INFORMATION TECHNOLOGY ADVISORY BOARD

Electronic Records Committee and Internet Task Force

Guidelines for Managing Records on Kansas Government Agency Web Sites

Version 1.0 January 2004

Guidelines for Managing Records on Kansas Government Agency Web Sites

These guidelines were drafted by the Kansas Web-based Records Management Task Force from the Fall of 2001 through the Summer of 2003. The task force, a broad-based group comprised of representatives from all branches and levels of Kansas government, was co-sponsored by the Electronic Records Committee (ERC) and the Internet Task Force (ITASK), subcommittees of the Kansas Information Technology Advisory Board (ITAB). The guidelines were endorsed by ITAB on October 14, 2003, and by the State Records Board on January 8, 2004, as best practices for managing government records in the web environment.

Table of Contents

1   Introduction

1.1   Purpose

This document is intended to assist Kansas state and local government agencies develop internal policies and procedures for creating, capturing, managing, and retaining web-based records for as long as those records have value. These guidelines apply when:

> materials exist on an agency web site **and . . .**
>
> the materials qualify as records[1], **and . . .**
>
> the materials have not already been adequately captured in an agency system capable of meeting records retention requirements.[2]

Recognizing the differences in the size, scope, and nature of agency responsibilities, the guidelines outline a range of potential strategies for managing web-based records. These guidelines recommend a risk-based approach to agency decisions regarding priorities, schedule, and level of effort for implementation.

Information technology has altered the way government conducts business, **but** digital technologies have not changed government recordkeeping responsibilities.   In this age of e-government, it is important to understand that all Kansas government agencies are accountable for resources delivered or made available over their web sites both now and in the future. They must be able to meet any legal obligations and community expectations for evidence of present and past positions, advice, guidance, transactions, or instructions delivered over their web sites.   When managed systematically, records are an asset; but when handled haphazardly, they are a liability. Agencies have the responsibility to all citizens of Kansas to manage government records effectively and efficiently, regardless of the technology used to create the records.

While this document focuses on records on government web sites, Kansas state and local government agencies should establish an overall policy for managing all electronic records. The broad electronic records management and preservation policy should include coverage of web sites, electronic mail or messaging, word processing files, spreadsheets, databases, optically stored images, audio and video recordings, multimedia files, and other topics of importance to the agency. An agency training program should ensure that all employees know how to recognize when materials qualify as records, what to do with records, and when to consult the agency records officer.

Terminology used in these guidelines as well as the legal framework for records management in the state of Kansas are contained in Appendices A and B respectively. Refer to Appendix C for a discussion of the variety of web sites.

---

[1] A definition of "records" is provided in Appendix A.
[2] Charles R. McClure and J. Timothy Sprehe, *Guidelines for Electronic Records Management on State and Federal Agency Web Sites*, February 1998, funded by the National Historical Publications and Records Commission, URL: http://slis-two.lis.fsu.edu/~cmcclure/guidelines.html.

1.2 Key Questions to Consider

The guidelines will help government agencies answer several key questions as a prelude to developing a records management policy for agency web sites:

- Which elements of the agency web site are records?
- How long is the agency legally required to retain each category (series) of records?
- If web-based records are not adequately captured elsewhere, how can these records be captured into an agency system capable of meeting records retention requirements?
- Can the agency be held accountable for something it had on its web site six months ago? A year ago? Five years ago?
- How can the agency build web-based records management into overall web site management?
- What are the records management responsibilities of staff members involved in web site development and management, especially the webmaster and content developers?
- Who in the agency should ensure records management responsibilities are fulfilled?
- How can the agency build staff awareness and compliance with web records management procedures?

2 Best Practices for Managing Records on Agency Web Sites

The following six Best Practices are recommended to apply to any agency web sites. In this document, **web site** is a generic term that refers to all types of web-based sites, including public web sites, extranets, and intranets. Appendices D, *Preservation Strategies*, and E, *Maintaining Web-Based Records Over Time*, provide details on strategies and options on maintaining these records.

2.1 Best Practice 1: Assess Risk of Accountability for Agency Web Resources

Government agencies should evaluate the risk regarding its web presence according to four main factors: purpose of site, public visibility, complexity and change management. See Section 3, Assessing Accountability, for more information.

2.2 Best Practice 2: Develop Policy

Government agencies are encouraged to include web-based records in their records management policies consistent relevant guidelines and laws (see Appendix B: *Legal Framework*). Strategies should be included for maintaining access to the records in compliance with their retention requirements.

2.3 Best Practice 3: Identify Records on Agency Web Site

Government agencies must identify records which exist on their web site(s), or which are about to be placed online. Government agency web sites may contain individual documents recording transactions or publications not created or kept in any other form.

2.4 Best Practice 4: Capture Records into a System

Government agencies should ensure that records posted to or received from their web site(s) and are not captured elsewhere are captured into a system capable of meeting retention requirements as specified in the general and/or agency retention/disposition schedule. These systems should guarantee the authenticity, reliability, and accessibility of the records.

2.5 Best Practice 5: Develop Training Program

An agency training program should ensure that all employees, but especially those who work in information systems and web site-related areas, know how to identify materials that qualify as records, what to do with records, and when to consult the agency records officer.

2.6 Best Practice 6: Conduct Periodic Reviews

Government agencies should develop a program for conducting regular reviews to identify the effectiveness of the web-based recordkeeping approach, measure performance, evaluate preservation strategies, identify any corrective action or enhancements required, and minimize the agency's exposure to risk through procedural failure.

Conducting a review is especially important when the web site is modified, or when business processes change. A review is an opportunity to identify any changes in circumstances that may mean a different recordkeeping approach is required.

3 Assessing Accountability

Ultimately, governmental agencies must cope with accountability for their web sites. At its core, "accountability means the ability to reconstruct an accurate picture of the past through records that can be used as evidence."[3] Agencies are accountable to many audiences. As tax-funded entities, they are ultimately accountable to the citizens of Kansas. They are accountable to the officials those citizens elect to represent them in the Legislature, and are responsible to abide by the laws enacted by that body. They are also accountable to the Governor or Supreme Court Chief Justice, and/or other specific elected or appointed department head which oversees each agency. And they are accountable for the content placed on agency web sites, which not only provide a face of the agency to the public, they also provide information to users who base decisions upon that information. Accountability includes the possibility of agencies being called into account – whether a court of law, in the court of public opinion, administratively, or in the court of history – for what appears on their web sites. To minimize an agency's accountability exposure, the agency must manage the information contained on its web site.

3.1 Factors to Consider

What factors should an agency consider when determining a course of action for keeping web-based records? There are four main factors which should guide the agency in evaluating the risk faced by the agency regarding its web presence. For a more detailed description of risk exposure analysis for agency web sites, refer to Appendix F.

---

[3] McClure and Sprehe

a. **Purpose of web sites**
Web sites are used for a number of purposes, each of which poses certain risks for a government agency, but the level of risk depends both on the nature of an agency's business and the extent of its dealings with the public.

b. **Public visibility of the agency**
The nature of an agency's business and the extent of its dealings with the public affect its public profile. For example, agencies that provide assistance and services to the public, and those which are involved in high-profile operations such as regulatory and legislative work are subject to much higher levels of public scrutiny (and criticism) than those which work behind the scenes.

Circumstances change and agencies therefore should review their level of public visibility periodically, and take the appropriate actions based upon this review. It should be noted that due to the current novelty of conducting government business in the web environment, all agencies establishing a web presence are more visible to the public, thereby raising the risk of being called into account.

c. **Complexity of web sites**
Web sites range from sets of static documents to dynamically generated sites that are more like applications than publications. From a recordkeeping perspective, web sites that are comprised of static documents and incorporate little or no interactivity are relatively simple to deal with. On the other hand, sites that incorporate high levels of interactivity and personalization and are made up of dynamically generated pages are very complex and prove more difficult to preserve effectively.

d. **Web site change management procedures**
Most web sites are periodically updated to provide current information. Changes may take place on a regular basis, with new or updated material uploaded to the site once a day, week, or month. In other cases, changes may be made on an "as required" basis or, in the case of database-driven sites, in real time. For the purposes of accountability and site maintainability, it is important that records of web site content changes are made and kept.

Web sites which require frequent changes that are not regularly scheduled are likely to pose the most degree of risk to the agency because such changes are the most difficult to track. Of course, the extent to which this factor contributes to the overall level of risk is dependent on the other risk factors.

3.2 Accountability Exposure Analysis

An agency should be able to demonstrate the likelihood of being called into account for the contents of its web site, and being called to prove or disprove the existence of particular content on its site. This applies to both past and present versions of the site.

The four factors outlined above interact to form a level of risk, and need to be considered in the context of the current state of recordkeeping within the agency. The key aspects of accountability are the ability to

reconstruct what occurred in the past and the exactness of reconstruction that circumstances may demand. Cost of reconstruction is an important factor, but so is the likelihood the reconstruction will be required. It makes no sense to create expensive procedures for reproducing exactly what occurred on a web site in the distant past when the agency is never likely to need an exact reproduction.[4]

Accountability is at the heart of archives and records management, and records management accountability consists of discharging responsibilities to keep records under the applicable statutes as outlined in Appendix B, *Legal Framework*, below. These statutes also require Kansas government agencies to follow State Records Board-approved general and agency-specific retention and disposal schedules. Exactly what web site records agencies should keep is a matter for discussion, study, legal opinion and senior management decision-making within each agency, but it is clear that there is an essential recordkeeping mandate contained in the cited statutes.[5]

Based upon the level of accountability exposure, agencies will need to take different approaches to manage the information contained in their web sites. A high level of accountability exposure indicates a higher likelihood of needing to reconstruct materials that were once available on the web site. It is recommended that agencies focus their efforts on addressing records management issues for high exposure web sites before moving on to web sites with less exposure. The analysis table and worksheet provided in Appendix F offer guidance to an agency on the likelihood of needing to reconstruct its web site, or determining the level of exposure an agency faces in relation to its web presence.

It is important to remember that the four risk factors (purpose, visibility, complexity, and change management) create a compounding effect in relation to each other. For example, an agency with numerous web sites that provide access to static records already available elsewhere, will not be as open to exposure as an agency with a single web site which is highly dynamic and creates records at the specific instructions of an individual web site user. When evaluating your own web presence, be sure to consider the total compounded weight of all risk factors, when combined with the other factors.

4 Developing Policy

Technical experts, content managers, and the agency records officer should all be involved in incorporating web-based records into an agency's records management policy. They jointly should determine each site's accountability exposure and seek advice, where appropriate, from legal counsel. Web sites may be classified as having low, moderate, or high accountability exposure. The web site technical experts, content managers, and records officer should jointly determine appropriate recordkeeping responses to low, moderate, and high accountability exposure sites.

4.1 Involving Internal Stakeholders

The agency's records management policy should address the activities of all staff, including third party web services providers, involved in web site creation, administration, and use. Key groups to include are:

- agency management,
- web content creators,
- web site technical experts,

---

[4] McClure and Sprehe
[5] McClure and Sprehe

- web site business process owners, and
- records officers.

4.1.1 Agency Management

Agency management should be aware of the agency's need to be accountable for its web-based activities. Their involvement should also include evaluating the levels of risk and prioritizing activities and investments to mitigate risk as it relates to various portions of the web site. An understanding of the retention and disposition requirements for web-based records can help determine priorities for mitigating risk.

This involvement is critical if agency staffing demands prevent the development and implementation of a comprehensive web-based records management policy. The agency director, division directors, and legal counsel should all understand which web-based activities are being documented and which ones are not, and the accountability implications of either having no formal policy or of having a limited records management policy.

4.1.2 Web Content Creators

Content creators and subject experts coordinate with technical experts to see that new information is added, obsolete information deleted and old content is maintained in conjunction with the appropriate metadata. For materials that are specifically prepared for web site posting, the content manager responsible for preparing the materials should coordinate with the appropriate information technology staff to ensure that the responsibility for keeping the record copy of the materials is fulfilled and that the record copy is transferred into a system capable of meeting retention requirements either before or after web site posting. If a third party service provider is used to create and/or maintain web content, contracts with the provider should include provisions for the management of appropriate web-based records.

4.1.3 Web Site Technical Experts

Web site technical experts work with the content creators, records officers, and business process owners to ensure that links within the system for retaining records maintain functionality. Web site technical experts should ensure that they are following established records management procedures for purpose-prepared materials. Again, if third party providers are used, they should be required to follow the same records management procedures.

4.1.4 Business Process Owners

Web site business process owners work with the web site technical experts, content creators, and the record officer to maintain current and operational business transactions on the web.

4.1.5 Agency Records Officers

The agency records officer coordinates with the stakeholders listed above to ensure that agency specific and general retention and disposition schedules are followed and to interpret State Records Board decisions. The records officer should participate in development of procedures for web content and confirm that content managers and web site technical experts are following procedures.

4.2 Considering External Stakeholders

Kansas government web sites usually have a number of external stakeholders whose needs and interests should be considered when establishing a web-based records management policy. External stakeholders include:

- the public – citizens have the right to access information;
- State Archives, Kansas State Historical Society - the preserver and disseminator of records and information;
- State Library - the disseminator of information;
- State Records Board - authorizes retention periods and disposition of records;
- agency managers and administrators – the personnel responsible for agency actions;
- Information Technology Advisory Board (ITAB) - the body which functions as the technical resource to the Chief Information Technology Officers (CITO) of each branch of Kansas government;
- Information Technology Executive Council (ITEC) – the body responsible for adopting information technology policies, project management procedures, the statewide technical architecture and the strategic information management plan;
- Internet Task Force (ITASK) - the body which promotes the development and coordinates policies, procedures, and best practices for application with the Internet; and
- Information Network of Kansas (INK) – the quasi-state agency which can assist with implementation of policies and guidelines.

Appendix A – Terminology

Access
The right and means to use, view, copy, and potentially modify a record in some manner.

Accountability
For this document, accountability is the ability to reconstruct an accurate picture of the past through records that can be used as evidence.

Accountability exposure
The real likelihood of being called into account for what appears on an agency web site.

Authenticity
The qualities of a record that ensure that it remains reliable over time.
Related terms: *reliability*
                   authentication

Authentication
The verification of the identity of a resource user; could be established through use of a digital or electronic signature.

Authorization
The process of giving users varying levels of access to resources based on their identity.

Checksum
A simple error-detection scheme in which the information that is to be reformatted (see section 8.3) is assigned a numerical value based on the number of set bits in the information. A formula is applied to the information after reformatting and a check is done to make sure the numerical value is the same. If not, it can be assumed that the information has been corrupted in the reformatting process.

Content
With regard to an electronic record, content is that which conveys information (e.g. text, data, symbols, numerals, images, and sound)
Related terms: *context*
                   structure

Content Management System
A system consisting of software that manages documents for Web sites. It provides for the storage, maintenance and retrieval of HTML and XML documents and all related elements.

Context
The circumstances surrounding the creation of records including background information that enhances understanding of technical and business environments to which the records relate (e.g. metadata, application software, logical business models) and the origin of the records (e.g. address, title, link to function or activity, agency, program or section).
Related terms: *content*
structure

Control
Actions which minimize risk and increase the likelihood of achieving agency objectives.
Related term: *risk*

Digital Object Identifier (DOI)
A persistent identifier of any of the various physical objects that are "manifestations" of intellectual property, such as printed books, CD recordings, videotapes, journal articles.
Related term: *persistent uniform resource locator*

Digital Signature
A cryptographic process used to assure information authenticity and integrity.
Related terms: *authenticity*
integrity

Disposition
The treatment of a record series at the end of its life cycle – it is either destroyed or maintained permanently by the creating agency or the State Archives.
Related terms: *retention period*
records retention and disposition schedule

Exposure
Risks which are not well controlled; that amount of risk left over after any existing controls have been implemented.
Related terms: *risk*
control

Extranet
An intranet, or portion of an intranet, to which an agency allows access by selected external entities, such as individuals or companies considered partners of the agency.

Government agency
Includes the state or any political or taxing subdivision of the state, or any office, officer, agency instrumentality thereof, or any other entity receiving or expending and supported in whole or in part by public funds appropriated by the state or by public funds of any political or taxing subdivision of the state. This definition does not include an entity that receives payment from public funds for property, goods or services of such entity, or any member of a Kansas judicial court.

Hash Function

An algorithm that turns a variable-sized amount of text into a fixed-sized output (hash value). The hash value is then used to check the text against corruption or tampering.

Intranet

A site, internal to an agency, which uses the same network technology and protocols as a public web site, but is for internal or restricted external use only.

Metadata

Metadata is usually defined as "data about data." It consists of information (e.g. creator, creation date, record series title, etc.) that is used to provide intellectual control of, and structured access to, other information. Metadata also contributes to ensuring the reliability and authenticity of electronic records.

Persistent Uniform Resource Locator (PURL)

A PURL points to an intermediate resolution service, rather than pointing directly to the location of an Internet resource as a URL does. The resolution service associates the PURL with the actual URL and returns that URL to the client, which can then complete the transaction in the normal fashion.

Related terms: *uniform resource locator*
                               digital object identifier

Privacy

A primary concern in the information age. State and federal privacy laws, including KORA, HIPAA and FERPA, cover areas including collection, use, disclosure (voluntary and involuntary), notification of disclosure, security (technical and physical) and storage (retention and disposition) of records that have personally identifiable information.

Public web site

A collection of Uniform Resource Locators (URLs) in the control of one administrative entity made accessible to the public via the World Wide Web.

Recordkeeping System

An information system that captures, manages and provides access to records over time. Recordkeeping systems are different from generic information systems in that they maintain linkages to the activities they document and preserve the content, structure and context of the records.

Unlike most other computer information systems, recordkeeping systems must often accommodate records that exist in more than one format (e.g. parallel paper case files and electronic case management systems). Recordkeeping systems should be able to identify all records, active and inactive, and the version of the computer software that supports access. They should be able to identify records stored off-line and off-site and on all media.

Records

In this document, the term "records" refers to government records which are "all volumes, documents, reports, maps, drawings, charts, indexes, plans, memoranda, sound recordings, microfilms, photographic records and other data, information or documentary material,

regardless of physical form or characteristics, storage media or condition of use, made or received by an agency in pursuance of law or in connection with the transaction of official business or bearing upon the official activities and functions of any governmental agency." (K.S.A. 45-402(d))

Records officer
In this document, the term records officer refers to the individual(s) within a government agency who has been given official or de facto responsibility for records management activities. It should be noted that K.A.R. 53-3-1 requires Kansas state agencies, including Regents institutions, to designate a staff member to exercise responsibility for preparing and updating agency records retention and disposition schedules and submitting the schedule to the State Records Board. The state agency records officer also oversees agency implementation of the schedule.

Records retention and disposition schedule
A records retention and disposition schedule is a timetable that identifies the minimum length of time a record series must be retained before it is destroyed or transferred to the State Archives. The State Records Board has statutory authority to approve records retention and disposition schedules submitted by Kansas state agencies and county government offices.
Related terms: *retention period*
                                   disposition

Record series
A record series is a set of records grouped together because they relate to a particular subject or function, or result from the same activity. All records fall into a record series, and each record series should be managed according to an appropriate records retention schedule.

Reliability
The qualities of a record that engender confidence that it is an accurate and trustworthy representation of a real event or activity.
Related terms: *authenticity*

Retention period
The length of time a given record series must be kept and made accessible, expressed as a time period (e.g., four years), an event or action (e.g., audit), or a combination (e.g., six months after audit).
Related terms: *disposition*
                                   retention and disposition schedule

Risk
Any occurrence or event that can have a significant adverse impact on the ability of an agency to achieve agency objectives.

Snapshot
A full and accurate copy of an agency's web site at a particular point in time.

State Archives
Official repository for government records that possess enduring value. The Kansas State Historical Society administers the Kansas State Archives. (K.S.A. 45-405)

State Records Board
A five-person board given statutory authority (K.S.A. 45-404) to set retention and disposition requirements for state and county government records. The board is comprised of a representative from the Attorney General's Office, a representative from the Department of Administration, the State Librarian, the State Archivist, and the Executive Director of the State Historical Society.
Related terms: *retention and disposition schedule*

Structure
The appearance and arrangement of the content of electronic records (e.g. relationships between fields, entities, language, style, fonts, page and paragraph breaks, links and other editorial devices).
Related terms: *content*
context

Uniform Resource Identifier (URI)
The generic term for all types of names and addresses that refer to resources on the World Wide Web.
Related terms: *uniform resource locator*

Uniform Resource Locator (URL)
The global address of documents and other resources on the World Wide Web. The first part of the address indicates what protocol to use, and the second part specifies the IP address or the domain name where the resource is located.
Related terms: *uniform resource identifier*
persistent uniform resource locator

Web portal
A user-centric web application that provides users access to resources they commonly use, in the language of their business. The portal might include, among other things, a search engine, proprietary content from a third party outside the agency (such as weather information), or links to related web sites, which are presented as one site and may be customizable for individual users.

Web site
In this document, "web site" is a generic term that refers to all types of sites built around the HyperText Transfer Protocol (HTTP) and HyperText Markup Language (HTML) including intranets, extranets, and public web sites.

Appendix B – Legal Framework

Statutes

Kansas

Government Records Preservation Act (K.S.A. 45-401 through 45-413)
This Act provides a definition of government records, declares them to be state property and prohibits their unauthorized destruction. It lists the duties and responsibilities of the State Records Board and designates the Kansas State Historical Society as the official State Archives for the State of Kansas. The Act also specifies the duties and responsibilities of the State Archivist and authorizes the State Archivist to access confidential information. It requires state and local agencies to cooperate with the State Records Board and the State Archivist while exempting legislative and judicial records from State Records Board control. The Act also stipulates the conditions for the destruction of records after microfilming.

Public Records Act (K.S.A. 75-3501 through 75-3518)
The Public Records Act defines records. It is the Act that establishes the composition and outlines the general responsibilities of the State Records Board. It requires agencies to comply with micrographic and optical disc standards established by the Board, and authorized the admissibility in court of micrographic and optical disc records. It establishes the State Records Center as the depository for inactive state government records, and provides guidelines for the use of acid-free and permanent paper.

Open Records Act (K.S.A. 45-215 through 45-223)
The Kansas Open Records Act (KORA) requires agencies to develop policies to provide prompt and convenient public access to government records for a reasonable fee. It describes 45 categories of records that are exempt from disclosure under KORA provisions, and provides that records exempted by KORA will be open to the public after 70 years unless closed by a specific statute or regulation.

Federal

Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g) (1974) (Buckley Amendment)
FERPA protects the privacy of student records. It protects any records kept by the institution that are directly related to the student. The records may be in any format including computer database, e-mail, or other medium. Universities may disclose "directory information". This includes name, address, telephone number, data and place of birth, participation in University-recognized activities, dates of attendance and degrees. This law protects rights of students to review and inspect educational records, to request to amend records, and to limit disclosure of personally identifiable information.

Health Insurance Portability and Accountability Act of 1996 (HIPAA) (PL 104-191)
A wide-reaching law designed to protect the privacy of health information; governs the acquisition, storage, use, and disclosure of health records; protects against "involuntary disclosure" due to network or other security breaches; mandates certain computer security practices along with physical security.

Privacy Act of 1974 (Amended in 1988) (PL 93-579)
Prohibits disclosure by Federal agencies of any record contained in a system of records, except pursuant to a written request by or with the prior written consent of the individual to whom the record pertains.

Government Records Defined

Kansas' records laws do not specifically address web-based records. However, the Kansas statutory definition of government records is sufficiently broad to encompass much of the information that exists on agency web sites:

> *"Government records" means all volumes, documents, reports, maps, drawings, charts, indexes, plans, memoranda, sound recordings, microfilms, photographic records and other data, information or documentary material, regardless of physical form or characteristics, storage media or condition of use, made or received by an agency in pursuance of law or in connection with the transaction of official business or bearing upon the official activities and functions of any governmental agency. (K.S.A. 45-402(d))*

Policies, guidelines, meeting minutes, reports and other static documents that frequently appear on agency web sites clearly meet the definition of government records and must be managed as such. Moreover, many agencies have implemented interactive technologies that facilitate dynamic web-based communications and transactions with citizens. Records of these communications and transactions should also be captured and maintained for as long as necessary to ensure agency accountability.

Records Retention and Disposition

K.S.A. 45-403 stipulates that government records are public property and cannot be destroyed without authorization from the State Records Board or through an approved retention and disposition schedule**. K.S.A. 45-404** grants the State Records Board the authority to "approve or modify retention and disposition schedules." The five-person board meets quarterly to consider schedules and specific records disposition requests submitted by state and county government agencies. Under the provisions outlined in the Government Records Preservation Act and the Public Records Act, a state or county government entity is legally obligated to retain all web-based records until it has explicit approval from the State Records Board to destroy such records.

Two types of retention and disposition schedules apply to government records in Kansas. General Records Retention and Disposition Schedules include guidelines for common record series maintained by many government agencies. Each agency should also have a specific Agency Records Retention and Disposition Schedule approved by the State Records Board, which contains requirements for record series that are unique to the organization and thus not listed in the General Schedule.

Access to Records

The Kansas Open Records Act requires that state and local government agencies make records available to the public unless they contain information that is exempt under K.S.A. 45-221 or other state and federal laws. The Act provides for public access to records regardless of their physical format or location. Web-based records are subject to all of the requirements of the Open Records Act. An agency must be able to provide access to records that existed on its web site at a specific point in time if their retention is required by the agency or general retention and disposition schedule, even if those records are not currently on the web site. Moreover, confidential records that are created and maintained on agency web sites must be restricted from unauthorized access. Robust security measures must be developed and deployed to protect confidential records on agency web sites. Agencies also should be aware of Americans with Disabilities Act requirements for web site accessibility, in particular, ITEC Policy 1210, Web Content Accessibility Guidelines for the State of Kansas - Guidelines by Priority (Version 2.0), and Web Content Accessibility Guidelines for the State of Kansas - Implementation Guidance.

Electronic Records Guidelines and Related Documents

This document on managing web-based records represents one in a series of guidelines related to electronic records management (ERM) developed through a partnership between the Kansas State Archives and the Kansas information technology community. Because of the technical expertise required to address records issues in the digital environment, the State Archives deemed it essential to work closely with government IT professionals in formulating electronic records management guidelines and best practices. Agencies have a responsibility to develop policies and procedures for the management of all of their electronic records, not just web-based records. Other documents that provide ERM advice to state and local government include:

Kansas Electronic Records Management Guidelines

Developed by the State Archives in November 1997 and updated in August 1999, these guidelines provide an overview of the problems associated with managing electronic records.

Kansas Electronic Recordkeeping Strategy: A White Paper

Published by the State Archives in December 1999 at the request of the Kansas Secretary of Administration, this document is an executive summary of the Kansas Electronic Records Management Guidelines. It provides a concise introduction to ERM issues.

Digital Imaging Guidelines

Issued by the State Archives in 1997, these guidelines offer advice to state and local government agencies to assist them in designing responsible digital imaging systems that may be used for creating or maintaining long-term or archival records.

Managing Electronic Mail: Guidelines for Kansas Government Agencies

The Information Technology Advisory Board approved this document in October 2001 to provide advice to government entities on the management of e-mail records. The Information Technology Executive Council approved the guidelines in May 2002.

ITEC Policy 2400 – Project Approval – Appendix E

The Information Technology Executive Council approved Policy 2400 in October 2000 to establish procedures governing the approval of State of Kansas information technology projects. Appendix E of the policy requires agencies seeking approval for a project to complete a series of questions related to electronic records management.

Kansas Statewide Technical Architecture (KSTA) – Chapter 20

The Electronic Records Committee developed a chapter for the KSTA that focuses exclusively on electronic records management issues. It is included as Chapter 20 of KSTA version 10.

ITAB Planning Guideline: Workflow, Imaging and Electronic Document Management

Also known as "State of Kansas Guide to Information Technology Investment," (October 1999) this report from the Statewide Workflow Subcommittee addresses the use of technology such as workflow, digital imaging and electronic document management systems.

ITAB Planning Guideline: Internet Site Development and Management

Also known as "State of Kansas Internet Technology Standards," (November 1999). Contact the Kansas Information Technology Office for more information.

Appendix C – Spectrum of Web-Based Resources

Web sites today come in a wide variety of forms and styles. They range from sites that present static information that does not change over time, to interactive sites that present information from databases that are updated by the minute and selected by the user. There are sites that collect data from the user, and then provide current and live information from a variety of sources reselected based upon the user's profile. Each of these sites may or may not be creating government records, but each site should be evaluated regarding the content presented on the site.

Static Web Sites

In its most basic form, a web site may be nothing more than a set of static documents within folders sitting on a server and tied together with hyperlinks. These documents share a common address – the domain name in the universal resource locator (URL), such as "www.accesskansas.org". A static web site maps URLs directly to file system locations, where the only interactivity provided is in the links that enable movement from one document to another or from one part of the site to another.

A variation of a static web site is the site that uses forms to collect information, such as comments or requests, from users. While these sites are still largely static in nature, agencies keeping records of such sites should also pay attention to:

- the information provided when the visitor fills in the form (usually stored in a "back end" system);
- the form itself; and
- the human-readable source code of the script or program which enables the form's functionality.

Dynamic Web Sites

Web sites that provide interaction or customization for the user, as opposed to presenting unchanging information, come in a variety of flavors. The simplest approach to a dynamic web site is a site that provides a user interface, to a database that changes over time and potentially presents new information to the user each time the site is viewed. Dynamic sites run the gamut from this simple interface to sites that present content based upon a user profile stored in a database that sits behind the web site, from sites that interact based upon user queries to accepting transactions from the user and providing services as part of transactions.

Non-Personalized Dynamic Web Sites

Non-personalized sites provide an unchanging interface to a behind-the-scenes database that is changing over time. A user will view different information on the web site during different sessions over time, while two users viewing the same site at the same time would see the same information. The changes could be made on a periodic basis by agency employees adding new, or changing existing, information to the database. Or the database might be changing in real-time as data is added or changed by an automated system, such as a web site that reports road conditions based upon remote sensors placed in the highway pavement.

Personalized Dynamic Web Sites

Another form of dynamic web sites presents unique information based upon data supplied by each user. The site stores data such as a user profile and provides information to the user based upon the profile; therefore two users would see different information if viewing at the same site at the same time. One example of this type of web site is the popular "web portal" which combine content from various sources into one screen for the user. This content might originate within the agency, or might be provided by a third party who owns and is responsible for the content it provides.

Business Transaction-Based Dynamic Web Sites

A third type of dynamic web site is a site where business transactions take place. A business transaction might include the purchase of a product or service by a user from an agency, the filing of a report by a user to an agency, or the renewal of a license by a user. With the trend toward "e-government", many more agencies will be creating these types of sites to provide goods and services to constituents.

Web Sites containing Non-HTML Content

A web site may also utilize such technologies as streaming audio and video, live chat, and the like that provide content within the web environment. Often (but not always) the content is non-textual and may be presented on either a static or dynamic site. The content may be offered to the user from a stored file, such as playing a stored video clip when the user clicks on a hyperlink, or it may be presented live, such as a live streaming audio track of a press conference. This content presents a new challenge to those responsible for managing web resources. Determinations should be made regarding the "recordness" of this content (e.g., is it a unique record or simply a copy of an existing record series in a different format?), the possible retention requirements if it is a record, and how to capture and store the content (especially in with streaming audio or video).

Appendix D – Preservation Strategies

This section outlines some options available for capturing and preserving information presented on web sites and records of web-based activity. It is important to remember that the options suggested are not mutually exclusive, but that in most cases agencies will pursue a combination of strategies designed to match their circumstances and requirements. It should also be emphasized that due to agencies existing in a variety of operational, technological and risk environments and with technology in a constant state of change and evolution, it is impossible to provide definitive guidance in this area. Not only will different strategies be required for different circumstances, agencies will also need to be flexible in the application of these strategies in order to deal appropriately with changes in their legal, technological and operational environments. Various commercial tools are available for managing the records of web-based activity. While it is impractical to list those tools in this document, the Kansas Statewide Technical Architecture lists those tools in Chapter 20, "Records Management and Preservation Architecture" (contact the Kansas Information Technology Office to obtain a copy of the entire Architecture).

The selection of appropriate strategies is contingent on the type or complexity of the web site being managed, the type of web-based activity in which the agency engages and the results of the agency's accountability exposure analysis of its recordkeeping requirements.

Agencies may consider strategies oriented toward static web sites or those aimed at dynamic sites, or, better still, a combination of the two. Appendix C discusses the variety of web sites that agencies might create or maintain.

Static Web Sites

This approach concentrates on managing the elements that constitute or are made available on a web site.

These strategies are well suited for sites that are made up of collections of HTML documents and do not rely upon complex interactivity with users of the site. The elements could be complete HTML documents that are stored and supplied to the user, or they could be the various elements that are assembled to create an HTML document when a user queries the site (e.g., headers, footers, agency logos, images and electronic documents).

This approach could entail taking periodic snapshots of collections of web resources in combination with tracking changes to the site and logging transaction details. Alternatively, elements or individual web resources could be separately captured and managed in association with metadata that describes the relationship between specified versions of the element and its unique Uniform Resource Identifier (URI) or Uniform Resource Locator (URL).

One method of capturing individual web elements is by using a content management system, which allows the web content to be created and managed in one repository that maintains the connections not only between the related elements but also the related metadata. (See *Content Management Systems* below for more detail.)

Dynamic Web Sites

These strategies focus on capturing the transactions that occur between the web site and the user. They are most suited to the situation where a dynamically generated site is database-driven and relies on stored user profiles, search mechanisms, SQL-HTML translation scripts, and other programs to enable full functionality.

Keeping records of a site that is dynamically generated would involve keeping track of changes to potentially enable its full reconstruction as it existed on any given date. This would require capture and storage of:

- user profiles;
- style sheets;
- search engine;
- scripts and programs;
- regular snapshots of the database itself; and
- database transaction logs.

Where the site is highly interactive or application-driven, it may be more feasible to capture "events" (such as single transactions between web site and user) rather than the elements that comprise the site at the time of the transaction (see below). An event-driven strategy would involve capturing:

- date and time of event;
- IP or domain address of user;
- user profile;
- query or other action performed; and
- the resource served to the user with relevant metadata attached.

Agencies need to remember that some of these elements, such as IP addresses or user profiles, may contain confidential information about the user. The *Kansas Open Records Act* outlines agency responsibilities with respect to the collection and security of personal information.

Capturing Metadata

Metadata can be described as "data about data". It provides descriptive data about a document, database, web site or other set of information. The application of metadata is central to the successful management of all records, including records of web-based activity. By using metadata and applying it consistently, an agency is able to describe the records of agency web activity and resources in an effective and efficient manner that facilitates future access to those activities and resources.

At this date, the State of Kansas has not adopted a metadata standard, but the Kansas Statewide Technical Architecture (KSTA) (see also the Kansas Electronic Records Management Guidelines) requires state agencies to maintain sufficient metadata to identify data for long-term retention. (The KSTA also provides in chapter 20 emerging and current metadata standards for recordkeeping and preservation, such as the Dublin Core Metadata Initiative, the Government Information Locator Service (GILS), and the Minnesota Recordkeeping Metadata Standard.) The metadata captured from agency web sites should be sufficient to meet the agency recordkeeping requirements, but should, at a minimum, include:

- record series title or unique identifier for the record;
- date and time of capture of the record into a system capable of meeting retention requirements;
- creator agency information;
- original data format;

- the use made of the record over time, including its placement on a web site;
- software and/or hardware dependencies to view the record;
- retention and disposal requirements for the record per agency or general records schedule; and
- any access restrictions associated with the record.

Snapshots

Taking a snapshot usually involves creating a full and accurate record copy of an agency's public web sites at a particular point in time. The snapshot should be captured into a system capable of meeting retention requirements and maintained over time for as long as the snapshot needs to be accessible, usually for the entire retention period of the records contained in the site.

When taking snapshots of collections of web resources, it is desirable to ensure (as far as possible) the continuing processability of the site and its component pages. This means that agencies should try to retain the capability to replicate the content, layout and functionality of the site across technological platforms without loss of data integrity.

This strategy is particularly useful for static web sites or collections of static objects that are essentially electronic publications. A snapshot should not be used to keep records of highly interactive dynamic sites or resources which are databases or transactional services. A drawback to this approach is that a snapshot only provides a picture of a site at a particular point in time. Snapshots may not capture all changes in content and structure.[6] If snapshots are captured in the absence of other records of web-based activity, it will be impossible to reconstruct the site together with its functionality at any other point in time. Since this method does not enable the agency to determine exactly when particular web resources were available, agencies that use this strategy should also create and maintain logs of changes made to sites between snapshots (see *Tracking Changes*, below).

For copies or snapshots of entire collections of web resources, metadata should include:

- date and time of capture;
- links to the universal resource indicator (URI) including information about version and date of link to specified URI;
- technical details about the web site design;
- details about the software used to create the web resources;
- details of the applications (including search engines) that supplement the web resources; and
- details about the client software needed for viewing the web resources.

Tracking Changes

Because government agencies are accountable for the information contained on their web sites, it is important to be able to show when changes occurred over time. This strategy involves tracking changes to the agency web site over time and creating a log that tracks those changes. The log is then captured into a system capable of meeting retention requirements and maintained for as long as needed to meet

---

[6] Anne R. Kenney et al., "Preservation Risk Management for Web Resources: Virtual Remote Control in Cornell's Project Prism," *D-Lib Magazine* 8, no. 1 (January 2002); available from http://www.dlib.org/dlib/january02/kenney/01kenney.html; accessed 19 January 2002.

retention requirements. Used in combination with snapshots of web sites, this method can be a reliable option for static sites.

The main problem occurring from this method is the capture of insufficient metadata about the site changes which result in the inability to interpret the log over time. It is important that metadata requirements are specified and sufficient metadata is captured into the same system as the records (see *Capturing Metadata*). However, it is suggested that at minimum the following data elements be captured in the change log:

- title or name of posting;
- version number;
- author or content manager responsible for creation;
- links embedded in the posting;
- date of initial posting;
- date of modification;
- date of replacement or withdrawal; and
- disposal information.

This is not a comprehensive list and agencies should review and adapt it to ensure their recordkeeping requirements are met.

In the case of a static web site, the log should capture changes to individual pages, documents or elements on the site. Changes to scripts, plug-ins, and forms used to present information, etc., will also need to be captured, as they will affect the functionality of the web-based records.

Managing Elements Separately

This strategy involves managing individual data elements with sufficient metadata to document how each element is associated with a particular URI/URL. The element-driven approach reduces the preservation burden by focusing on the preservation of the data elements and associated metadata instead of attempting to preserve entire systems that support web sites.

An agency should maintain a register or list of URIs/URLs that have been made available on its public web site and capture the data elements made available from these URIs/URLs. The relationship or association between the data elements and the URIs/URLs should be documented and maintained in a separate metadata store. The information about each URI/URL and data element that should be captured includes:

- the absolute URI/URL;
- the data element;
- the mime type of the data element;
- the start and end time of the association; and
- possible relationships to other records that document the administrative processes by which the resource was authored and published.

This information enables the agency to accurately track web resources at any point in time. Although it is possible to read the metadata and then view each element individually, this method of preservation does

not enable reconstruction of the web site because the data elements have been separated from other components of the site.

Content Management Systems

Content management systems can be used to maintain both the content of a web site as well as the code that was used to create the site. Content management systems allow content owners to control versioning of the information for which they are responsible. Because the system can control the changes in content over time, and allows for archiving of the information, this approach provides another option for the management of dynamic sites in particular. The different types of programming tools used to develop the site alongside the content, allowing web site owners to document how their sites both looked and behaved at a given point in time. While this approach might solve the issue of managing the site itself, remember that the records created by the activity on the site should also be maintained.

Web Transaction Logs

An agency web site that allows transactions to take place will both generate and be made possible by records. In the absence of a record, there is no evidence of the transaction having transpired. In the absence of legally sustainable evidence of a transaction having transpired, the transaction may be repudiated and/or deemed by a court of law to have not taken place. It is therefore essential that agencies capture full and accurate records of web-based transactions into systems that can guarantee the authenticity, reliability and accessibility of the records.

This strategy involves creating a log of site visitors, capturing the logs and any other records of transactions into a system capable of meeting retention requirements together with sufficient metadata, and maintaining them as long as required by an authorized retention and disposal schedule. The logs that capture the raw data from the hypertext transfer protocol (HTTP) header during a user visit to a web site ("Web Server Transactions Logs") are listed on the General Retention/Disposition Schedule, which requires that this data be retained for six months, although agencies may need to retain these logs longer if the records they track require a longer retention.

Although all web servers generate these log files, the raw data that is captured is often confusing and difficult to translate. It may be difficult to extract sufficient information to satisfy recordkeeping requirements. In such cases reliance on site logs may be considered too risky and alternative recordkeeping strategies will need to be pursued. These logs were designed to meet web site administration needs, not agency records requirements. It is therefore necessary to establish procedures to ensure that visitor logs satisfy recordkeeping requirements and are captured and maintained in appropriate systems.

A number of software packages are available to compile comprehensive reports in the form of tables and graphs. The elements that can be captured include:

- date and time;
- IP address or domain name;
- pages visited;
- actions performed;
- queries made; and
- web browser used.

Agencies will need to select the elements that fulfill their recordkeeping requirements.

Another issue to consider when capturing logs of web site transactions is privacy. Most logs of web site activity will contain personal information about the user. Kansas government agencies need to be aware of their responsibilities under the *Kansas Open Records Act* with respect to the collection and security of personal information.

Online Archive

An online archive is an attempt to capture all material posted to the active web site. The archive should capture both past and current postings to the site. To be used as a recordkeeping strategy the online archive should be linked to a system capable of meeting retention requirements with sufficient metadata. The main advantages of this option include facilitating maximum functionality of archived postings; providing greater accessibility to archived postings; providing a more appropriate option for complex sites; and, when used in conjunction with that system, enabling the reconstruction of the site at any point in time.

When an online archive is carefully planned and implemented, there are few deficiencies with this option. It relies on linking the system with a dedicated web server. The costs and other practical considerations associated with establishing and maintaining such a dedicated server will need to be balanced against the benefits of such a system. In particular, maintaining all past and current postings may require large amounts of storage space.

Security and Integrity of Archived Web-Based Records

Agencies need to ensure the authenticity of their archived records of web activity and resource. By protecting the authenticity of archived records, agencies can reduce their risk of accountability exposure and preserve the records for future access. The systems that create and maintain web sites must remain secure, and agencies should adhere to security standards established by the Kansas Technical Executive Council and the KSTA. These standards should also be applied to any web server or enterprise storage system that is used to maintain and preserve records of web activity, whether or not the server or system also houses the active web site.

Transfer of Electronic Records into State Archives Custody

The rapid pace of technological change, the diversity of hardware and software used by government agencies, and the State Archives' limited resources have made it essential for state government agencies to assume increased responsibility for the management and preservation of electronic records, including those on or generated via agency web sites. The State Archives is the official repository for state government records with enduring value. However, the principles outlined below are applicable to all levels of government in Kansas and to the entity that houses their records of enduring value.

As a general rule, the Kansas State Archives will accept physical custody of web sites only in cases where:

- the agency that created the records is about to or has become defunct and no agency is identifiable as its successor to the function, or
- the state archives enters into an agreement with an agency to take custody of the web site.

However, it is recognized there may be situations where certain temporary value web sites have to be taken into custody of the State Archives. Each proposed transfer of enduring or temporary value records will be considered on a case-by-case basis. Given the wide range of potential formats, volume, standard of preservation which could be involved; and that technology as well as formats and media are subject to constant and rapid change, it is not possible to draw up prescriptive rules governing standards for all proposed transfer of web sites. At the present time, the State Records Center is unable to store electronic media due to inadequate environmental controls at its facility.

As stated, a decision by the State Archives to accept custody of web sites will be on a case-by-case basis. The following matters will be taken into consideration for any transfer proposal. It is emphasized that the list is not comprehensive:

- The records have been appraised and have enduring value.
- The records have sufficient metadata and contextual information to meet the archives descriptive standards for web sites.
- It has been demonstrated that the administrative function which the records document has been discontinued by the Kansas legislature and
  - there is no successor agency for the function or activity;
  - there is no other agency or institution which could take custody of the records; or
  - the records have a security classification that would preclude them from being stored by another agency or institution.
- The resource impact is assessed and is manageable.
- The records proposed for transfer conform with media and formats that the state archives can support at the time.
- The transferring agency will meet all processing requirements (including duplication) necessary for the records to meet the state archives' standards for transfer and access.

The basic principle for management of electronic records in the State Archives custody is that the records must conform to standards and media that the State Archives can accommodate. This is so that the records can continue to accessed and preserved for the period required.

A transfer of custody may result in a modified or reduced functionality of the records (i.e. other than which they may have had when created). Similarly, the structure of the records may have to be modified to allow management of the web sites and access to them. Any such change will occur prior to transfer and will be specified in the agreement with the transferring agency. The transferring agency will meet all costs associated with any conversion action necessary to convert the records to an acceptable format and will provide any necessary new media or duplication required.

Appendix E – Maintaining Web-Based Records Over Time

It is important for agencies to manage the web-based records over time to meet the retention requirements for the records. The amount of time each record needs to remain accessible will depend on the requirements of the State Records Board approved retention and disposition schedule for the agency. This section outlines a number of storage and preservation issues that are likely to arise as a result of the need to maintain web-based records in an accessible form over time. Some of these issues remain unresolved and are the subject of industry research. It is crucial for agencies to be familiar with current preservation issues and best practices for web-based records; and as research becomes available, the Kansas State Historical Society will continue to provide guidance to agencies in this area.

Hardware and Software Dependency and Obsolescence

All web sites, regardless of their complexity, are dependent on particular pieces of hardware and software to enable full functionality. For example, a site may require a number of applications to function properly, such as Adobe Acrobat Reader, scripts, applets, media players, or a search engine. Often these applications are customized for a specific environment, such as a particular browser with plug-ins added, and will only run on a particular hardware configuration or operating system platform.

Computer technology is subject to ongoing technological obsolescence, with both software and hardware quickly becoming outdated as new upgrades and versions come onto the market. Agencies need to consider a number of interrelated software and hardware factors when maintaining web-based records, including:

- The evolutionary nature of standards for markup languages, such as different versions and types of HyperText Markup Language (HTML), and the increasing use of eXtensible Markup Language (XML).
- The proprietary, platform-specific nature of many search engines and database query tools;
- The embedding or linking of the correct versions of applications required for functionality, including applets, JavaScript, and software plug-ins.
- The limitations of some, particularly older, browsers, which cause different browsers to produce radically different or incomplete views of web pages.
- The estimated physical and/or commercial life of the medium on which web-based records and related descriptive metadata are stored.
- The long-term availability of the hardware and operating system platforms needed to access information stored on different types of media.

Agencies should plan for obsolescence by ensuring that records can be copied, reformatted or migrated. This includes hardware, software, operating system and media obsolescence. Section 8.3 below discusses reformatting, migration and conversion.

Standards

Web developers can decrease the need to migrate or convert web documents, media and applications by designing and implementing new programs and services with long term use and accessibility in mind. Designing in accordance with standardized best practices and using common, standardized file formats will be a much better long term investment in the initial creation of new content than using newer,

---

unproven and in some cases unstable technology. This is not to say that newer technology should or cannot be used to enhance a web page, simply not to represent vital data in a web application.

Markup Standards

It is important to ensure that documents are created with markup standards in mind. This practice will help to promote a document's usability with future versions of a standard and to help eliminate "code rot". Code rot is the phenomenon many application developers have experienced where older programming instructions will no longer function properly when migrated into a newer system or framework.

Markup language is the base of the Internet and predates the modern browser based applications that most users regard as "the Internet". Modern markup languages include HTML, XML, XHTML, and SGML. Due to its age and popularity, almost every markup language is very standardized and critically reviewed.

Unfortunately, from a records management perspective anyway, the user agents that are being used to display or represent modern markup languages are very flexible and can handle errors by the document author. These user agents are typically graphical web browsers, but also exist in the forms of such things as screen readers and text-based browsers. The fact that the newer user agents masks errors can promote poor coding practices and the use of proprietary markup that does not exist in formal standards because the published web documents will display as the author intended despite errors.

An example of this can be seen if a document authored in HTML 3, with text marked up with <p> tags at the beginning of each paragraph, but never closed. Almost any web browser would ignore this problem without error. If, at a future time, this text were moved into another document using <div> tags and stylesheets for positioning, the text may be out of position, not display at all, or even crash a standards compliant browser.

Multimedia, in this context, does not include content processed by some sort of interpreter such as a web browser plug in. Flash movies, Adobe Acrobat files and that sort of content will be addressed in the next section. Rather, multimedia content is considered to be image, audio and video formatted files that can either be displayed on a standalone basis, or interpreted by an application common to most modern operating systems.

Image Content Standards

Image formats are probably the most common format for representing information next to marked up text on the Internet. Web authors regularly use image formats such as JPEG, GIF, and TIFF to enhance a web document. It is important to test various image formats in target browsers to ensure that the layout, presentation, and coloring of the images is as intended across multiple formats when creating documents. The use of well thought out values of the **alt** attribute of the <img> tag will also help maintain an image's intent when it cannot be displayed.

Most image formats have been fairly stable for some time now. Nonetheless, if significant information is being represented in the image, or if the image is dynamically created, alternative long term storage of the data may need to be considered.

Audio and Video Content Standards

Audio and video formats are rapidly evolving and as such, critical information presented in audio or video should be accessible in other formats as well.

Standards for All Other Formats

Web content contained in a document that must be interpreted by a plug-in or application with works as a subset of the user agent will need to be handled with more care than the more common image, audio and video formats. Flash, Adobe Acrobat files, Java applets and VRML applications are good examples of this type of content.

Standards for Extracting Information

It is suggested that some means of extracting the information be stored and made available along with the rest of the web archive. An example may be to either store an application that can render Adobe Acrobat (PDF) along with the archive or a version of the information in a more common format, such as text. The archive, at best, would contain versions of the application for multiple platforms.

While the data in web documents is static, simply storing that document and ensuring that the document will be able to be interpreted at any given point in the future should be sufficient. If the data is generated at the time it is processed, however, additional steps may be in order to ensure that the information accessed from a web document can be retrieved.

Standards for Dynamic Web Environments

There are usually three key components, or layers, in a dynamic web environment.

The first layer is the presentation layer. This is the layer that controls the end result of the processing how the user agent presents information to the user. Typically, this will include HTML, images, and forms on a web page, along with embedded scripts that pull in the results of the middle layer.

The middle layer is programs or scripts that model the business logic of the application. These programs and scripts perform such duties as generating database queries, processing user input, error checking and business processing. This type of application will usually be executed on the web server at the time the user inputs information or submits a query. These scripts or programs are reliant on other programs that interact with the databases and web servers. These types of programs are collectively known as the application server for that program or script. Typical common web application servers are IIS, which can process ASP scripts, and Apache with mod-Perl or a PHP module which can process PHP or Perl scripts.

The final, or back end layer is a database or representation of data in some format, typically a relational database such as Oracle, MSSQL (Microsoft SQL server), or MySQL. The back end layer can be also anything from a text or XML file to a simple Microsoft Access database.

In order to properly store dynamic web applications, it will usually be necessary to consult a web developer. This should help identify which components of the web application will be needed to recreate transactions on that web application in the future. Storing vital components of all three layers will typically be necessary. Committing the presentation HTML pages, server-side scripts and an ANSI SQL dump of the vital parts of the database on a periodic basis are examples of this procedure.

It is worth noting that the best possible scenario for records preservation would be a complete representation of the database and user inputs to the web application. This may not always be possible in practice, however, due to either constantly modified or sensitive data. In these cases storing selected data, or perhaps transaction logs of the database may be considered instead of the preserving the entire database.

Most application servers and databases have achieved relative stability in their life cycles and most changes are either made with backwards compatibility in mind, or are well documented. Even so, as systems are upgraded over time, a plan to be either to roll back or recreate a past application server and database environment may be the simplest and most reliable strategy for continued web records reliability.

A listing of common formats in the web environment can be found below.

Reformatting, Conversion, and Migration

While adopting standards can delay an agency's need to respond to technological changes, preserving access to web-based records with long term value inevitably will require reformatting, conversion, and/or migration to new generations of hardware and software.

Reformatting electronic records typically involves transferring the underlying data from one storage medium to another, for example, transferring a web site snapshot from a CD-ROM to a DVD.[7] Whenever records are copied or reformatted, it is recommended that techniques such as "checksums" and "hash functions" be used to confirm the maintenance of data integrity.

Conversion involves automatically exporting or importing electronic records from one software application or version to another with no loss of record content and context and little or no loss of record structure.[8] In word processing, moving a document from Word Perfect to Microsoft Word would be an example of conversion.

Migration, the most challenging response to technological change, requires the creation of special purpose code to transfer electronic records to hardware or software.[9] A future web environment based on XML or some other markup language, for example, could require the migration of web-based records created with HTML 4.0 to the new platform.

Web-based records and their associated metadata should be reformatted, converted, and/or migrated as often as necessary to avoid technological obsolescence for as long as the records are required to remain accessible. All preservation actions should be noted in the recordkeeping metadata. Any loss of record content, context, or structure that occurs as a result of reformatting, conversion, or migration should be fully documented in the metadata.

Reliability and Authenticity

As defined above in Appendix A, a record is reliable if it engenders confidence that it is an accurate and trustworthy representation of a real event or activity. An authentic record has qualities that ensure that it remains reliable over time. Authenticity and reliability therefore are linked together in such a way that

---

[7] Charles M. Dollar, *Authentic Electronic Records: Strategies for Long-Term Access* (Chicago, IL: Cohasset Associates, Inc., 1999), 26-27

[8] Dollar, 29

[9] Dollar, 30-31

offers assurance on the "recordness" of the records in question. For researchers working with paper-based material, the presence of a document in a body of similar material that is known or believed to have been produced in accordance with standard procedures carries a presumption of reliability.[10] The same applies to electronic records, including those created, received or used in the web environment. Therefore, reliability corresponds to the business context in which it was created and used.

In order for agency records to remain reliable throughout their retention periods, the agency must ensure that the records are not altered over time. As Charles Dollar states, ""Over time" encompasses the circumstances of the initial transaction of the electronic record as well as any reproduction as a means of preservation. Ensuring the authenticity of electronic records over time, while maintaining their processibility or migrating them from legacy information systems, therefore, should be part of the core mandate of any organization responsible for providing access to electronic records."[11] Agencies must protect the records of web-based activity from the moment of creation, in the course of reference use, and through any possible migrations to new media or hardware and software platforms required to keep the records accessible until ultimate disposition either by destruction or transfer to the State Archives. Agencies must be able to demonstrate how they protect the authenticity of their web-based records, especially if called into account for the information contained in those records.

Persistent Identifiers

The World Wide Web has produced a revolution in the availability of electronic information for users of all types based on the distributed network nature of the Internet. The strengths of this distributed nature is that it leverages an architecture that is open and universally deployed. Its major weakness relates to the almost universal use of web identifiers that are based on the location of the digital resources they identify and which are therefore subject to change as resources move and systems change. One of the keys to maintaining web-based resources is the ability to rely on the links between the digital resources to remain static over time, not only those identifying the location of the original resource but any embedded links as well.

It is recommended that agencies adopt the practice of using persistent identifiers for their online resources. For as long as a given resource is available online it should have the same Uniform Resource Identifier (URI), Digital Object Identifier (DOI), PURL (Persistent URL) or online identifier. This means that users of online resources can cite or bookmark resources, confident that the resource identifier that they have quoted will not change.

Storage Media

Depending on recordkeeping requirements, agencies need to decide whether to capture and maintain web-based records on an offline or online storage medium. The size of the records is one of the determinants of the choice of storage media. Snapshots of sites and activity logs, for example, are likely to consume large amounts of storage space.

A second determinant is the desired speed of access. There is usually some delay in accessing records stored offline. Options for offline storage include optical disk or magnetic tape. In contrast, online storage provides instantaneous access in the form of a hard drive. However, instantaneous access is more expensive to maintain, especially if the agency is storing large quantities of data.

---

[10] Dollar, 24.
[11] Dollar, 25.

Online

Online storage makes it possible to back up or share important files for access through the web. The focus point of online storage should be to ensure that agencies can access their data anytime from any Internet-enabled location including being able to backup data off-site and collaborate with citizens and co-workers or any other possible entity.

Online storage provides easy and secure information access and may include:

- flexible 24/7 access from any computer with an Internet connection;
- scalable service to meet storage and access needs of agencies;
- automatic encryption and compression to protect information and increase efficiency;
- compatibility with firewalls and proxy servers to allow authorized users to access data from anywhere; and,
- automatic daily backup to enhance reliability.

Offline

Offline storage is contained in separate media devices. It provides a means of storing the information content of a Web site, and helps an agency manage recorded information through 1) controlled deletion or 2) preservation of archival records. It is the responsibility of the agency to ensure that information stored offline is readable for as long as the agency retention schedule requires.

Widely Used Storage Media

The most commonly used storage media on the market can be categorized into two types: magnetic and optical. Below is a brief description of these two types of storage media.

Magnetic Media

Magnetic media has developed along two paths: direct access media and sequential access media. Both types of magnetic media consist of a magnetic recording material and a binder that holds the recording material to a substrate or backing. An electromagnet, in the form of a read/write head, is then used to store or access the information on the media. Direct access media store data in available sectors and tracks without regard for physical location on the media, thereby placing "parts" of a file in available space on the media. A directory allocation table keeps track of the physical addresses of data on direct access media so that a specific set of data can be retrieved. On sequential access media one set of data follows another set, so an entire file is stored in the same physical location on the media. This requires the drive device to "look" at each set of data the proceeds the desired data.[12]

Direct access media includes both "hard" disks, usually encased within a computer CPU case, and "floppy" disks, which are manufactured to be portable. Hard disks are primarily used for rapid online storage, while floppy or removable disks have been used for short-term portable storage.

Sequential access media in the form of magnetic tape was introduced in the early 1950s with the first commercial computers. Over time tape technology evolved to provide better tensile strength and better oxide coatings to prevent flaking off of the tape base. Tape technology also allowed more compact tape

---

[12] Dollar, 164.

formats: open 10.5-inch reels of half-inch wide tape, storing 100 bits per inch, gave way to 4.5-inch square tape cartridges holding up to 10 gigabytes of data.[13]

Optical Media

There are three types of optical storage media available: read only memory (ROM), recordable (Write Once Read Many – WORM), and rewriteable (RW). These types of optical media share several features. The shared physical media characteristics include a rigid translucent polycarbonate substrate that contains thousands of tracks, a reflective coating over the substrate, and "pits" on the tracks that represent binary 1s and "lands" (space without pits) that represent 0s, and a protective clear lacquer or acrylic overcoat. These three types of media also share in common the use of a low power laser beam to read the 1s and 0s. ROM, WORM, and RW optical storage media can be differentiated by the way a binary stream of 1s and 0s is "recorded" and whether the recorded bit stream can be changed.[14]

Read Only Memory

There are only two read only technologies in use today. One is CD-ROM and the second is DVD that is "…an emerging storage technology with considerable future potential for the long-term storage of electronic records."[15] The two technologies are similar, in that they both use disks that have the same physical form, are fabricated the same way, and are read by a low power laser. Both technologies employ a high power laser to etch microscopic pits on a master glass disk, which is then used to create metal stamper disks that replicate the pits as microscopic bumps on their surface. Molten polycarbonate plastic is injected into the stamper disks to create plastic disks, which after cooling are covered with a very thin, highly reflected aluminum coating. The pits and lands on CD- and DVD-ROM disks are non-reversible so that they cannot be altered, making the disks "read only".

There are two major differences between CD and DVD technologies. DVD uses a laser with a shorter wavelength than CD-ROM technology, which means that a smaller pit is recorded. Smaller pits mean increased pit density per track and narrower tracks, resulting in a storage capacity for a single-sided DVD disk of 4.7 Gbytes (vs. 650 Mbytes for single-sided CD-ROM). The second major difference is that the data transfer for DVD disks is 1.4 Mbytes per second or approximately nine times for that of conventional CD-ROM. This is made possible by the high pit density per track and a greater rotational disk speed for the DVD. As of yet there is not an approved international standard for DVD. Due to the differences stated above CD-ROM and DVD disks are inherently incompatible. However, the differences can be overcome with a DVD drive with dual drive functionality.[16]

Write Once Read Many

Write Once Read Many (WORM) media have the same features of read only media except for the key difference of how the 1s and 0s are recorded on the media. In this technology a focused laser causes microscopic irreversible physical alterations on the surface of the recording material. Four different alterations are widely used: (1) ablative, (2) thermal-bubble, (3) bimetallic-alloy, and (4) organic dye. In ablative, the focused laser beam melts a pit in a thin film layer, which signifies a binary "1" and absence of a pit signifies a binary "0". The reverse occurs with a thermal-bubble alteration where the focused laser causes a bubble or blister to form, which signifies a "1". The third kind, bimetallic-alloy alternation, used the focused beam to fuse two different metallic layers with a known reflectance into a new alloy which has

---

[13] Dollar, 165-166.
[14] Dollar, 171.
[15] Dollar, 171.
[16] Dollar, 172.

a different reflectance. Spots where the new alloy has been formed are read as "1s" and places where no alternation has occurred are read as "0s". The fourth type of alteration, organic dye, exposes an opaque dye recording material to a high intensity laser. The laser heats a microscopic area to a temperature that changes it from an organic state to a translucent state.[17]

WORM media are available in four sizes: 4.72 inches, 5.25 inches, 12 inches, and 14 inches. The 4.72-inch (120 mm) disk is referred to as CD-R and can store approximately 650 Mbytes with a standard data transfer rate of 150 Kbytes per second. A very important feature of CD-R technology is adherence to International Standard (ISO) 9660 that covers both the physical layout of the disk as well as the format of the recorded information. An ISO 9660-compliant CD-R should be capable of being read by any CD-R or CD-ROM drive that also conforms to the same standard. As noted above, DVD-ROM and DVD-R do not yet conform to an international standard. However, "industry" standards do exist through the vendor-created DVD Forum, which is a non-proprietary standard.[18]

Rewriteable Optical Media

The essential characteristic of rewriteable media is that the surface deformations are reversible. CD-RW and DVD-RW allow users to erase previously recorded information and then to record new information onto the same physical location on the disk. This is accomplished by using "phase change" technology that allows a laser beam to change the media material from amorphous, where there is low reflectance, to crystalline in which there is high reflectance. The power of the focused laser beam is modulated between high and low power. When the beam is at high power the film forms an amorphous structure and when the beam is at low power it causes the amorphous structure to relax and return to a crystalline state.[19]

Environmental Controls and Monitoring

The storage environment for media is very critical to long-term preservation of whatever media is used to store web-based records. This might include:

- ensuring optimal temperature and humidity levels;
- protection against magnetic fields;
- using air filtration units to protect against air pollutants;
- prohibiting the consumption of food in storage areas; and
- planning for disaster preparedness.

As with other electronic records, procedures for regular backups of all web sites and associated repositories such as the local site repository should be undertaken. Best practices to be considered are:

- a "new generations" backup rule should be applied in which the seven day most recent backup storage media are kept at all times and the oldest backup copy is used for making a new backup copy; and
- backup media should be stored in a safe location away from the office – at minimum in another building.

---

[17] Dollar, 173.
[18] Dollar, 173.
[19] Dollar, 174-175.

Agencies should also consult their records officers and the system administrator for essential record requirements as a part of the agency disaster planning process.

Common Web Formats

Application

ZIP - A popular data compression format. Files that have been compressed with the ZIP format are called ZIP files and usually end with a .ZIP extension. A special kind of zipped file is a self-extracting file, which ends with an .EXE extension. You can unzip a self-extracting file by simply executing it

PDF – Short for *Portable Document Format*, a format developed by Adobe Systems for representing documents in a manner that is independent of the original application software, hardware, and operating system used to create those documents. A PDF file can describe documents containing any combination of text, graphics, and images in a device-independent and resolution independent format. The advantage of PDF over PostScript is the better accessibility of text within PDF files, which can still be searched for text and can also contain structural information like a table of contents. PDF files were originally intended to preserve printing formatting in an electronic format.

ps – *PostScript* is an interpreted, stack-based Programming Language. Its primary application is to describe the appearance of text, graphics, and images on printed or displayed pages. A program in PostScript can communicate a document description from a composition system to a printing system in a device-independent way. PostScript is an unusually powerful printer language because it is a full programming language, rather than a series of low-level escape sequences. The latest version of PostScript, version 3, fully integrates PDF.

Image

bmp - A native bitmap format of MS Windows and it is used to store (virtually) any type of bitmap data. Most applications running under MS Windows (MS DOS) and under other operating systems support read and write to BMP files

gif - Stands for *graphics interchange format*, CompuServe's standard for graphics image data exchange. There are currently two versions (GIF standards): GIF87a and GIF89a. The differences between two formats are minor

jpg - Short for *Joint Photographic Experts Group,* an Image Compression algorithm that is designed for compressing either full-color or grayscale digital images of natural, real-world scenes. It does not work very well on non-realistic images, such as cartoons or line drawings. JPEG does not handle compression of black and white (1 bit per pixel) images or moving pictures. JPEG itself does not describe an Image Format, it only specifies the compression algorithm. Some Image Formats for exchanging images compressed with the JPEG algorithm are the wide-spread JFIF and the less popular SPIFF.

tiff - Acronym for *tagged image file format*, a format that can characterize almost any form of 2D raster data using either ASCII or binary coding. "Private" tags may be used to allow additional parameters to be added to the descriptor. "Standard" TIFF allows the use of PackBits, LZW, G3 Fax, G4 Fax, and JPEG compression schemes within transmitted images. Four photometric classes are supported: TIFF-B for monochrome, TIFF-G for grayscale, TIFF-P for palette-based coding, and TIFF-R for RGB coding

png - Short for *Portable Network Graphics,* and pronounced *ping,* an extensible file format for the lossless, portable, well-compressed storage of raster images. PNG provides a patent-free replacement for GIF and can also replace many common uses of TIFF. Indexed-color, grayscale, and truecolor images are supported, plus an optional alpha channel for transparency. Sample depths range from 1 to 16 bits

Text

HTML – *HyperText Markup Language,* a simple markup language used to create hypertext documents that are platform independent. HTML documents are SGML documents with generic semantics that are appropriate for representing information from a wide range of domains. For example, HTML markup can represent Usenet News, Email, documentation, menus of options, database query results, simple structured documents with in-lined images, and hypertext views of existing bodies of information

XML - Short for *Extensible Markup Language,* XML is a markup language for structuring arbitrary data. XML was designed to replace HTML, which was deemed too restricted with its fixed set of elements and attributes. Because HTML is based on SGML, but SGML itself was considered as being too complex for direct application on the WWW, XML was defined as a functional subset (a "profile") of SGML. XML defines data types (called "schemas") with DTDs, which originate from the document-centered view of SGML. However, XML is very successful in B2B scenarios, and as such is increasingly used for data exchange (as opposed to document exchange). Because data description requires other features than document description (eg, built-in data types and type derivation), XML Schema has been defined as a replacement for DTDs, which makes XML more usable for B2B scenarios

SGML - Short for *Standard Generalized Markup Language,* a markup language for structured documents. Being the foundation for HTML, SGML today is the most frequently used language for structuring documents. The rules for how documents of a certain type may be structured are specified in a DTD, and every application of SGML (such as HTML) defines such a DTD. Even though SGML has been very successful, it is also rather complex and contains a lot of obscure features which are rarely used (and implemented). Thus, when a new language for replacing HTML on the WWW was needed, rather than directly taking SGML, a functional subset of SGML was defined, which has become known under the name of XML.

Streaming Media (Audio and Video)

aiff – *Audio Interchange File Format* is a proprietary Audio Format developed by Apple. The format can store monaural or multichannel sampled sounds in a range of sample rates and sample resolutions. Although originally AIFF did not support compressed audio data, a new version of the format called AIFF-C has been defined which allows compression

mp3 – A file extension for MPEG (*Moving Picture Experts Group),* audio layer 3. Layer 3 is one of three coding schemes (layer 1, layer 2 and layer 3) for the compression of audio signals. Layer 3 uses perceptual audio coding and psychoacoustic compression to remove all superfluous information (more specifically, the redundant and irrelevant parts of a sound signal. The stuff the human ear doesn't hear anyway).

WAV - A proprietary standard for audio files developed by Microsoft and IBM. The format can store monaural or multichannel sampled sounds in a range of sampling types, sample rates and sample resolutions. Support for WAV files was built into Windows 95 making it the de facto standard for sound on personal computers. WAV sound files end with a *.wav* extension and can be played by nearly all Windows applications that support sound.

midi – Short for *musical instrument digital interface*, a format which enables users to use multimedia computers and electronic musical instruments. There are actually three components to MIDI, the communications "Protocol", the Hardware Interface and a distribution format called "Standard MIDI Files". In the context of the WWW, the most interesting component is the Audio Format. In principle, MIDI files contain sequences of MIDI Protocol messages. However, when MIDI Protocol messages are stored in MIDI files, the events are also time-stamped for playback in the proper sequence. Music delivered by MIDI files is the most common use of MIDI today

mpeg - the Moving Picture Experts Group (MPEG) a working group of ISO/IEC in charge of the development of standards for coded representation of digital audio and video. Established in 1988, the group has produced MPEG-1, the standard on which such products as Video CD and MP3 are based, MPEG-2, the standard on which such products as Digital Television set top boxes and DVD are based, MPEG-4, the standard for multimedia for the fixed and mobile web and MPEG-7, the standard for description and search of audio and visual content. Work on the new standard MPEG-21 "Multimedia Framework" has started in June 2000. So far a Technical Report and two standards have been produced and three more parts of the standard are at different stages of development. Several Calls for Proposals have already been issued

QuickTime (mov) - Apple's architecture for handling multimedia data. First versions of QuickTime were basically a Format for audio and video, newer versions integrate streaming capabilities as well as more media types, including support for 3D and virtual reality. Being a proprietary technology, QuickTime can be compared to Microsoft's ASF

avi - Short for *Audio Video Interleave,* the file format for Microsoft's Video for Windows standard is used for storing audio and/or video information. It is a common format for audio and video files within PC environments. Being a proprietary technology, AVI can be functionally compared to Apple's QuickTime

asf - Short for *Advanced Streaming Format,* an extensible file format developed by Microsoft to store synchronized multimedia data. It supports data delivery over a wide variety of networks and protocols while still proving suitable for local playback. The explicit goal of ASF is to provide a basis for industry-wide multimedia interoperability. Each ASF file is composed of one or more media streams. The file header specifies the properties of the entire file, along with stream-specific properties. Multimedia data, stored after the file header, references a particular media stream number to indicate its type and purpose. The delivery and presentation of all media stream data is synchronized to a common time-line. ASF's functionality is similar to the one provided by Apple's QuickTime

Sources

http://wildesweb.com/glossary/
http://www.rasip.fer.hr/research/compress/algorithms/adv/gif/
http://www.rasip.fer.hr/research/compress/algorithms/adv/bmp/
http://www.rasip.fer.hr/research/compress/algorithms/adv/jpeg/
http://www.w3.org/Graphics/PNG/
http://mpeg.telecomitalialab.com/
http://www.webopedia.com/

Appendix F – Accountability Exposure Table

The following table, based loosely upon the research of Charles McClure and Timothy Sprehe, suggests the level of accountability exposure an agency is open to on its web sites. See Section 3, *Accessing Accountability*, for a discussion of this topic.

		Low Exposure	**Moderate Exposure**	**High Exposure**
Purpose	*Agency's website(s) is/are mainly used to:*	Provide access to materials available elsewhere or already captured in agency recordkeeping systems capable of meeting retention requirements.	Provide access to ephemeral "bulletin board" postings, official agency publications not captured elsewhere in recordkeeping systems, official hearings and other agency business created interactively in real time.	Transact business between users and agency.
Visibility	*Experience shows that agency publications:*	Are not controversial, generate no unfavorable press reaction.	Provide potential for adverse and controversial public interest.	Receive frequent press coverage and are watched closely by public interest groups.
	Experience shows that agency:	Has never been the subject of litigation	Has sometimes been the subject of litigation	Is frequently the subject of litigation
	Legal counsel advises:	Web sites demonstrate little to no legal risk.	Web sites demonstrate some capacity for legal liability.	Web sites carry substantial liability for agency, or sites not reviewed by counsel.
	Public desires the agency to:	Maintain of existing online information services.	Increasing electronic access to agency information holdings and expansion of web site offerings.	Provide e-commerce and e-government transactions.
Complexity	*Web site functionality*	Static materials only; no	Both static and	Interactive materials

	demonstrates:	interaction between users of web sites and a supporting application or database.	interactive real time materials; minimal user interaction required to display information generated from a supporting database.	only; information generated specifically for each user per individual request.
Change Management	*Frequency of changes:*	Not frequent (e.g., two months or more between changes)*.	Moderately frequent (e.g., once a month).	Very frequent (e.g., daily or weekly).
	Scheduling of changes:	Automated updates.	Regularly scheduled.	Irregular, not scheduled at all or on an "as needed" basis.
	Ability to make changes:	Is controlled through a centralized agency team or individual.	Is available to separate offices identified by the agency.	Is not restricted or managed by the agency.
	Quality control demonstrates:	Information on web site(s) is accurate and consistent.	Several web sites overlap in content and may present conflicting information.	Information on web site(s) is inaccurate and inconsistent.
		* Agencies should be aware that while infrequent changes in an agency web site might indicate low accountability exposure, no changes over time to a web site will more than likely expose an agency to high risk of being called into account for out-of-date or superceded information on the site.		
		Low Exposure	**Moderate Exposure**	**High Exposure**

Self-Assessment Tool for Determining Accountability Exposure

To assist agencies in determining their level of accountability exposure for a web site or set of web sites, the following tool has been developed. The tool is presented in two formats (narrative and table) to better facilitate the evaluation of agency web sites. Both formats present the same information from the Accountability Exposure Table and should provide the same feedback on the level of exposure on agency web sites.

PLEASE NOTE: This tool is intended to serve only as a potential indicator of the level of exposure. Agencies should still employ their best judgment in determining what actions should be taken to ensure that records and other information maintained on their websites are accurate, consistent, and complete.

Narrative Format

Answer all of the questions below, understanding that "a" corresponds to "low exposure", "b" corresponds to "moderate exposure" and "c" corresponds to "high exposure".

Purpose

_____ Your agency's web site(s) is (are) used to (may choose more than one):
a) Provide access to materials available elsewhere or elsewhere already captured in agency recordkeeping systems
b) Provide access to ephemeral "bulletin board" postings, official agency publications not captured elsewhere in recordkeeping systems, official hearings and other agency business created interactively in real time
c) Transact business between users and agency

Visibility

_____ In your experience, agency publications:
a) Are not controversial, generate no unfavorable press reaction
b) Provide potential for adverse and controversial public interest
c) Receive frequent press coverage and are watched closely by public interest groups

_____ In your experience, the agency:
a) Has never been the subject of litigation
b) Has sometimes been the subject of litigation
c) Is frequently the subject of litigation

_____ Review of agency web site(s) by legal counsel indicate that:
a) Web site(s) demonstrate little to no legal risk
b) Web site(s) demonstrate some capacity for legal liability
c) Web site(s) carry substantial liability for agency or not reviewed by legal counsel

_____ The public indicates a desire for the agency to:
a) Maintain existing online information services
b) Increase electronic access to agency information holdings and expansion of web site

offerings
c) Provide e-commerce and e-government transactions

Complexity
Web site functionality demonstrates:
a) Static materials only; no interaction between users of web site(s) and a supporting application or database.
b) Both static and interactive real time materials; minimal user interaction required to display information generated from a supporting database
c) Interactive materials only; information generated specifically for each user per individual request

Change Management
Please indicate the frequency with which updates are made to your web site(s):
a) Not frequent (two months or more between changes)
b) Moderately frequent (once a month)
c) Very frequent (daily or weekly)

Please indicate which of the following best describes the method with which site changes are scheduled:
a) Automated updates
b) Regularly scheduled
c) Irregular, not scheduled at all or on an "as needed" basis

Ability to make changes to web site(s):
a) Is controlled through a centralized agency team or individual
b) Is available to separate offices identified by the agency
c) Is not restricted or managed by the agency

A review of your web site(s) for quality demonstrates:
a) Information on web site(s) is accurate and consistent
b) Several web sites overlap in content and may present conflicting information
c) Information on web site(s) is inaccurate and inconsistent

Table Format

Space is provided to check the appropriate column for low, moderate and high exposure for each question. This form will provide a picture of where the agency stands regarding the exposure posed by its web site(s).

		Low Exposure	**Moderate Exposure**	**High Exposure**
Purpose	*Agency's website(s) is/are mainly used to:*	Provide access to materials available elsewhere or already captured in agency recordkeeping systems.	Provide access to ephemeral "bulletin board" postings, official agency publications not captured elsewhere in recordkeeping systems, official hearings and other agency business created interactively in real time.	Transact business between users and agency.
		☐	☐	☐
Visibility	*Experience shows that agency publications:*	Are not controversial, generate no unfavorable press reaction.	Provide potential for adverse and controversial public interest.	Receive frequent press coverage and are watched closely by public interest groups.
		☐	☐	☐
	Experience shows that agency:	Has never been the subject of litigation	Has sometimes been the subject of litigation	Is frequently the subject of litigation
		☐	☐	☐
	Legal counsel advises:	Web sites demonstrate little to no legal risk.	Web sites demonstrate some capacity for legal liability.	Web sites carry substantial liability for agency, or sites not reviewed by counsel.
		☐	☐	☐

	Public desires the agency to:	Maintain of existing online information services.	Increasing electronic access to agency information holdings and expansion of web site offerings.	Provide e-commerce and e-government transactions.
		☐	☐	☐
Complexity	*Web site functionality demonstrates:*	Static materials only; no interaction between users of web sites and a supporting application or database.	Both static and interactive real time materials; minimal user interaction required to display information generated from a supporting database.	Interactive materials only; information generated specifically for each user per individual request.
		☐	☐	☐
Change Management	*Frequency of changes:*	Not frequent (e.g., two months or more between changes)*.	Moderately frequent (e.g., once a month).	Very frequent (e.g., daily or weekly).
		☐	☐	☐
	Scheduling of changes:	Automated updates.	Regularly scheduled.	Irregular, not scheduled at all or on an "as needed" basis.
		☐	☐	☐
	Ability to make changes:	Is controlled through a centralized agency team or individual.	Is available to separate offices identified by the agency.	Is not restricted or managed by the agency.
		☐	☐	☐
	Quality control demonstrates:	Information on web site(s) is accurate and consistent.	Several web sites overlap in content and may present conflicting information.	Information on web site(s) is inaccurate and inconsistent.
		☐	☐	☐

		* Agencies should be aware that while infrequent changes in an agency web site might indicate low accountability exposure, no changes over time to a web site will more than likely expose an agency to high risk of being called into account for out-of-date or superceded information on the site.		
		Low Exposure	**Moderate Exposure**	**High Exposure**