documents available to the public that are located on the agency web site. Processes that automate snapshots are referred to as web harvesting.
- r. Stakeholder Program -- LM organization that handles dealings with intergovernmental organizations, states, Congress, and other federal agencies; interacts with stakeholders to develop outreach strategies to promote LM policies,

programs, and initiatives; and serves as a focal point for interactions with state and local regulatory authorities.

- s. Unclassified Controlled Information (UCI) -- Unclassified information that may be exempt from public release under FOIA and for which disclosure, loss, misuse, alteration, or destruction may adversely affect national security, governmental interests, or personal privacy.
- t. Web Site -- NARA defines a web site as a related collection of web content identified by a domain name. For the purposes of this procedure, web sites include the principle public LM web site (www.lm.doe.gov), web sites transitioned to LM control (e.g., rfets.gov), and the LM intranet portal (lportal.gjo.doe.gov).
- u. Web Site Administrative Records -- Federal records generated by the web site program office in the course of its operations.
- v. Web Site Content Records -- Web site content determined by the agency to meet the statutory criteria in 44 U.S.C. 3301 for being deemed a Federal record.
- w. Web Site-Related Records -- Web site-related records include (1) web content records, which represent information presented on a web site, and (2) web site administrative records, which provide evidence of the management and operations of the web site.
- x. Webmaster -- Person responsible for the design, implementation, management, and maintenance of LM web sites.

5. QUALITY CONTROL.

The RLO reviews this procedure as necessary, but at least annually, to accommodate changing conditions within LM and to ensure compliance with applicable laws, regulations, and DOE requirements.

6. RESPONSIBILITIES.

- a. The PRO ensures that designated LM Personnel maintain and retain web site-related records according to established procedures and ensures that appropriate risk assessment activities are conducted for LM web site-related records activities.
- b. The RLO:
 - Is responsible for ensuring that web site-related records are identified and are maintained according to the LM File Plan and protected accordingly based on the identified risks.

- Provides guidance to the Webmaster concerning identification of web site-related records.
- c. The Webmaster:
- Ensures that content not appropriate for public release (e.g., FOIA-exempt information, UCI) is not made available via an LM web site.
 - Ensures that copyrighted material is not published to an LM web site unless permission is secured from the material's owner and documented.
 - Captures and maintains records necessary to document LM's web site-related activities, while ensuring that content changes are properly tracked and documented.
- d. LM Personnel, when designated by the RLO or Webmaster, are responsible for supporting web site risk assessment activities.

7. TRAINING REQUIREMENTS.

Personnel with responsibilities for creating, maintaining, or dispositioning web site-related records should be cognizant of the applicable requirements and processes within this procedure and related DOE directives and NARA guidance.

8. DOCUMENT CONTROL.

- a. The current and official controlled hard-copy version of this document shall reside with the Directives Manager.
- b. The Directives Manager shall place the most current version of this procedure on the LM Intranet in a read-only format.

9. PROCEDURE.

LM uses official web sites to assist in carrying out its mission and to disseminate information to the public and its stakeholders. LM web site content and administrative information meeting the definition of a Federal record must be preserved for as long as needed for business operations and according to records management regulations.

Attachment A. illustrates the process of managing web site-related records.

a. Web Site Risk Assessment

The RLO, working with the Webmaster and other designated LM personnel, assesses web site risk as appropriate to identify records management controls,

record capture frequency, and record retention requirements. See Attachment B. – Risk Assessment.

b. Identification, Creation, and Maintenance of Web Site-Related Records

(1) The RLO:

- (a) Identifies all web site-related records. See Attachment C. for more information on the types of information that constitute web site-related records.
- (b) Reviews inherited web sites as they are transitioned to LM to identify web site-related records.
- (c) Notifies the Webmaster of web site-related records generated or inherited.

(2) The Webmaster:

- (a) Captures all identified web site-related records using an appropriate method, such as:
 - Web harvesting, or periodic snapshots, of all web site content.
 - Use of a CMS that automatically records all web site additions, modifications, and deletions.
 - Managing the live version of the web site along with a detailed log of all site changes.
- (b) Verifies that permission is secured and documented before using copyrighted material on LM web sites.
- (c) Consults with the Stakeholder Program concerning new content to be posted to an LM web site when there is a concern about the information's public release.

(3) The Stakeholder Program ensures new content considered for posting to an LM web site does not contain UCI or information exempt from public release due to FOIA or Privacy Act regulations.

- (4) The Webmaster:
 - (a) Maintains both previously captured and new web site-related record material (electronic or hard copy) according to the LM File Plan.
 - (b) Works with the RLO as necessary to ensure that web site-related records maintained electronically are stored in LM's approved RMA. Records captured electronically must remain viewable throughout their records retention periods.

- c. Disposition

The RLO dispositions web site-related records in accordance with the LM File Plan.

10. ATTACHMENTS.

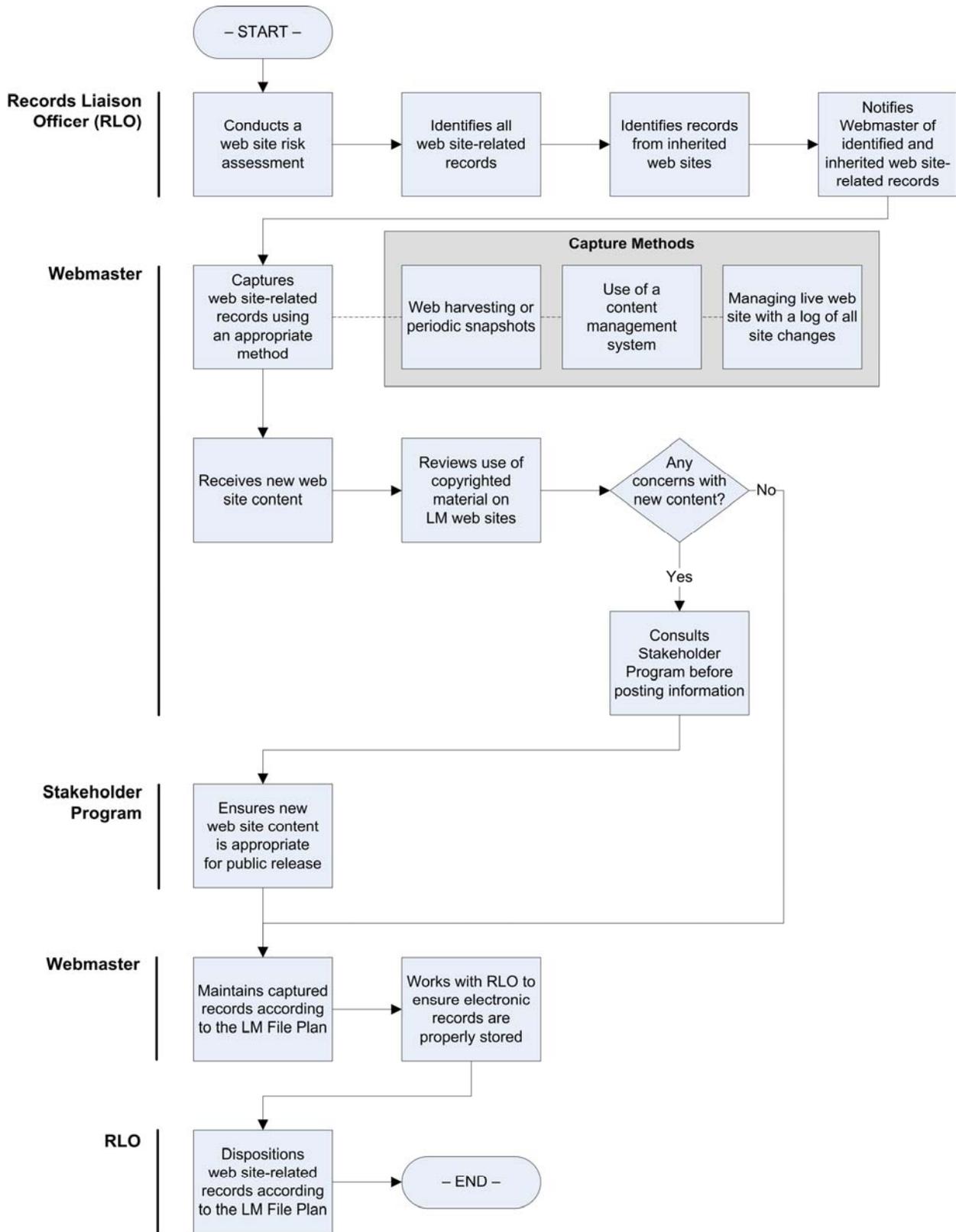
- a. Attachment A. – Web Site-Related Records Flowchart
- b. Attachment B. – Risk Assessment
- c. Attachment C. – Identifying Web Site-Related Records

Approved: Original signed by Celinda H. Crawford
 Director
 Office of Business Operations

Distribution: As required

The most recent and official controlled hard copy version of this document resides with LM's Directives Coordinator. An electronic version of the controlled document has been placed on the LM Intranet for employee use. Printed hard copies of this electronic version are considered uncontrolled documents.

Attachment A. – Web Site-Related Records Flowchart



The most recent and official controlled hard copy version of this document resides with LM's Directives Coordinator. An electronic version of the controlled document has been placed on the LM Intranet for employee use. Printed hard copies of this electronic version are considered uncontrolled documents.

Attachment B. – Risk Assessment

This attachment summarizes U.S. Department of Energy (DOE) Office of Chief Information Officer (OCIO) and National Archives and Records Administration (NARA) recommendations concerning risk assessments.

While not prescribing a uniform approach to risk assessment, the DOE OCIO recommend that assessments include certain elements: the nature and sensitivity of the web site's information; who uses the information; and problems that might arise if the information is incorrect, out of date, or lost. The risk assessment also should indicate the appropriate capture frequency and security required to maintain the trustworthiness of the web site records.

Many records management-related risks associated with web sites relate to technical vulnerabilities that could result in negative programmatic consequences. These consequences include the inability to verify site information at any given time, inability to document agency decisions relating to web operations, and dissemination of misinformation.

a. Assessments

A records management risk assessment addresses the possible consequences of untrustworthy or unrecoverable records, the likelihood that a damaging event will occur, and the costs associated with corrective actions. LM considers the following four factors when determining levels of risk:

- (1) *Threats* – Include unauthorized destruction of web site records, inability to reconstruct web site views, and inability to document web site policy decisions.
- (2) *Visibility* – The level of public awareness of the web site.
- (3) *Consequences* – The negative impact if web site records are untrustworthy or unrecoverable.
- (4) *Sensitivity* – The Office of Legacy Management's (LM) assessment of the importance of the web site records and operations.

NARA guidance for risk assessment and analysis suggests using a combination of techniques to generate a list of potential risks to evaluate. The techniques include interviewing personnel with subject matter expertise and conducting brainstorming sessions in which personnel consider risks to analyze. NARA further recommends using a rating system to calculate a risk factor for each potential risk event.

b. Mitigating Risk

LM personnel use the risk factor ratings calculated during the assessment to determine how to most effectively protect web site records. Risks with high overall rankings should be addressed before considering solutions to risks with lower rankings.

In addition to taking action to address specific risks, the DOE OCIO recommends the following measures as part of a broad effort to mitigate overall risk:

- (1) Review existing web management policies and procedures and determine whether additional controls are needed.
- (2) Develop a plan to address records issues (e.g., types of records needed to document web-based activity) and cyber security issues.
- (3) Develop and revise retention schedules for web records as appropriate and obtain NARA approval. (While the current NARA-issued General Records Schedules [GRS] do not include specific disposition authority for web site-related records, some items in the GRS can be applied to web site management and operations. These include GRS 14, GRS 21, and GRS 24.)
- (4) Document the systems used to create and maintain web records.
- (5) Ensure that web records are created and maintained in a secure environment that protects the records from unauthorized alteration or destruction.

c. Management of Web Site Content

When LM web sites are significantly changed or redesigned, changes must be captured as web records. LM personnel may decide to create the new web site record using a harvesting technique, a content management system (CMS), or a basic change tracking option.

- *Harvesting* – A harvesting tool or snapshot captures a web site as it existed at a particular point in time. If the harvesting function is employed, personnel include all content pages and a site map showing the page relationships. Risk rankings influence harvesting frequency and the need to track changes between snapshots.
- *CMS* – A CMS keeps the web site updated by managing the creation, modification, and removal of content from the website. A CMS must support creation of audit trails. A backup, or mirror, of the site must be available in case of equipment failure or other catastrophe. CMS software tracks versions of web content. Major changes to the site's directory structure can be captured by producing a new site map at the time of major revision.
- *Alternative for low-risk sites* – An option available for low risk web sites is to manage the live versions of the web site content pages while the pages are up on the web site. The current posted version of the web site and the standard operating procedures and change log may be sufficient to document web records.

When using the harvesting or CMS method, a certified records management application (RMA) must be used to store electronic versions of web site-related records. Certification means that the RMA complies with DOE Standard DOE-STD-4001-2000, Design Criteria Standard for Electronic Records Management Software Applications.

It is possible that hyperlinks will not continue to function in the preservation copy of the web content records. When generating hyperlinks to web sites not under LM's control, web site developers must insert hypertext markup language (HTML) comments after the hyperlink that include the name of the site and a content summary. Alternatively, LM personnel may develop a listing of a web site's external hyperlinks. The listing includes each hyperlink and a description of the content on the hyperlinked page.

d. Maintaining Trustworthiness

The risk assessment and web site management activities discussed in this section are designed to ensure the trustworthiness of LM's web site-related records. The amount of effort invested in ensuring an individual record's trustworthiness is based on the value of the information and the level of risk associated with the record.

A trustworthy record has four important attributes: reliability (the content is accurate), authenticity (the record is what it purports to be), integrity (the information is complete and unaltered), and usability (the record can be located and retrieved).

A trustworthy record preserves the content of the record and information about the context in which it was created and used. It also may be necessary to preserve information about the record's structure or format.

Attachment C. – Identifying Web Site-Related Records

LM must manage records from web site operations, including those needed to ensure trustworthiness (reliability, authenticity, integrity, and usability) and to document agency programs. Web site-related content that meets the Federal definition of a record must be managed as such. Common categories of web site-related records include:

a. Administrative Records

- *Web site program operations records* – Program management files that document policies and procedures for agency web site operations, including those that:
 - Provide detailed procedures 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;
 - Demonstrate the development of policies and procedures to ensure web sites are accessible to people with disabilities (compliance with Section 508 of the Rehabilitation Act of 1973); or
 - Otherwise explain or illustrate site development and management procedures.
- *Web site design records* – Records such as graphic design layouts for pages or embedded image maps and meeting minutes documenting the production of such.
- *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 assessment and selection of any commercial software;
 - Records describing customization of web-related software; and
 - Documentation relating to web-related software.
- *Web site logs and statistical compilations* – Records including raw data or summary logs of user access, site posting logs (indicating when pages were posted, updated, or removed from the site), system load and search result statistics, and reports containing such items.

- *Software configuration files* – Files used to operate the web site and establish its style including server environment configuration specifications.

b. Content Records

- *HTML-encoded pages* – The content pages composing LM web sites, inclusive of the HTML language.
- *Records generated interactively on the web site* – Records created dynamically in real time when a user interacts with the LM web sites (e.g., on-the-fly, text-based page creation, forms completed online).
- *Additional content files referenced within HTML-encoded pages* – Files having the ability to “self-execute” (e.g., scripts, applets) as well as files that are static (e.g., these include graphic files, multi-national character sets). Both self-executing and static pages are external to the HTML-encoded content pages but referenced in the HTML syntax.
- *Comprehensive list of URLs referenced by a web site’s hyperlinks* – 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.
- *Copyrighted web content and records documenting the use of such content* – Include works for hire, such as custom produced graphics files. Additional records may be necessary to document appropriate use of copyrighted material.
- *Web site map* – A linked, graphic, or text-based display of a web site’s hierarchy, similar to an organization chart.