Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government

NATIONAL ARCHIVES
OF AUSTRALIA

CONTENTS

ABBREVIATIONS

AGLS	Australian Government Locator Service
CD-R	compact disk – recordable
CD-ROM	compact disk – read only memory
HTML	hypertext mark-up language
IP	Internet protocol
ISO	International Standards Organisation
IT	information technology
OH&S	occupational health and safety
PURL	persistent universal resource locator
RDA	records disposal authority
URI	universal resource indicator
URL	universal resource locator
URN	universal resource name
VPN	virtual private network
W3C	World Wide Web Consortium
WebDAV	web distributed authoring and versioning
WORM	write once read many
XML	extensible mark-up language

1.   INTRODUCTION

1.1 Purpose

This document provides Commonwealth agencies with a set of guidelines for keeping records of web-based activity. The purpose of the guidelines is to help Commonwealth agencies determine and implement appropriate strategies for creating, capturing, managing and retaining records of web-based activity for as long as they are required.

Directions for making and keeping records of online resources are available in a companion document, Archiving Web Resources: A Policy for Keeping Records of Web-based Activity in the Commonwealth Government. While the policy document gives agencies directions on *what* they are required to do, these guidelines provide practical advice on *how* to comply with the directions. The guidelines do not advocate a one-strategy-fits-all approach. Rather, they outline a range of possible strategies, suggest combinations of strategies that might be appropriate for different sets of circumstances, and discuss the considerations that may lead to the adoption of particular combinations of strategies.

The online resources covered by these guidelines are:

- public websites;
- virtual private networks (VPNs);
- extranets;
- intranets;
- individual web-based documents and publications; and
- records of web-enabled activity, including electronic commerce.[1]

1.2 Scope

These guidelines are applicable to all Commonwealth agencies which are subject to the *Archives Act 1983* and which:

- maintain a public website, intranet, extranet or virtual private network;
- use the Internet to provide services and products to the public; and/or
- conduct business with other organisations and/or individuals over the Internet.

The challenge of archiving web resources

With government agencies conducting a significant and rapidly increasing proportion of their business over the Internet, it is vital that they set in place policies, procedures and systems which ensure that full and accurate records of web-based activity are created and retained. This need has been recognised in the Government Online Strategy, which identifies the National Archives of Australia's recordkeeping guidelines as a key enabler of the strategy. Agencies are required to

---

[1] Detailed guidelines for keeping the records resulting from electronic commerce (e-commerce), including the use of Public Key Technology, are not provided in this document. A National Archives policy and guidelines for keeping e-commerce records are due for release in 2001.

follow the principles outlined in these guidelines if they are to be successful in delivering appropriate services over the Internet.

The creation and maintenance of authentic, reliable, accurate and durable evidence of web-based activity is essential if agencies are to retain corporate memory and meet legal obligations and community expectations. Good web-based recordkeeping is also essential if agencies are to maximise both the efficiency of online operations and the return on investment in information assets. From a broader perspective, online resources constitute a vital component of the documentary record of Australia. If future generations of Australians are to be able to understand life in Australia and the role of government in the early 21st century, they will need to have access to preserved copies of significant government online resources and records of government web-based activity. The National Archives of Australia has a responsibility to work with government agencies to help ensure that these legal, operational and historical imperatives are met.

1.4   Responsibilities for preserving online material

Each Commonwealth agency has primary responsibility for creating and maintaining full and accurate records of its activities. Under s24 of the *Archives Act 1983*, Commonwealth agencies must retain or dispose of Commonwealth records in accordance with instructions issued by the National Archives. Consequently, agencies have an 'up-front' preservation responsibility to capture and maintain records, including web-based records, in accordance with authorised disposal authorities, such as the *Administrative Functions Disposal Authority* or specific Records Disposal Authorities.[2]

The National Archives of Australia and the National Library of Australia, in partnership with the Info Products Branch of the Department of Finance and Administration, have complementary roles in preserving the documentary record of the Commonwealth Government. In the online environment the responsibilities of the National Library and the National Archives overlap, and this has implications for Commonwealth agencies and their recordkeeping responsibilities. *Archiving Web Resources: A Policy for Keeping Records of Web-based Activity in the Commonwealth Government* outlines the relationships among these organisations and their specific preservation responsibilities.

The National Library of Australia, for example, selectively collects web-based publications from various Australian websites, including Commonwealth agency websites, as part of its PANDORA project. However, the collection policy for PANDORA is selective and it does not supersede an agency's responsibility for ensuring the retention of copies of their web-based publications in a recordkeeping system.

---

[2] Further guidance on determining and satisfying recordkeeping requirements can be obtained from *Designing and Implementing Recordkeeping Systems: Manual for Commonwealth Agencies*, developed by the National Archives of Australia in collaboration with the State Records New South Wales.

1.5 Key definitions

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.

Gateway site

A public website which comprises links to webpages and other websites around a common theme. Also known as a 'metasite', 'portal' or 'umbrella site'.

Intranet

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

Public website

A collection of electronic files, usually under common administrative control, linked together and made accessible to the public via the World Wide Web.

Publication

Something which is published; or

the function of having works, irrespective of format, issued for sale or distribution internally or to the public.[3]

Records

Recorded information, in any form, including data in computer systems, created or received and maintained by an organisation or person in the transactions of business or the conduct of affairs and kept as evidence of such activity.[4]

Snapshots

Full and accurate record copies of an agency's public web resources captured at particular points in time. For records retention and disposal purposes, it is considered that the equivalent of snapshots can be delivered by a recordkeeping system that captures and retains full and accurate records of the data objects made available via an agency website, together with adequate metadata documenting the management history and context of those objects. Entries 1933 to 1937 of the Administrative Functions Disposal Authority, issued February 2000, provide advice on disposal of snapshots of an agency's public website.

Virtual private network (VPN)

A private data network that makes use of the public telecommunication infrastructure (ie the Internet). Privacy is maintained by employing secure protocols and security procedures such as data encryption. Agencies can choose to use VPNs as extranets or wide-area intranets. For example, FedLink is a

---

[3] National Archives of Australia, *Administrative Functions Disposal Authority*, Canberra, February 2000.

[4] Australian Standard AS 4390–1996, *Records Management*.

Commonwealth VPN which will be used as a wide-area intranet, linking all Commonwealth departments and agencies together in a secure online environment.

Website

In this document, a generic term which refers to all types of web-based sites, including public websites, VPNs, extranets and intranets.

2.   THE DIVERSITY OF WEB-BASED RESOURCES

Websites today come in many different forms. These range from simple collections of static pages that display the same information to all visitors, through to pages which are created and displayed dynamically in response to specific queries. In addition, many websites now *do* something – they enable visitor details to be captured, online orders to be taken and personalised information to be displayed based on user profiles. Thus, some websites can be considered to be more like a software applications than a set of publications.

Each form of web-based activity poses a different set of challenges for those who seek to keep and preserve records of them over time. Sections 5, 6, and 7 of these guidelines outline strategies and advice for dealing with some of these challenges.

2.1   Static websites and web resources

In its most basic form, a website may be nothing more than a collection of static documents sitting in folders 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 'naa.gov.au'. A static website maps URLs directly to file system locations. The only interactivity provided by static sites is in the links which enable movement from one document to another or from one part of the site to another.

Static sites are relatively simple to archive, for example by making and keeping:

- copies or snapshots of the site 'as a whole' (ie the functionality of the internal links is not destroyed); and

- logs of changes (eg when and how documents are removed, replaced or updated).

2.2   Static websites and web resources with form-based interactivity

Many websites utilise forms to collect information such as comments and requests, from visitors. While these sites are still largely static publication mechanisms, agencies keeping records of such sites must also pay attention to:

- the information provided when the visitor fills in the form (usually stored in a 'back end' information system);

- the form itself; and

- the human readable source code of the script or program which enables the form's functionality.

2.3   Websites and web resources based on dynamic data access

Websites are sometimes used as front ends, or user interfaces, for accessing an organisation's database(s). Site users search prepared lists or put together their own searches which, in turn, query the content of a database. The information returned from these queries is displayed as an HTML (hypertext mark-up language) document to the user.

In many cases, documents exist as objects in a database. Each document will have its own unique identifier, usually reflected in the URL. This means that a user can bookmark the particular document and return to it later without reconstructing

the original search query (provided the document has not been deleted from the database).

Even if the site's main or top-level pages are static, dynamic data access websites raise some additional issues for agency recordkeepers.

- Not all users 'see' the same website. At designated levels, the pages displayed on users' browsers are based on what they ask for, therefore user queries are an integral part of generating the website and may need to be captured.

- Information contained in databases behind the site may be continually changing.

2.4 Dynamically generated websites and web resources

An increasing number of websites are being built which generate all of the pages 'on the fly'. This means that the component parts of each individual page – its content, structure and presentation – are generated dynamically using a combination of databases and style sheets based on:

- a stored set of user preferences;

- a stored set of access profiles;

- a user query; and/or

- the capabilities of the user's browser.

In these situations, the website does not exist in any single or easily capturable form. Each user sees a different 'site' based on their stored preferences and access rights, current needs, and the capabilities or limitations of the technology they are using.

Although the end result for the user might be a set of static pages, the processes which build the pages involve the use of a number of software tools. This is the point at which websites become more like software applications than electronic publications. Agencies need to consider how to archive dynamically generated web resources in a fully functional state.

The major issues these sites raise for agency recordkeepers is the need to choose whether to use an object-based or an event-based approach to keeping records of web resources and activities. That is, an agency needs to determine whether it wishes to focus on keeping records of:

- the individual transactions between clients (users) and servers (agencies); or

- the objects that comprise the content of the site at any given time.

3.   FUNDAMENTALS OF GOOD WEB-BASED RECORDKEEPING

In keeping records of web-based activity, there are certain fundamental procedures that all agencies should observe. These rules are not unique to web-based recordkeeping. They are commonsense approaches which organisations, as a matter of course, should implement as part of their regimes for managing information resources – whether these are web resources, electronic or paper-based records, or data in legacy systems.

3.1   Take a systematic approach

Effective web-based recordkeeping relies on persuing a systematic approach that is generally applicable to all records, regardless of format. The approach advocated by Australian Standard AS 4390–1996, *Records Management* involves:

- taking responsibility for recordkeeping;
- identifying the functions and the environment in which the organisation operates;
- analysing the need for records relating to those functions; and
- building systems and managing records according to those needs.

In line with this approach, the following strategies are applicable to all electronic records, regardless of format.

Promulgate agency policy on making and keeping records of web-based activity

The companion policy document to these guidelines provides a policy framework that Commonwealth agencies should adapt and expand to suit their unique environments and circumstances. With the support and endorsement of senior management, Commonwealth agencies should develop and promulgate agency-specific policy on making and keeping records of web-based activity.

Formulate a plan for capturing and maintaining these electronic records

A strategic approach to recordkeeping involves careful planning over a range of issues. Australian Standard AS 4390–1996, *Records Management* outlines the need for organisations to develop strategies that ensure 'full and accurate records' are made and captured into recordkeeping systems and maintained for as long as required. In March 2000 the National Archives, in collaboration with the State Records Authority of New South Wales, released [Designing and Implementing Recordkeeping Systems: Manual for Commonwealth Agencies (DIRKS Manual)](). This manual provides practical guidance on the design and implementation of best practice recordkeeping systems. The methodology involves identifying recordkeeping requirements, assessing existing systems to determine if they meet those requirements and designing new recordkeeping systems to meet identified gaps. Issues such as capturing and maintaining electronic records are addressed in this manual.

Formulate an electronic records preservation plan

Agencies should ensure that storage media and related technologies used for maintaining electronic records over time are designed so that the evidential elements of electronic records are accessible but cannot be altered. This includes

considering issues such as implementing security controls, data administration measures, audit trails and adopting media and open standards that meet longevity and migration requirements. These issues are fully considered in Section 6, Storage and preservation issues.

Implement specific website maintenance procedures

The final stage of a systematic approach is implementation. Agencies should take the necessary action to ensure procedures are in place for a successful implementation process and monitor compliance with strategies and procedures that have been established.

3.2   Assign and document responsibilities

Three or four different groups of agency staff may have responsibility for making and keeping records of web-based activity and resources. They are:

- content authors;

- website administrators;

- recordkeeping practitioners; and

- information technology staff, such as network managers or data administrators.

Although the spread of responsibilities may vary from agency to agency, the important point is that responsibilities need to be assigned to individuals or positions, and documented. If an agency has a high public profile and is particularly open to public scrutiny, it is liable – and may be called to account for – the material on its public website. A major component of an agency's internal accountability process should be assigning and documenting responsibilities for web-based recordkeeping.

One example of the type of responsibility that must be assigned and properly documented is the capturing of individual records of web-based activity into a formal agency recordkeeping system. It would make sense to assign this responsibility to either the content author (or responsible section), or to agency publications staff or to recordkeeping staff, rather than to the website administrator (who manages the website itself). However, an agency might choose to extend some recordkeeping responsibility to the website administrator.

In this scenario, a procedure might be written which requires the website administrator to inform responsible staff when material has been posted to or removed from the website. This would help to ensure that the relevant administrative metadata (eg management and use history) is appended or linked to the original record in the recordkeeping system.

The identification of responsibilities to be documented and assigned is explored further in Section 5, Strategic and technological options. Some of the main responsibilities that need to be assigned include responsibility for:

- identifying recordkeeeping requirements for web-based activity;

- determining whether existing systems can satisfy these requirements or whether it is necessary to design and implement new systems or improve existing recordkeeping systems;

- establishing policies and procedures governing the control and management of the agency's website;

- carrying out a risk assessment to determine the level of acceptable risk posed by an online website presence;

- raising the profile and general awareness within the organisation of the general recordkeeping responsibilities of all staff, especially the need to keep records of web-based activities, and providing necessary training;

- selecting and implementing an appropriate strategy or combination of strategies to ensure recordkeeping requirements are satisfied (eg establishing clearly when records need to be created and captured and how long they need to be retained);

- documenting procedures and processes to ensure strategies are carried out;

- selecting appropriate storage media and ensuring procedures and processes for long-term preservation are instituted; and

- setting-up data management procedures to ensure the integrity of the system and the authenticity records.

3.3   Determine requirements for records

Before determining specific strategies and actions for keeping records of web-based activity, agencies need to know which records they need to create, and how long those records should be retained. The National Archives of Australia has developed a number of tools to assist agencies to ascertain recordkeeping requirements and to determine the need to create, capture, retain or dispose of records. These include the DIRKS Manual and the Appraisal Guidelines for Commonwealth Records.

Agencies must decide what records of their web-based activities are needed to support operational needs and to satisfy broader organisational accountability requirements and community expectations. An agency's decision to keep or not keep records of its web-based activity must, as with all records, be made on the basis of an assessment of business risks, costs and benefits. In particular, agencies must assess the business risks of not enabling full accountability for their actions.

Agencies also need to be aware of their responsibilities under the *Privacy Act 1988* with regard to capturing and maintaining records containing personal information. These responsibilities extend to keeping records of web-based activity.

Web publications

Policy Statement 3.2 of *Archiving Web Resources: A Policy for Keeping Records of Web-based Activity in the Commonwealth Government* notes that some web resources may be publications. Web resources that are primarily designed to publish and disseminate information, and are relatively static informational resources, are to be treated as publications. Requirements for records that document web resources as publications or web resources that contain publications are identified in the National Archives of Australia's *Administrative Functions Disposal Authority* (*AFDA*), issued February 2000.

The PUBLICATION/Production functional activity set of the *AFDA* identifies the following recordkeeping requirements.

Entry	Description of Records	Disposal Action
1933	Publications produced only in an electronic format on an agency's public website	Retain as national archives (1) capture in agency's electronic recordkeeping system, and (2) transfer to National Library's Pandora project.
1935	Snapshot of agency's public website taken in accordance with National Archives policy on archiving websites	Retain as national archives
1936	Snapshot of agency's intranet site, extranet sites and websites hosted on behalf of other organisations taken in accordance with National Archives policy on archiving websites	Destroy when references ceases
1937	Agency publications other than those paper publications lodged with the National Library of Australia (NLA) under legal deposit or electronic publications lodged with NLA's Pandora project eg procedures, manuals and circulars	Destroy according to disposal action under a relevant function in *AFDA* or an agency's RDA

It important to note, from the above entries, that *AFDA* states that agencies must retain, as national archives:

- publications available *only* on public agency websites; and

- records of *public* websites, such as site snapshots.

Where a website publication *exactly* replicates a printed publication, it is sufficient to maintain the printed version of the publication as the record that is captured into the agency's recordkeeping system. If, however, the website version contains significant web-based functionality that is not adequately replicated in the printed version, then web-based records of the publication must be captured into the agency's recordkeeping system.

Decisions regarding how often an agency needs to take snapshots of public web resources will vary from agency to agency. These decisions should be based on an analysis of the risks faced by the agency in relation to its web-based activities (see Section 4, Assessing risk). The PUBLICATION/Risk Management functional activity set in *AFDA* deals with records arising from risk management activities.

Entry	Description of Records	Disposal Action
1947	Records documenting risk management relating to the publication function (eg risk analysis to determine timing for a snapshot to be taken for the agency's web/intranet site). Includes records covering each stage of the process.	Destroy 7 years after last snapshot of website was taken.

Other web resources

Web resources which go beyond being a publication, include those that are highly interactive, or serve as the front end of organisational databases and provide a unique response to queries and requests. In these circumstances, the records that should be captured are the requests made and the unique resources that are delivered in response to the queries. This event-based strategy is further explored in Section 5, Strategic and technological options. These records are subject to functional appraisal in line with all other core business records of the organisation.

Under entry 1937, the *AFDA* states that an individual record on a website must be captured and retained in accordance with the relevant disposal authority. For example, an agency may create an internal occupational health and safety (OH&S) policy and set of procedures. These records must be captured and retained in accordance with the disposal authority which covers the OH&S function, *regardless of whether or not they are placed on the agency intranet.*

3.4   Apply metadata

The application of metadata is central to the successful management of all records, including records of web-based activity. Metadata needs to be applied:

- at different levels of resource aggregation;
- at different times; and
- for different purposes.

The Government Online Strategy requires all Commonwealth agencies to describe their web-based resources – including websites, individual webpages and web-based services – using metadata that is compliant with the Australian Government Locator Service (AGLS) metadata standard for online resource discovery. Consistent application of AGLS metadata to online resources across the Commonwealth will assist members of the public to locate and access the Government information and services they need.

The AGLS User Manual and detailed guidelines for the application of the standard are available from the National Archives website. National Archives staff can, upon request, conduct a quality assessment of the AGLS metadata applied to the web resources of an agency. Interested agencies should contact the Archives at agls@naa.gov.au.

Records of web resources and activity – including site snapshots – should be further described using recordkeeping metadata as recommended in the National Archives' Recordkeeping Metadata Standard for Commonwealth Agencies.

While compatible with the AGLS metadata set, this standard specifies additional metadata that agencies should use to describe the context, management, use, preservation and disposal of information resources – including online resources – which are also records.

For individual records on websites and for other records of web-based activity, this means using metadata to describe:

- date and time of creation and registration of the record into a recordkeeping system;
- organisational context;

- original data format;

- the use made of the record over time, including its placement on a website;

- mandates governing the creation, retention and disposal of the records; and

- management history of the record following creation – including sentencing, preservation and disposal.

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

- data and time of capture;

- links to the universal resource indicator (URI) including information about version and date of link to specified URI;[5]

- technical details about the website 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.

3.5 Capture records into a recordkeeping system

The capture of record of web-based activity, together with sufficient recordkeeping metadata should occur at the time the resource is posted to the website and the record created. To achieve this, agencies will need to have clearly defined procedures, appropriately assigned responsibilities and adequate systems to support this requirement.

Staff awareness is the first step – agencies should take proactive steps to raise the general awareness of the need to create and capture records of web-based activity.

Assessing recordkeeping needs is also crucial. Organisations need to assess the complexity of the records and recordkeeping requirements to ensure 'full and accurate records' are captured and maintained. Agencies should assess the ability of existing recordkeeping systems to capture and maintain records of web-based activity and to satisfy online recordkeeping requirements. The DIRKS Manual provides the methodology to enable this assessment and practical guidance in designing new systems if existing systems fail to meet recordkeeping requirements.

The extent of maintenance required to preserve the functionality of electronic records will also need to be determined by this assessment. A relatively static website comprising simple documents with low interactivity will have different requirements for maintenance than a complex web-based document or highly interactive website. Each circumstance may require a different strategy to ensure record functionality can be maintained.

---

[5] It is worth distinguishing between three concepts:
1. Universal resource indicator (URI) – a general purpose namespace mechanism
2. Universal resource locator (URL) – an instance of a URI that is the address of some resource, accessible by means of a protocol such as HTTP
3. Universal resource name (URN) – an instance of a URI that, unlike a fragile URL, is guaranteed to remain available (Jon Udell, *Practical Internet Groupware*, O'Reilly, Sebastopol, CA, 1999, p. 471).

Some websites will offer significant functionality requiring additional web-based records to be created and captured (ie a multimedia document may require the capture of plug-ins to maintain its functionality). A number of technological options are available to ensure that additional functionality (and the generated website records) are captured into the agency's recordkeeping system(s) (refer to Section 5, Strategic and technological options).

Many agencies currently operate in a mixed environment, dealing with both paper-based and electronic records. The capacity to capture and manage web-based electronic records in their native formats during this transitional period may be limited if the appropriate infrastructure does not exist. In such cases, suitable hybrid transitional strategies will need to be developed and implemented.[6]

---

[6] For example, a transitional strategy might consist of taking snapshots of collections of web resources and creating and maintaining activity logs and site visitors logs. Capturing these records into traditional paper-based or electronic recordkeeping systems will involve consideration of issues such as the storage capacity of systems and the ability of the recordkeeping systems to absorb large formats. Records capture may take the form of ensuring adequate links are created and maintained between the agency's recordkeeping system and web-based records that are stored either online or offline.

4.   ASSESSING RISK

[Archiving Web Resources: A Policy for Keeping Records of Web-based Activity in the Commonwealth Government](#) states that 'any Commonwealth agency that administers a website must assess the business risk it faces in maintaining an online presence…' The policy does not explicitly prescribe the extent of recordkeeping required, because the results of a risk assessment will guide agencies' selection of recordkeeping strategies. Where an agency faces high levels of risk exposure, comprehensive records will need to be captured and maintained, possibly over long periods of time.[7]

Further guidance on the assessment of risk in the broader context of recordkeeping in an organisation is explored in the [DIRKS Manual](#). In particular, readers are referred to Step C, Identification of recordkeeping requirements, and the section on Recordkeeping risk analysis in Part Three. In addition, Australian Standard AS 4360–1999, *Risk Management* provides general guidance on how to undertake risk assessment.

4.1 Factors to consider

What major factors should an agency consider when determining a course of action for keeping records of its web resources? Four main factors should guide the assessment of the level of risk faced by an agency.

1. 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, consider Commonwealth agencies which provide assistance and services to members of the public, and those which are involved in current high-profile operations or investigations. These agencies are subject to much higher levels of public scrutiny (and criticism) than those which work behind the scenes conducting, for example, government analysis and research.

The higher public visibility of the first group of agencies means that they are at greater risk of being held accountable for the content of their public websites than the second group. This means that an agency should assess its current level of public visibility – categorising it as either high, medium or low – and take the results of this assessment into account when:

- assessing the level of risk associated with its web-based activity; and

- determining an appropriate course of action for keeping records of its web-based activity.

Circumstances change – new operations or activities are commenced and public concerns are directed at new targets. Agencies must therefore review their level of public visibility periodically, and upgrade or downgrade their assessment as necessary.

---

[7] Australian Standard, AS 4390.5–1996, *Records Management*, 6.3.

2. Purpose of the website

Websites are used for a number of purposes, including:

- disseminating (publishing) and sharing information;
- providing advice;
- working collaboratively;
- advertising goods and services;
- providing online services;
- conducting business transactions;
- soliciting responses and feedback; and
- providing a public record of special political, social or cultural events.

This list can be grouped into three main categories of purpose, namely to:

- publish information;
- communicate and collaborate; and
- provide (access to) goods and services.

Each of these purposes poses certain risks for a Commonwealth agency, but the *level* of risk entailed depends both on the nature of an agency's business and its public profile (the extent of its dealings with the general public).

3. Complexity of the website

As discussed in Section 2, Diversity of web-based resources, websites range from sets of static documents to dynamically generated sites that are more like applications than publications. Site complexity can therefore be gauged on the basis of:

- whether it is static or dynamically generated; in combination with
- whether it is document-centred or application-centred (that is, its level of interactivity).

From a recordkeeping perspective, websites that comprise static documents and incorporate little or no interactivity are relatively simple to deal with. On the other hand, sites which incorporate high levels of interactivity and comprise dynamically generated pages are very complex and prove more difficult to archive effectively.

4 Frequency and regularity of content change

As one of the purposes of websites is to provide current information to visitors and customers, the content presented on a website changes. Changes may take place on a regular basis, with all new or updated material uploaded to the website once a day, week or month. In other cases, changes may occur on an 'as required' basis or, in the case of database-driven sites, in real time. For the purposes of both accountability and site maintainability, it is important that records of website content changes are made and kept.

In determining the best way for an agency to deal with tracking and recording content change over time, it might be helpful to use the following terms to refer to site changes:

- frequent – three months or less between changes;

- infrequent – more than three months between changes;

- regular – changes are made on a planned basis (eg once hourly, twice daily, every Thursday, or on the 28th of each month);

- irregular – changes are not planned, but made on an 'as required' basis.

An agency website's rate of change could then be categorised using one of four labels:

- regular and frequent;

- regular and infrequent;

- irregular and frequent; or

- irregular and infrequent.

Websites that change irregularly but frequently 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 impacts on the overall level of risk is dependent on the other risk factors.

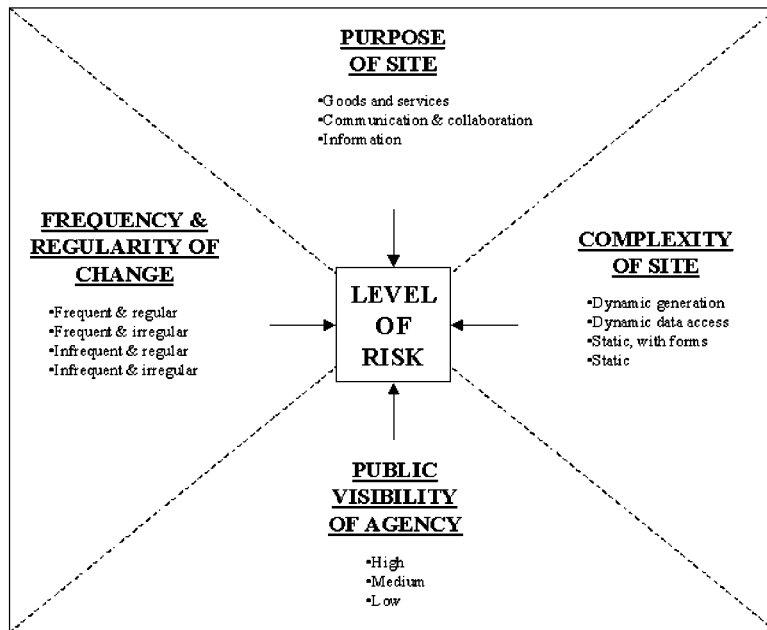## 4.2   Determining an acceptable level of risk

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

Figure 1, Risk factors to consider, illustrates how the four factors outlined above interact to form a level of risk. These factors need to be considered in the context of the current state of recordkeeping within the agency. Where the recordkeeping environment has significant weaknesses, the strategies adopted to capture and maintain web-based records may need to relatively more stringent. The agency's ability to satisfy recordkeeping requirements across its organisation, including requirements to maintain web-based records, will influence the level of risk posed by the website to the agency.

Once the level of acceptable risk is determined, an agency should select recordkeeping strategies that are appropriate to that level of risk.

The constant evolution of web-based technologies, emerging trends and other changing environmental factors mean that it is necessary to periodically re-examine the level of risk faced by an agency in maintaining a web presence.

FIGURE 1   RISK FACTORS TO CONSIDER

PURPOSE OF SITE
- Goods and services
- Communication & collaboration
- Information

FREQUENCY & REGULARITY OF CHANGE
- Frequent & regular
- Frequent & irregular
- Infrequent & regular
- Infrequent & irregular

LEVEL OF RISK

COMPLEXITY OF SITE
- Dynamic generation
- Dynamic data access
- Static, with forms
- Static

PUBLIC VISIBILITY OF AGENCY
- High
- Medium
- Low

5.   STRATEGIC AND TECHNOLOGICAL OPTIONS

This section outlines some options available for capturing and preserving web-based resources and records of web-based activity. It is important to note that the options suggested are not mutually exclusive. In most cases agencies will pursue a combination of strategies designed to fit their circumstances and requirements. It should also be emphasised that, with agencies existing in a variety of operational, technological and risk environments and with web-based technology in a constant state of flux 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.

Object-driven versus event-driven strategies

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

Agencies may consider object-driven or event-driven approaches or, better still, a combination of the two.

Object-driven approach

This approach concentrates on managing the 'objects' that constitute or are made available via a website.

Object-driven strategies are well suited for websites that comprise collections of HTML documents and do not rely on complex interactivity with users of the site. Objects could be complete HTML documents that are stored and served up to the user. Alternatively, objects could be the various objects that are assembled to create an HTML document when a user queries the site (eg headers, footers, corporate logos, images and text content).

This approach could entail taking periodic snapshots of collections of web resources in combination with tracking changes to the site and logging transaction details (see Section 5.3, Snapshots). Alternatively, objects or individual web resources could be separately captured and managed in association with metadata that describes the relationship between specified versions of the object and its unique URI (see Section 5.2, Managing objects separately).

Adopting an object-driven approach for websites that primarily provide transactional services may be futile. In such circumstances, an event-driven approach may be more appropriate.

Event-driven approach

This approach focuses on capturing 'events' or transactions that occur between the website and the user.

This approach is most suited to the scenario 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 using the object-driven approach would involve keeping track of changes to enable its full reconstruction as at 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' – single transactions between website and user – rather than the objects that comprise the site at the time of the transaction (see Section 5.5, Logging and keeping records). An event-driven archiving approach 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.

Managing objects separately

This strategy involves managing data objects separately together with sufficient metadata to document how each object is associated with a particular URI. The object-driven approach reduces the preservation burden by focusing on preservation of the data objects and associated metadata instead of attempting to preserve entire systems that support web resources.

Procedures for managing objects separately

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

- the absolute URI;

- the data object;

- the mime type of the data object;

- 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 object individually, this method of archiving does not enable reconstruction of the website because the data objects have been separated from other components of the website.

Responsibilities

The main responsibilities to assign include:

- determining if this is an appropriate option and whether it should be supplemented by other recordkeeping strategies;

- creating and managing the metadata store, including determining what metadata should be collected to document the relationship between each URI and data object (a relational database could be used to manage the metadata);

- maintaining the data object store over time;

- ensuring sufficient links to the recordkeeping system are established; and

- identifying preservation implications to ensure that web resources are accessible over time for as long as required (eg storage media and migration requirements).

5.3 Snapshots

A snapshot usually involves creating a full and accurate record copy of an agency's public web resources at a particular point in time. The snapshot should be captured into a recordkeeping system and maintained over time for as long as the snapshot needs to be accessible.

When taking snapshots of collections of web resources, it is desirable to ensure (as far as possible) the continuing processability of the website 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 viable where risk assessments have been carried out to determine the frequency at which copies should be created and captured into recordkeeping systems. It is particularly useful for static resources or collections of static objects that are essentially an agency's electronic publication(s). Used in combination with other options, this strategy will facilitate compliance with *Archiving Web Resources: A Policy for Keeping Records of Web-based Activity in the Commonwealth Government.*

A snapshot is an object-driven approach and should not be used to keep records of highly interactive dynamic sites or resources which are databases or transactional services. A deficiency of this approach is that a snapshot only provides a picture of a website at a particular point in time. 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 archiving method does not enable the agency to determine exactly when particular web resources were available, agencies that use the snapshots strategy should also

create and maintain logs of changes made to web resources between snapshots (see Section 5.4, Tracking changes).

The *Administrative Functions Disposal Authority* developed by the National Archives provides guidance on the need to create and retain snapshots of an agency's website. It also provides guidance on the creation and retention of the records documenting the risk analysis to determine the frequency of creating snapshots (see Section 3.3, Determine requirements for records).

Procedures for creating and capturing a snapshot

A snapshot should include all aspects of the website to ensure that a fully functional site can be reconstructed. For example, the snapshot should also include scripts, programs, plug-ins and browser software, that is, all components that make the snapshot fully functional. The snapshot should be captured into the recordkeeping system with sufficient descriptive metadata (see Section 3.4, Apply metadata).

It may be necessary to make some modifications once the snapshot is created. For example, a CGI script for site counters will need to be disabled. If site counters are not disabled, there will be no accurate or authentic record of the number of visitors to the site at the time the snapshot was created. In effect, the record is no longer a snapshot of the site.

Responsibilities

The main responsibilities to assign include:

- determining if this is an appropriate option and whether it should be supplemented by other recordkeeping strategies;

- determining how frequently copies of web resources should be created;

- creating the snapshot;

- capturing and maintaining the snapshot in the recordkeeping system including the capture and maintenance of sufficient metadata;

- selecting an appropriate storage medium and undertaking data management tasks and quality control checks; and

- identifying preservation implications and ensuring the records are accessible for as long as required.

Website administrators or information technology staff may already carry out the task of creating 'back-ups' of the website as part of normal data management activities. However, because these back-up copies are created for the purpose of data management activities, they are usually overwritten regularly with more recent versions, or deleted. They are not captured or maintained for recordkeeping purposes. To be used as a viable recordkeeping strategy, it is necessary to intervene and establish processes and procedures to ensure that snapshots are created and captured into the recordkeeping system and maintained over time for as long as required. The above responsibilities represent the minimum list of responsibilities that should be documented in agency's procedures and assigned to recordkeeping practitioners, website administrators and information technology staff.

5.4   Tracking changes

This strategy involves tracking changes to the web resources over time and creating a log of changes or activity. The activity log needs to be captured into a recordkeeping system and maintained to satisfy requirements for accessibility for as long as needed. Used in combination with snapshots of the web resources, this approach can be a reliable option for static sites.

The main problem arising from this option is the creation of insufficient metadata of the activity log, resulting in the inability to interpret the log over time. It is vital that metadata requirements are specified and sufficient metadata is captured into the recordkeeping system (see Section 3.4, Apply metadata).

Procedures for creating and capturing activity logs

Suggested data elements that can be captured in an activity log include:

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

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

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

It may be possible to use emerging web technologies to track changes. Web robots, spiders or crawlers are automated programs that visit sites for the purpose of indexing sites for search engines. These programs may be useful for tracking changes, provided they gather sufficient information to satisfy recordkeeping requirements.

Responsibilities

The main responsibilities to assign include:

- determining the list of data elements that should be captured in an activity log;
- establishing procedures and processes to ensure the activity log is created, updated and maintained over time;
- capturing and maintaining the activity logs in the recordkeeping system, including the capture and maintenance of sufficient metadata;

- selecting an appropriate storage medium and undertaking data management tasks; and

- identifying preservation implications and ensuring the records are accessible for as long as required.

It is essential that agencies assess their systems, policies, procedures and other tools, and customise this list of responsibilities to ensure that identified recordkeeping requirements can be satisfied.

5.5 Logging and keeping records of transactions

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

This event-driven strategy involves creating a log of site visitors, capturing the logs and any other records of web-enabled transactions into a recordkeeping system together with sufficient metadata, and maintaining them as long as required. This option enables the capture and maintenance of evidence of site use, particularly any queries or transactions enabled by the site.

A problem may arise if appropriate log analyser software is not available. Although all web servers generate log files of server activity, raw log files are confusing and difficult to decipher. 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.

A final issue to consider is privacy. Most logs of website transactions will contain personal information about the user. Commonwealth agencies need to be aware of their responsibilities under the *Privacy Act 1988* with respect to the collection and security of personal information.

Procedures for creating and capturing logs of site visitors

Although website administrators routinely create log files of site visitors, these logs were designed to meet website administration needs, not agency recordkeeping needs. It is therefore necessary to establish procedures and processes to ensure that visitor logs satisfy recordkeeping requirements and are captured and maintained in the recordkeeping system.

A number of software packages are available that have the capacity to configure raw log files and produce comprehensive reports in the form of tables and graphs. The elements that can be logged 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 satisfy their recordkeeping requirements.

Responsibilities

The responsibilities for this option are the same as those listed in Section 5.4, Tracking changes – creating and capturing activity logs.

5.6 Maintaining online archives

An online archive is intended to replicate, at the time of posting, all material posted to the active website. Unlike an active website, the online archive must capture past as well as present postings. Postings to the online archive should be linked to a recordkeeping system with sufficient metadata.

The main advantages of this option include:

- facilitating maximum functionality of archived postings;

- providing greater accessibility to archived postings, instead of reconstructing sites from an offline storage medium;

- offering a more appropriate option for complex sites that are more than simple static electronic publications; and

- enabling the reconstruction of the site at any point in time rather than only when a snapshot is created and therefore providing greater accountability for the agency.

Where an online archive is carefully planned and responsibilities are documented and assigned, there are few deficiencies associated with this option. It relies, however, on linking the recordkeeping system with a dedicated web server. Costs and other practical considerations associated with establishing and maintaining such a server will need to be balanced against benefits. In particular, maintaining all past and current website postings may require large amounts of storage space, with negative cost implications for this strategy.

Procedures for establishing an online archive

The creation of an online archive (a dedicated web server) and the links to the recordkeeping system should be carefully planned and designed to ensure both recordkeeping and system requirements are fully considered. Agencies should ensure the necessary consultation occurs between recordkeeping practitioners, website administrators and IT staff. The DIRKS Manual provides guidance on how to ensure full and accurate records are created, captured into recordkeeping systems and maintained for as long as required. Modifications to website page counters will also need to be considered.

Responsibilities

The main responsibilities to assign include:

- determining whether this is an appropriate option, perhaps undertaking a cost-benefit analysis and comparisons with other options;

- identifying system specifications;

- setting up procedures to ensure that postings are captured;

- ensuring that links are established with the recordkeeping system and sufficient metadata is captured;

- identifying procedures and practices to ensure the system integrity of the online archive (eg data management tasks are identified and responsibility allocated);

- identifying preservation implications and ensuring the records are accessible for as long as required; and

- implementing compliance programs to facilitate post-implementation review.

The above responsibilities represent the minimum list of responsibilities that should be documented in agency's procedures and appropriately assigned. It is possible for agencies to add to and alter this list based on their specific requirements. The important point to emphasise in this approach is that there is a significant reliance on the early establishment of clear procedures. Consultation with various areas of the agency is essential for ensuring that the dedicated servers to meet the requirements for the online archive and that links with the recordkeeping system are established and maintained. To ensure the success of this approach, compliance with established procedures needs to be monitored.

6.   STORAGE AND PRESERVATION ISSUES AND STRATEGIES

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 further industry research.

In 2001/2002 the National Archives will begin an infrastructure development program for accepting electronic records into custody. This program will include research on electronic preservation issues, preferred storage media and recommendations on procedures for transferring electronic records to the Archives.

It is crucial for agencies to be familiar with current preservation issues and best practice recommendations for web-based records.

6.1   Hardware and software dependency and obsolescence

All websites, 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 (eg software plug-ins such as Adobe Acrobat Reader, scripts, applets, search engine). Often, these applications are customised for a specific environment and will only run on a particular hardware configuration or operating system platform.

Computer technology is subject to ongoing technological obsolescence, with both hardware and software quickly becoming outdated as new upgrades and versions come onto the market. Electronic material created under older systems becomes unreadable (and hence inaccessible) in the original form after relatively short periods of time. Agencies taking website snapshots for online or offline storage need to plan for technology obsolescence.

There are a number of interrelated software and hardware factors which agencies need to consider when maintaining snapshots of websites as records, including:

- the evolutionary nature of the standards for markup – the existence of different versions and types of HyperText Markup Language (HTML), each with different functionality, and the increasing use by software developers 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 webpages (for example, older browsers are not frames-capable and are unable to execute JavaScript, leading to a loss of intended functionality);

- the estimated physical and/or commercial life of the medium on which a website snapshot and its related descriptive metadata are stored; and

- the long-term availability of the hardware and operating system platforms needed to access records stored on different types of media.

6.2   Maintaining web-based records over time

Ensuring the accessibility of web-based records over time raises the same issues that apply to other electronic or paper-based records. Australian Standard AS 4390–1996, *Records Management* emphasises the need to maintain the functionality of a record over time for as long as required.[8] The following accessibility issues should be considered.

Ensuring records are carefully managed

This might include:

- maintaining records in widely accepted technology-neutral storage and data interchange formats such as XHTML and avoiding the use of non-standard HTML tag extensions;[9]

- where appropriate, keeping records in two different formats;

- maintaining preservation master sets and storing these in a separate location;

- exercising and refreshing media on a regular basis; and

- carrying out random spot checks to monitor the functionality and integrity of records.

Planning for obsolescence by ensuring that records can be copied, reformatted or migrated

This includes hardware, software, operating system and media obsolescence. Whenever records are copied or reformatted, it is recommended that technology tools such as 'checksums' and 'hash digests' are used to confirm the maintenance of data integrity. Web-based records and their associated metadata should be migrated as often as necessary to avoid technological obsolescence for as long as the records are required. Any preservation actions such as copying, reformatting or migrating should be documented in the recordkeeping metadata. Any loss of functionality, content or appearance that occurs as a result of reformatting or migration to standard formats should be fully documented in the metadata.

Using widely supported standards (eg open architecture standards)

When designing and building websites, agencies should plan to use software tools and applications which meet accepted (or de facto) standards, and which are readily available and fully supported.

---

[8] Australian Standard AS4390.3-1996, *Records Management*, 8.7.2 & 8.7.3.

[9] According to Charles Dollar: 'Currently, the W3C [World Wide Web Consortium] is promoting "data independence" of Websites and Web pages through its support for XHTML as a technology neutral "open" format. The proponents of XHTML believe that this strong base of support will encourage the developers of software tools for Websites and Web pages to ensure that their products conform to XHTML. The support of W3C for XHTML, the availability of multi-vendor products that conform to XHTML, and a wide base of user support suggest that XHTML and its variants are likely to persist over time.' Charles M Dollar, 'Archival preservation of websites and web pages: strategy, principles and guidelines', Draft 22 November 2000, p. 12.

Implementing security measures to protect records against either deliberate or accidental alteration

Some possibilities include:

- maintaining controlled access to a secure storage facility that enables only authorised staff to access the records;

- compliance and audit programs to ensure security procedures are maintained; and

- providing 'read-only' access to the records.[10]

Using persistent identifiers

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 URI, DOI, PURL 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. It is recommended that agencies observe the advice issued by the National Library of Australia in its publication, *Managing Web Resources for Persistent Access*.[11]

Ensuring environmental control and monitoring

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.[12]

6.3   Selection of storage media

Depending on recordkeeeeping requirements, agencies need to decide whether to capture and maintain web-based records on an offline or online storage medium. The size and complexity 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.

Below is a brief description of four widely used storage media.

---

[10] See Charles M Dollar, *Authentic Electronic Records: Strategies for Long-Term Access*, Cohasset Associates, Chicago, 1999, p. 97.

[11] National Library of Australia, *Managing Web Resources for Persistent Access*, exposure draft, February 2001, available at:
http://www.nla.gov.au/guidelines/2000/persistence.html

[12] For further reading on these issues, refer to: Dollar, *Authentic Electronic Records,* 1999.

- CD-ROM (compact disk – read only memory)
  Optical disk technology is capable of storing large amounts of data that can be read but not altered. CD-ROMs all conform to size and format standards and are well suited for colour, large software applications, graphics, sound and video. CD-ROM technology adheres to ISO 9660 which covers both the physical layout of the disk and the format of the recorded information.

- CD-R (compact disk – recordable)
  Based on WORM (write once read many) technology, a CD-R can store large amounts of data. CD-R technology also adheres to ISO 9660. CD-R drives have been improved to enable multi-session recording (that is, data can be added over time). Standard error checking techniques should be used to assess the quality of the blank discs being used for storage.

- Magnetic tape
  A magnetically coated strip of plastic on which data can be encoded, magnetic tape provides relatively inexpensive and large storage capacities. Because tapes do not allow random access to data, access time is slower on tape than on disks. Tapes are available in a range of sizes and formats.

- Magnetic hard disk
  A hard disk (as opposed to floppy disk) is a magnetic disk that can store large quantities of data. However, hard disk storage is more expensive than other storage media.

7.   DETERMINING THE BEST OPTION

7.1   Map actions to circumstances

There is no generic solution for creating and maintaining records of web-based activity. The best option will depend on the outcome of an analysis of the particular circumstances. Each agency should assess a number of factors, including:

- the type of web-based activity – information dissemination or transaction or both;

- the complexity, diversity and sophistication of the web-based activity;

- the frequency and regularity of change to the web-based activity;

- the frequency of challenges to the validity of the information on the website;

- the level of risk and public visibility of the agency;

- the agency's recordkeeping requirements;

- the agency's technological environment; and

- the availability of resources.

The most appropriate recordkeeping strategy is likely to use a combination of approaches. For example, taking a regular snapshot of the site, maintaining a log of each change made to the website and maintaining a log of transactions.

Careful consideration of the deficiencies and advantages of each approach, along with other practical considerations, such as feasibility and cost-benefit analysis, will determine choices. The rationale for selecting a particular approach should be documented.

7.2   Produce action plan and procedures

When the most appropriate strategy has been identified, a team comprising recordkeeping practitioners, website administrators and information technology staff should be selected. In some agencies, the team might include communication managers, as they usually have overall responsibility for managing the content and design of websites. The team will be responsible for developing an action plan that includes policies and procedures. The plan should cover the implementation process and timeframe, resources required, maintenance and training requirements, and a monitoring and review regime. It should be documented and approved by senior management and be widely promulgated throughout the agency. The plan should raise the awareness of the recordkeeping implications of websites and define and communicate staff roles and responsibilities. Additional training may be required to ensure that each person is aware of and understands their individual roles and responsibilities. Mechanisms for effective communication between all areas assigned roles, responsibilities and tasks should be established to ensure effective coordination in archiving web resources. The action plan and policies and procedures should be integrated into the broader recordkeeping strategy and policies of the agency.

7.3   Conduct regular reviews

A review of the agency's policies and procedures for web-based recordkeeping should be conducted shortly after implementation. The timing of the initial review may depend on the strategies adopted. At a minimum, the review should be conducted within six months of implementation. The purpose of the review is to identify the effectiveness of the approach, measure performance, identify any corrective action or enhancements required, minimise the agency's exposure to risk through procedural failure and develop a program for conducting regular reviews.

The review is an opportunity to identify any changes in circumstances that may mean a different approach is required (eg purpose and complexity of the website, frequency and regularity of content change, level of risk/public visibility of the agency). The review is also an opportunity to determine whether records that exist on websites or are about to be placed online are being captured into a recordkeeping system and to review the frequency of snapshots.

The review team should include recordkeeping practitioners, website administrators, information technology staff and content authors or managers. Each of these groups will bring a different perspective to the review process. Including content authors or managers in the review process provides an opportunity to increase their awareness of recordkeeping responsibilities and strategies and to identify any changes to the purpose and complexity of the website that have occurred or are about to occur.

There are two aspects to the review: system performance and risk assessment.

Recordkeeping system performance

Assess and test the performance of the system to verify that:

- the system is complying with appropriate standards;

- the objectives of the system and the needs of the organisation are being met;

- personnel involved in the management and ongoing operational activities associated with archiving web resources are carrying out their specific roles and responsibilities; and

- policies and procedures are being followed.

The specific performance testing measures used in the review will be determined by the approach adopted for archiving web resources. They will need to address the following questions.

- Is the transactional log working effectively and capturing details of who visited the website, when and what actions they performed?

- Are changes to individual documents able to be tracked effectively?

- Is it possible to identify all changes over a specified period of time?

- Is the material being archived retrievable, accessible, readable, adequately preserved and functional? (For example, does the search engine work?)

- Are links functional?

- Are copies of scripts and correct versions of plug-ins included to facilitate functionality?

- Does the archived website have all the functionality of the original site?

- Do snapshots enable the reconstruction of the website for a given date, together with its essential functionality?

- Have people experienced any problems with the system?

- Is there an adequate record of the quality control checks?

Risk assessment

The purpose and complexity of the website should be reviewed regularly to determine if the adopted approach is still appropriate. Whilst a website may start out static, new uses may increase complexity and mean that the original approach to archiving no longer meets the needs of the agency. This is particularly true if a website moves from a static to a dynamically generated site. For example, a website may be established initially for the purpose of information dissemination, but subsequently be redeveloped to enable electronic service delivery.

The frequency and regularity of content change should be monitored and reviewed to evaluate the continuing appropriateness of the adopted approach (eg the frequency of creating and capturing snapshots of the site). The level of public visibility of the agency and the consequent level of risk posed by the website also should be reviewed and upgraded or downgraded as necessary. The level of public visibility is an important factor in determining the appropriate course of action for capturing and maintaining records of the website. Changes in the level of public visibility may mean that a new approach is required.

At the conclusion of the review, a report should be prepared for management that document the findings and recommendations. Any corrective action required should be undertaken and the action documented.

7.4   Monitor developments

Website technology is changing dramatically and rapidly. New technologies, tools, formats and applications are developed almost daily and are regularly implemented by website administrators and content authors. Websites are increasingly becoming more complex and increasingly being used to facilitate business processes rather than merely disseminating information. Website programming is becoming more complex. It is important that agencies monitor web development technologies, emerging trends and business applications, as they will continue to impact on strategies and techniques for recordkeeping and for archiving websites.

One trend is the merging of document management systems with web technologies. Although at present the Web is primarily a read-only medium, new developments such as web distributed authoring and versioning (WebDAV) will provide users with the ability to read-write documents over the Web. Distinctions between electronic document management systems, authoring systems and web-based distribution systems may well dissolve, with applications increasingly residing on the Web and work increasingly being performed in a web environment. Emerging technologies include XML (eXtensible Markup Language), a markup language for documents containing structured information which will enable groups of people or organisations to define their own

customised mark-up language for exchanging information in their particular domain.[13]

In addition to monitoring technological developments, agencies should maintain contact with the National Archives to ensure they are aware of current guidelines covering web-based records, metadata and amendments to disposal authorities such as the *Administrative Functions Disposal Authority.* These guidelines will be updated regularly to reflect technological change and advances in our understanding of strategic and technical best practice in managing web-based records.

8.   COMMENTS

The National Archives welcomes comments on these guidelines, particularly with respect to their practicality and implementability within agencies. Comments and queries about these guidelines may be addressed to:

Adrian Cunningham
Director, Government Recordkeeping Standards and Policy
National Archives of Australia
Phone: (02) 6212 3988
Fax: (02) 6212 3989
Email: adrianc@naa.gov.au

9.   ACKNOWLEDGEMENTS

The National Archives would like to acknowledge the input of the following organisations during the development of these guidelines: InfoProducts Division of the Department of Finance and Administration, Department of Primary Industries (Queensland), CSIRO and the National Library of Australia. Within the National Archives, the input of the following staff members during the formulation of the guidelines was particularly valuable: Karen Skelton, Karuna Bhoday, Adrian Cunningham, Simon Davis, Jenni Davidson, Andrew Wilson and Brad Headland.

10.   FURTHER READING

Australian Archives, *Managing Electronic Records: A Shared Responsibility*, July 1997

Dollar, Charles M, *Authentic Electronic Records: Strategies for Long-Term Access*, Cohasset Associates, Chicago, 1999.

Dollar, Charles M, 'Archival preservation of websites and web pages: strategy, principles, and guidelines', draft, 22 November 2000.

---

[13] 'The successor to SGML (standard generalised markup language, an ISO specification widely adopted as a de facto standard for electronic document management), XML is considered more accessible, less expensive, and easier to implement than EDI [electronic data interchange] as a platform for e-commerce. XML version 1.0 provides specifications for document linking, security and other functions associated with intercompany exchanges of electronic documents.' (David O Stephens, 'International standards and best practices in RIM', *The Information Management Journal,* April 2000, pp. 68–71).

IM Forum – Internet and Intranet Working Group, *An Approach to Managing Internet and Intranet Information for Long Term Access and Accountability*, 24 September 1999, available at: http://www.imforumgi.gc.ca

IM Forum – Internet and Intranet Working Group, *Managing Internet and Intranet Information for Long Term Access and Accountability – Implementation Guide*, 24 September 1999, available at: http://www.imforumgi.gc.ca

McClure, CR and Sprehe, JT, *Guidelines for Electronic Records Management on State and Federal Agency Websites*, February 1998, available at: http://istweb.syr.edu/~mcclure/guidelines.html

Myburgh, S., 'Intranets and the records manager', *The Records & Retrieval Report*, Vol. 13, No. 9, November 1997.

National Archives of Australia, *W*hy records are kept: directions in appraisal, March 2000, available at: http://www.naa.gov.au/records-management/publications/Why-records-are-kept.aspx

National Archives of Australia, *Administrative Functions Disposal Authority*, February 2000 (particularly the Publication and Information Management functions), available at: http://www.naa.gov.au/records-management/keep-destroy-transfer/authorities/AFDA.aspx

National Archives of Australia, *Archiving Web Resources: A Policy for Keeping Records of Web-based Activity in the Commonwealth Government*, re-released January 2001 (the companion document to these guidelines) available at: http://www.naa.gov.au/records-management/publications/Archiving-web-policy.aspx

National Archives of Australia, *A*ustralian Government Recordkeeping Metadata Standard Version 2.0, July 2008, available at: http://www.naa.gov.au/records-management/publications/AGRkMS.aspx

National Archives of Australia and Office for Government Online, *Australian Government Locator Service (AGLS) Manual for Users*, Version 1.1, July 1999, available at: http://www.naa.gov.au/records-management/publications/AGLS-Element.aspx

National Archives of Australia in collaboration with the State Records Authority of New South Wales, *Designing and Implementing Recordkeeping Systems: Manual for Commonwealth Agencies*, September 2001, available at: http://www.naa.gov.au/records-management/publications/DIRKS-manual.aspx

National Library of Australia, *Guidelines for the Selection of Online Australian Publications Intended for Preservation by the National Library of Australia*, last updated 21 December 1999, available at: http://www.nla.gov.au/scoap/guidelines.html

National Library of Australia, *Managing Web Resources for Persistent Access*, exposure draft, February 2001, available at: http://www.nla.gov.au/guidelines/persistence.html

Phillips, ME, 'Ensuring long-term access to online publications', *Journal of Electronic Publishing*, Vol. 4, Issue 4, June 1999, available at: http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main; idno=3336451.0004.405

Powell, TA with Jones, DL and Cutts, DC, *Website Engineering: Beyond Web Page Design*, Prentice Hall PTR, Upper Saddle River, NJ, 1998.

Standards Australia, AS 4390–1996, *Records Management*, 1996.

Stephens, David O, 'International standards and best practices in RIM', *The Information Management Journal*, April 2000.

Udell, Jon, *Practical Internet Groupware*, O'Reilly, Sebastopol, CA, 1999.