managing web resources

MANAGEMENT OF ELECTRONIC RECORDS ON WEBSITES AND INTRANETS: AN ERM TOOLKIT

VERSION 1.0

DECEMBER 2001

First version published 2001

Public Record Office
Ruskin Avenue
Kew
Surrey
TW9 4DU

Designed and typeset by Format Information Design

www.format-info.co.uk

Contents

1 Policy statement

Central government websites (including intranets) contain public records that need to be managed by their creating departments and others who have custody of them. As public records, some government web resources may be of historical value and interest and be appropriate for selection and transfer to the Public Record Office.

Records management principles and procedures need to be applied to government websites in the legitimate interests of all stakeholders. This is for a number of important reasons:

- to support the conduct of public business by ensuring that business requirements for the capture and management of records are met;
- to promote good information management in government (records being an important subset of public information and records management a vital subset of information management);
- to assure accuracy, authenticity, audit and admissibility of web content and web based transactions;
- to preserve accountability;
- to ensure that authentic records can be sustained through time so they are kept for a period appropriate to their use and importance; and
- to support the identification, survival and preservation of website records of historical and research interest in the national archives.

Implementing the policy

The remainder of this records management toolkit is concerned with practical steps to implement the policy as regards active records management and sustainability within departments.

Selection for the national archives will take place within the existing framework of the Public Record Office's acquisition policy and thematic/departmental Operational Selection Policies.

2 About this toolkit

Audience, roles and responsibilities

Website management has often been seen as the preserve of IT specialists, press/communications functions and librarians. In government, it also needs rigorous records management input. This is a point that has often been overlooked.

The primary intended audience for this toolkit is records managers in government, web project managers or IT and information managers with information and records management responsibilities. Some aspects may be of assistance to business managers. It assumes a reasonable level of general IT and information literacy but is not written from a technical IT perspective.

Records managers will need to become involved in systems design, risk management, surveying, retention scheduling, sustaining through time, exporting to new technological platforms as well as appraising for archival purposes and in some cases transfer to the Public Record Office. This requires involvement from a far earlier stage in the records lifecycle than with paper records. It also means having some involvement in the management of information that may not be considered as formal records.[1]

It requires a collaborative approach with IT managers, web project managers, 'webmasters'/site editors, content providers and general users. It also needs the support of senior management. Suggestions on how records managers might get involved in website management and provide added value are in the model action plan in Section 6.

1 *This could manifest itself in a variety of different ways according to the organisational placement and remit of the records management function. The very minimum might be sufficient involvement to ensure the interface between official records in websites and other content are fully understood and catered for. The other end of the spectrum (and increasingly the most desirable one) is for a joined up approach to strategic information management including records and document management, embracing both the content and the technology issues.*

3 Introduction

Purpose

This toolkit outlines the existence of public records in websites and intranets and is designed to explain specific requirements for the management of information and records on websites. It provides pointers to how this can be achieved effectively even where a website is already up and running without rigorous procedures already in place.

The toolkit is also designed to complement the Office of the e-Envoy's Guidelines for UK Government Websites, which gives guidance on the production of effective websites, their management, their content and their accessibility.

It does not replicate the more generic guidance on the management, appraisal and preservation of electronic records already published by the PRO, nor the e-Envoy's Guidelines and other documents. References to these and other useful resources are at the end of this toolkit. Both departments' guidance documents are all available through their respective websites and the addresses (URIs) are given in these references.

Scope and terminology

This toolkit provides guidance for the management of both public websites and networks using web technology to make information available between narrower groups of people, through a departmental intranet, within the government community across the Government Secure Intranet (GSI), or as part of an Extranet linking public and private sectors. For the purpose of brevity, the word 'websites' is used throughout, usually to denote all three categories.

The wider electronic records management picture

At the time of writing (autumn 2001), departments are at varying stages in their progress towards enterprise-wide electronic records management by 2004.

This specific target is related to the general target for electronic service delivery by 2005 and also needs to be seen in a wider electronic information management context that also includes the management of information that does not constitute a formal corporate record (including document management).

There will be a number of direct relationships apparent between the contents of this toolkit and other guidance already published by the Public Record Office to assist with this transition. Web records may (should) have been identified during the compilation of information inventories and fed into appraisal and preservation strategies. Many will have been created in a relatively unmanaged environment, outside formal records systems.

As with other records for which this is true, the ultimate goal for web-based records must be to bring their creation and management within the control of a fully functioning electronic document and records management system (EDRMS) when that is implemented across the organisation. It is to be hoped that the next few years will see an awareness of this need in the design of content management solutions so that records management disciplines are built into operational processes and systems and not seen as either an additional overhead or an 'optional extra'.

What is a 'website'?

A variety of definitions are possible.[2] For the purposes of this publication, we are defining it as: 'A collection of electronic resources:

- **that is made available** in a particular domain of the internet, for the communication of information and/or the conduct of business transactions; and
- **that share a common domain name**, normally belonging to a single or defined group of organisations and having as their virtual location (or Uniform Resource Identifier) a hierarchical (or other) relationship with the main domain content (often referred to as the 'home page'); **and**
- providing a body of interlinked information resources that is navigable using browser technology.'

The extent of user access is not a part of this definition. This policy relates to government intranet and extranet resources as well as public websites on the part of the internet known as the 'world wide web' [3]

These resources may exist on a single server or be distributed among other information resources owned and hosted by one or many different organisations. Hosting content on the distributed model raises additional challenges to effective records management. If content is not under the direct control of the information owners owing to outsourced web hosting or other IT services etc., additional management procedures will need to be put into place to ensure records are controlled according to corporate policy.

Typically, the website content might include a combination of:

- static pages;
- compound pages formed by displaying together content from a variety of sources;

---

2 *The working definition used by the World Wide Web consortium runs 'a collection of information, documents or databases that is provided to a user community using World Wide Web formats and protocols'. The National Historical Publications and Records Commission (NHPRC) funded guidelines for US Federal websites use 'a set of URLs [uniform resource locators] that fall under a single administrative control'. The proviso is entered that the definition is operational rather than analytic and there was no consistency found in patterns of administration. A single organisation with more than one umbrella domain name for its web resources would by this definition have several websites. This may not be seen as helpful but in reality and (as the NHPRC study says) the administrative permutations are endless.*

3 *Some of the supporting records management guidance may also be applicable to private networks.*

- dynamic pages formed according to the user's expressed preferences from a variety of sources;
- active server pages (for the entry of simple information to interact with databases that are an integral part of the site such as a search facility or a telephone directory);
- web forms capturing information for processing in separate database applications;
- graphics, audio files, video clips and virtual reality;
- linked documents such as publications.

Many of these are very different in nature from the traditional image of a 'record'; so much so that it can tend to give the impression that no records are present. This can be highly misleading.

Records in websites

It is no longer a sustainable view that all website content is merely ephemeral.

Early official websites were mainly used to publish documents also available in hard copy. Now they are used to make information available in more imaginative ways (that often have no direct hard copy equivalent) and to conduct business in real time.

Government's rôle in providing information is a vital part of the Knowledge Economy. Access to official information is being dramatically increased under the Freedom of Information Act and the internet will be one of the key enabling technologies in this as well as in implementing electronic government. Implementation of the Freedom of Information Act 2000 will also mandate on all public record bodies compliance with the Code of Practice on Records Management. Meanwhile, the Data Protection Act 1998 has made more stringent requirements for processing personal data to protect individual privacy.

Some examples follow illustrating the diversity of the records management issues that can arise with websites:

Example 1: A department uses its website to conduct a policy consultation

Public comments are invited on a consultation draft and some of these arrive through a comments form on the site itself, as well as by 'snail mail'. The issue is a controversial one raising important questions of public policy. An online discussion forum is part of this process.

Example 2: A public inquiry into a human tragedy uses its website to publicise its proceedings

The site is used not just to raise awareness, but also to air public safety issues, provide the general public with transcripts of the public hearings and invite comment. It is also intended to prompt witnesses of the event to come forward.

Example 3: A department's core business is the provision of information to the public

Owing to the nature of the business, on the face of it there is little to be done in the way of records management (as opposed to web publishing).

However, the department licenses some of its intellectual property rights to commercial organisations and has charging regulations enabling it to recover some of its costs when it makes certain information available. These transactions are partly web-enabled (small payments, standard contracts).

Substantial portions of the information made available through the site are of continuing cultural and, potentially, research interest. Some of the rest (it is difficult to tell from the outset which) will be taken up in the press and could be subject to parliamentary scrutiny.

Example 4: A department's website is used to capture data in a transactional database about social attitudes and conditions, including those of private individuals

Whilst the immediate business issues are to be producing statistical surveys to assist with economic, fiscal and other planning [and data protection might be the first information management priority], the resultant datasets have great potential historical value.

Some of this will only be realisable once the data is no longer about living identifiable individuals, though redaction techniques and other manipulations of the dataset could present possibilities far sooner.

To summarise the effects of all this, there is a paradigm shift occurring in where the prime record occurs. For example:

- if web pages are what the public, user, etc will refer to when making a decision such as entering into a contract or deciding to travel abroad, then both they and the organisation making the information available have a continuing stake in it and there is a business record present; or
- if the website itself hosts dynamic transactions, including purchasing or filling in an official form, then either it or any underlying application needs to capture a record of the transaction and for it to be maintained in accordance with its importance and value; and
- in addition, there is potential record value in the content, transactional functionality and user experience being presented to the public by government in the course of discharging its functions.

As noted above, these web-based records often cannot accurately be represented by printing out to paper. To attempt to do so would not only lose aspects of the user's experience (e.g. video clips, links from one part of a document to another) but potentially even the content itself (in the case of linked and compound documents).

Stakeholders

Virtually everyone has an interest in the management of public sector web resources:

- creating organisations
- the general public
- business
- other public bodies
- the Courts
- the media
- academics
- historical researchers.

The lifecycle of the resources could be summarised broadly as having three stages:

- creation and active use (typically being made available to a wide group of users for information and the conduct of business);
- medium term sustainability to meet business needs for continued retention of the records after they have ceased to be referred to regularly for current activity (e.g. for audit, accountability, information reuse/management reasons);
- archival preservation.

The balance of different stakeholders' interests will vary according to the stage in their lifecycle the records have reached. For example, historical researchers will mainly be concerned with archival preservation. Creating organisations will be interested in all stages, business primarily in the active use phase, and so on. Members of the general public will be concerned with the provision of up to date information and services, but also the privacy of their personal data.

What is done (or not done) at the earlier stages in the life of electronic resources can have a profound effect on what happens to them later. Unless electronic resources are managed right from their creation, there is a likelihood that they might not survive at all. This is because of the fragility of electronic media compared with the traditional hard copy environment and issues encountered with systems migration and software obsolescence.

Arguably (and in common with other electronic records), management of the records needs to begin at the point when the system is designed and well before any records are created in the first place.

4 How websites differ from other electronic resources

Standards for the management, appraisal and preservation of electronic records (including both principles and procedures) are already available on the PRO's website (second editions, 1999). These apply in full to website records. The guidance in this toolkit is consistent with those principles and procedures and they are only reproduced here insofar as they serve to clarify particular issues with websites.

The main issues

The particular records management challenges posed by websites are:

Immature technology

- the main driver for the technology to date has been mainly one-way communication, with less thought given to underlying management mechanisms needed for more involved use, such as e-business;
- the use of interactive and dynamic displays to make websites more informative, attractive and entertaining has compounded this tendency;
- the technology is rapidly evolving, resulting in unstable technological standards.

Content management solutions

Important caveats need to be entered about some of the 'content management' solutions currently on the market in terms of their records management capabilities:

- some websites have very rapidly changing content as full advantage is taken of the 'immediate publication' potential of the web. Content is not always properly considered before posting;
- the facilities for publishing to websites directly from standard office software can encourage uncontrolled publishing;
- version control can be haphazard;
- there is a perception that such 'content management' solutions actually take care of records management considerations. In general, they do not: they either provide document management facilities with little or no records management functionality[4] or are merely a rapid publishing mechanism that can sometimes lessen the likelihood of these aims being achieved or even considered.

4 *A succinct summary of the difference between ERM and EDM appears in the e-Government policy framework for records management, available on the e-Envoy's website.*

Records capture

- capturing the content of websites in other formats, electronic or otherwise, will usually result in loss of functionality or present a degraded reproduction of the original user's experience, especially over time;
- compound documents, drawing together in the user's browser a selection of resources from different electronic objects, pose particular challenges. These are heightened where the selection itself is the result of some 'intelligent' system rather than the users' expressed preferences;
- website content will for many departments represent the first or one of the earliest types of record that cannot be accurately represented by a hard copy printout[5]. Thus a departmental policy that may apply to certain other records of printing electronic records to paper is not a satisfactory option;
- website resources that have the status of records may have been produced and retained in an unmanaged environment, often with little input from records managers;
- appraisal is particularly challenging given the amount of ephemeral information on websites and the existence of some or all of the information they contain in other forms and formats.

Sustainability

- long-term sustainability is an untried area. It is unlikely that public bodies will be able to tackle this by preserving a museum of current computer hardware to enable the exact replication of the website user's experience for posterity. This is owing to the resource and the technological implications;
- a variety of strategies for website sustainability have been suggested, including taking periodic snapshots of entire or partial websites and the contrasting approach of concentrating on the transactional records generated through web hosted business. Both approaches have potential costs and pitfalls that need to be carefully considered.

---

5 *Printing to paper might be considered 'better than nothing' for the representation of a few single-version static pages with very rudimentary linkage, but is of little other use. At the time of writing, as with managing electronic documents on a local area network, it is the view of the Public Record Office that effective records management is not promoted by such an approach. This is because business pressures will be producing change in web deployment and, de facto, changing the record content of websites, whatever the wording of formal records policies.*

5 Management control mechanisms

There are three broad steps to establishing a robust management mechanism where one does not already exist:

- identify which categories of website material need to be captured as corporate records;
- identify methods to address this immediately where necessary (e.g. to manage uncontrolled business risk exposure);
- identify a management strategy and processes to develop for future management within a structured environment.

The next section (Model action plan, page 21) contains guidance on implementing the choices that this analysis will present.

Determining the 'record-ness' of content

It is essential to clarify which website content has the status of a corporate record and which does not. Departmental inventories produced to assist in planning for the 2004 target will assist in clarifying this. Some useful questions to ask are:

- is the website content a unique instance?
- **if so**, what is its importance? Or:
- **if not**, is the website version of business importance *in its own right* (although also held elsewhere?

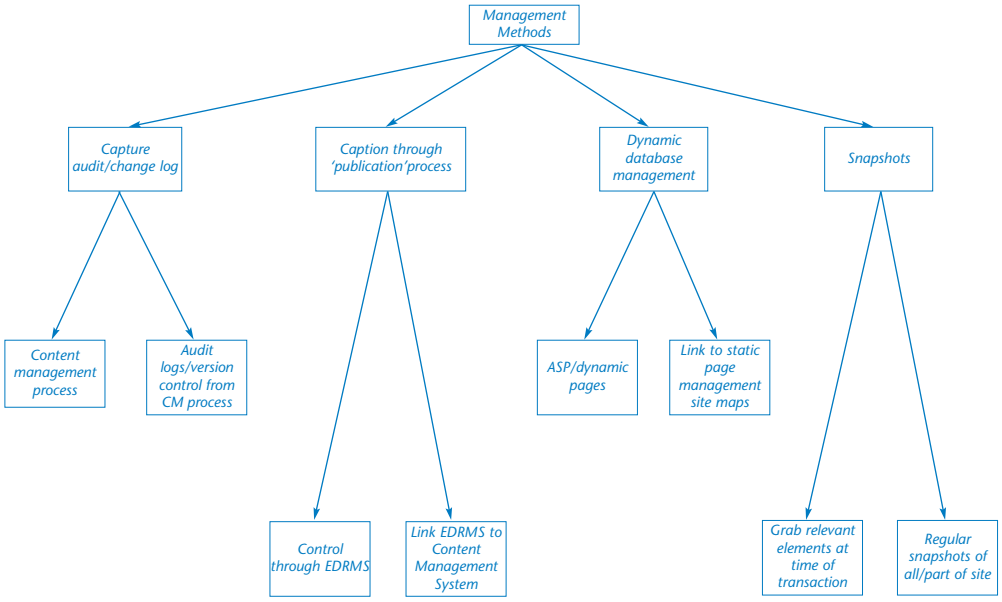
When considering the third bullet point, it helps to examine how useful the information contained in the records is (or what use it could potentially be put to) by the people who view it: see also the heading 'Risk assessment' in the next section starting on page 21.

Website documents rarely exist in isolation – they have been produced by and for a business function that will have created other records to record its activities for business and accountability reasons. There may also be records of the web publishing process itself, but preserving the source documents may not provide evidence of what the website said without some evidence of the publishing process that demonstrates its integrity.

You should clarify for each record type whether your records management procedures need to capture an individual event (e.g. what the website user sees at the crucial moment) or a physical object (that are available to them for access). There is potential danger in attempting to do both half-heartedly and actually ending up achieving neither.

There will still be a need in most cases to capture the point at which the web content was published and removed and to retain that information for a length of time appropriate to its

Figure 1: Schematisation of possible management processes



significance and value (in accordance with an official retention schedule). This can either be done in an EDRMS or in a document management system with version control facilities as appropriate for the status of the content (see below).

Management processes

A variety of management options are possible, from instituting the sort of version control and structure recommended in the PRO's toolkit *Good practice in managing electronic documents using Office 97 on a local area network*, to implementing version control (with or without a full document management solution) and tightening the handling of structured information in databases with full audit logging and capturing content in a fully-functioning EDRMS. A possible structure for these options appears in figure 1 (above).

The solution should be fit for purpose according to the business requirements rather than being driven solely by technology. As already stated, it should also fit into the wider information management strategy for the organisation. It is important that procedures are agreed, spelt out and properly understood within the organisation.

Content management – capturing website records in a fully functioning electronic document and records management system (EDRMS)

Website content that has the status of records should ideally be captured at the point of creation and managed throughout its life in an EDRMS to ensure the capture of records management metadata and preserve the contextual and evidential value of the material. The PRO has published functional requirements for ERMS systems and currently tests off-the-shelf systems against these (where this has not been possible from the outset, some pointers in improving management disciplines are provided in the next section). A number of the approved systems have document management facilities or are integrations of a records management client with an enterprise-wide document management system.

There are early signs that software suppliers are beginning to respond to the potential market for linking content management solutions to EDRMS. The approach taken is likely to involve maintaining page structure and change history in the EDRMS and the objects themselves in the content management solution, thereby embedding the records management disciplines into the business publishing process.

This is one of the areas where we expect the situation to change over the next couple of years.

Document management and versioning

Other content also requires careful management that should be complemented by the records management considerations. For example:

- non-current material may be redundant and require removal;
- there may be a requirement to retain access to outdated versions, whilst making their status clear to users (for example, departments are required[6] to maintain policy consultations until the next change of government even if public policy has moved on);
- alternative versions of documents that are substantially the same may be required to promote accessibility (e.g. HTML rather than just *.pdf*, minority languages or plain English versions).

This requires proper version control mechanisms and compliance with the e-Government Metadata Framework resource description metadata standard (e-GMF/e-GMS)[7].

Managing content common to both website and intranet

One of the major advantages of web technology advocated in the private sector is the ability to make the information resources that exist inside an organisation available outside. This can be limited to modest websites or embrace comprehensive knowledge systems.

---

6 e-Envoy's guidelines

7 On the e-Envoy's website

Where security considerations allow, it is worth public sector organisations considering the potential benefits of having common source files for both information views (i.e. being made available both within and outside a single organisation).

Extensible mark-up language (XML) and other technology can enable different views of the same text files to be made available to users in different domains. Alternatively, it may be possible to control the content at source file level and publishing into both domains. For material with the status of record, the integrity of this process will have to be assured by robust, contemporaneous record and metadata capture. This requires the deployment of an EDRMS as stated on the previous page and careful metadata configuration (i.e. of publishing domains).

The role of site maps and control of uniform resource identifiers (URIs)

Site maps are an invaluable tool to assist in the technical management of large websites. Used with an inventory[8], they can assist the records manager and others to understand how both the logical and the actual structure of the site have been implemented and developed over time. Sophisticated tools are available for the automatic generation of site maps. These have an obvious relevance to the management of dynamic content (see page 18) and extended retention of them may be essential to website records management.

Some system of tracking the development and allocation of uniform resource identifiers (URIs or 'web addresses') within a site needs to be put into place. This could prove critical if the objects forming the user's experience of the site are distributed across different servers or even organisational and national boundaries. This is essential for the management of content purely as a current information resource – otherwise users will experience very rapidly broken links. Software is available to check periodically whether hyperlinks are still working.

8 *For example the inventory of records departments have recently prepared to establish what records need to be migrated across to the environment with enterprise-wide EDRM and which can be scheduled for disposal or disposed of when the immediate operational use has expired.*

Using site maps to manage record relationships

It is even more important to preserve a record of the relationships between various content if any of it has been found to have the status of a formal record. These relationships will not be based on a simple hierarchy.

There may well be a conventional hierarchical structure apparent in paths users might take to navigate from lists of links on 'home pages' to broad topics or business areas and then down to narrower subjects or individual transactions. However, this is unlikely to be either the only means of navigation, nor the only logical structure of the information: good web material is often authored differently from that always intended for hard copy publication. Other structures will be implemented by Hyperlinks from parts of documents to parts of others and navigation enabled by these, Freetext, subject field and other types of searching.

Overall site structure needs to be borne in mind for managing information and records that might be:

- visible to current users;
- removed from view but retained for record purposes;
- not viewable by accessing the site in the normal way but present and accessible through links from elsewhere (e.g. departmental information asset registers accessed through HMSO's *Inforoute*); and/or
- never viewable by the site's users but related to the administration of the site (e.g. technical metadata, etc[9]).

Managing static content

The model of early government websites with infrequent changes to a small quantity of static content is perhaps no longer the norm, nor compliant with best practice (see the e-Envoy's Guidelines). Some public functions will nonetheless have less dynamic sites than others, including ones that have either a short lifespan or no public policy remit[10].

Some straightforward methods for managing records in this scenario might be appropriate. For example:

- documenting publishing processes thoroughly (including changes to existing pages);
- taking snapshots of small sites formed of static pages (noting the caveats about snapshots in the final section of this guidance).

9 Specific issues relating to the management and preservation of website technical documentation (including for preservation purposes) are intended for the next edition of this guidance

10 An example might be the site of a public inquiry, though these now sometimes invite interaction from the public.

Collaborative working applications on intranets[11] (e.g. newsboards, instant messaging, chatrooms)

Interaction with users through a website (i.e. interactions causing changes to some content rather than just displaying for information or entertainment) produces records of some sort; for example, submitting a feedback form or ordering a publication.

Much of this is likely to be purely ephemeral, for example:

- newsboard or discussion group content is likely to be ephemeral unless it is feedback to a public policy consultation (in which case it requires to be captured with the other records of that function);
- this sort of material can rapidly clutter up web pages. Whilst open discussions may require the other contributions to remain visible and this is a valuable contribution to the openness of debate, government departments could potentially be broadcasting libel if their websites host defamatory statements posted by users;
- communications such as requests for the webmaster's assistance or free publications ordering that may not really require formal recording mechanisms.

Underlying database applications

Records managers need to be clear whether users' interaction is with the website itself or with an integral underlying application that happens to be interfaced through the website. This means of access may not be exclusive.

An underlying application may have its own audit trails and its own records capture mechanisms. Many of these will involve databases that may have been around for quite some time but are being used in new ways. Where the application is an integral part of the website itself, the content will require managing according to its business importance. A transactional website specifically designed for substantive business through a web interface will probably require full audit logging functionality.

Possible approaches to managing compound documents on websites

An analogy could be drawn between dynamic web content (e.g. active server pages or '.asp' files – that are populated dynamically from content existing in a separate series of files or links to them from a separate database) and electronic files containing macro fields that

11 *Management processes will need to put into place in circumstances where there is a danger that instant electronic messaging might be used for substantive business. This is because the technology is not fit for this purpose. Organisations have worked hard in recent years to deal with a similar challenge posed by the electronic mail and will need to use a similar combination of policy, persuasion and procedure to manage this risk. At the time of writing (Autumn 2001) actual technological solutions to these problems are not yet apparent. The lack in many cases of a viable printing facility even removes a temporary coping mechanism available in the past with electronic mail*

update according to external conditions in force when the file is manipulated. The website .asp file might not produce the same result if changes have been made to the relevant content it retrieves when the query is run.

The Code of Practice for Legal Admissibility and Evidential Weight of Information Stored Electronically (DISC PD0008, 2nd Edition, 1999) includes handling instructions for 'self-modifying' files[12]. The extra challenge in the context of websites is that the source material is not an integral part of the page used to display the results but is held elsewhere, for example in another directory or even another server which may not even be sited within the UK.

This difference could be managed in a variety of ways, broadly amounting to robust records management of the content potentially made available to the results page file:

- capture of each individual component in an EDRMS;
- capture of associated metadata to include the .asp files, frames etc that the object could potentially have been made available to establish the ways in which these components could have been assembled 'on the fly';
- there may also be a need for metadata elements such as: 'made available [published] on [date]' & 'discontinued from [intranet/internet] domain on [date]';
- an EDRMS with web publishing functionality may have the capability to be configured to facilitate this capture demonstrably *without* human definition of the metadata.

In some business areas, this could be sufficient to manage business risks. Make a careful assessment of how important the information content is to the conduct of business or informing decisions by people to whom a department might owe a duty of care. Taking regular snapshots, audit logging or even maintaining static listings instead might be a more appropriate solution for some applications in higher risk environments.

Example: Some departments use common gateway interface (or 'cgi') files to keep up to date publications lists on their websites. The .cgi file retrieves either the information content itself or a list of active links to the information, which is generated from an unstructured database. Because the database is updated when publications are either discontinued or added, the results from launching the list are different as it does not hold the information independently but runs a query on the database in real time.

One approach might be to preserve snapshots of such lists each time an update is made. Another would be to ensure that the database had a comprehensive audit log function. Either or both of these might be required in high risk environments. An alternative would be to rely on the metadata and the capture of the individual electronic objects.

12 Paragraph 4.16 Self-modifying files (Documentation of procedures for controlling macros)

There comes a point with dynamic content (especially with general .asp format search screens) that it is probably pointless to try to pin down what content would have appeared or did appear in the user's browser at a particular time. It may be as much as could be achieved to define what could potentially have appeared there. Whilst it may not be possible to reconstruct actual user experience, this may give a better chance of maintaining the content, the information and perhaps a 'fit for purpose' evidential record.

Where business risks are particularly high, it may be prudent to avoid this type of content completely where it constitutes a formal record.

6 Model action plan for bringing existing website records under corporate control

A table and checklist for this implementation plan is in Annex 1 of this Toolkit. It assumes that the records manager has only just become involved in the management of the website. Records managers with prior involvement may wish to use the checklist to identify which stage they have reached and what remains to be done.

Risk assessment of the website and its use

Your department will need to conduct a risk assessment to ascertain which of its web content is a vital record, which a corporate record and which merely ephemeral and/or a version of something adequately captured elsewhere that can be disposed of once it is no longer required on the site. As a department's use of its website will develop quite rapidly, you need to repeat the risk assessment at regular intervals. In the present period of rapid change, a minimum of every six months is recommended.

Your risk assessment needs to take into account:

- business transactions conducted directly through the website; and
- the interests of *all* stakeholders in the information presented there.

For example, some departments use the internet to provide information on the strength of which external individuals and organisations take decisions.

Depending on the importance of those decisions and the part played in them by the information, a department could expose itself to financial and other liabilities should it find itself unable to provide evidence of what information was on its website at a particular time.

- The risk assessment is likely to be the best opportunity to convince senior managers of the need for records management of websites.

Example scenario 1: A government department or agency uses its website as an official channel to warn of serious public health or personal safety risks.

Risk: If the department was unable to demonstrate what the relevant page of the site said at a particular time, it might be vulnerable to civil claims should citizens fall ill or find themselves in jeopardy because individuals might be able to assert that the warning was not made available to them at the relevant time.

Example scenario 2: A regulatory body has a provision in its founding legislation that indemnifies the directors of client organisations from personal liability should a breach of the civil law occur whilst following its advice.

Risk: If the regulatory body has put guidance on its website on the subject but is unable to demonstrate exactly what it said at the relevant time, it might be unable to argue that directors had not acted in good faith on its advice and consequently unable to use its regulatory and enforcement powers to call them to account.

Content audit

- audit website content as you would other records and information. Guidance is available in *Guidance for an inventory of electronic record collections*. This will produce a full picture of what records are currently being captured;
- consider the results of the content audit carefully in the light of the risk assessment. Specifically, map the information flows involved in exercising the relevant public function as there could be areas where records capture should be happening but is not. In the electronic environment this issue can be hidden. Urgent corrective action may be required.

Introducing records management procedures

Records capture and standards

- design new records capture procedures where these have been shown to be missing during the content audit;
- in the electronic environment, it is important that such issues as records appraisal, retention scheduling and metadata standards are applied as early as possible. Ideally, this would have been at systems design stage. Where this has not been possible, it needs to be as close as possible to (and preferably *at*) the point of records creation. Records managers will need to take the lead in setting policies and standards for these;
- there may be additional records capture issues highlighted in the next section (*Sustaining web records over time*, page 25) for records that are to be required for long periods or even permanent preservation;
- there is additional information in the *Toolkit on Sustainable electronic records: strategies for the maintenance and preservation of electronic records and documents in the transition to 2004;*
- ensure that metadata standards for the website comply with the e-GMS metadata standard published by the Office of the e-Envoy and any records have additional metadata required for records management.

Publishing process

- do you need to maintain a separate directory of material that has had its format converted for posting on the website? Your publishing processes may make this desirable in case any mishap occurs and versioning becomes unclear (or, alternatively, for back-up and information reuse purposes;

- if the separate directory option is chosen, the *Good practice in managing electronic documents using Office 97 on a local area network*, may contain useful guidance on its management. Document management and/or version control software may also help;
- alternatively, reexamine the publishing process: there maybe ways of maintaining more robust audit trails of what content was accessible on the site and when, for example using an EDRMS to capture the content itself.

Terminology

- when IT colleagues use such terms as 'metadata' and 'archiving', they may not be using them in the way you would. Discuss to clarify your understanding.

Disposal scheduling

- it is irritating and potentially misleading from the website users' point of view to be looking at out of date information. Your concern for managing the records has a direct benefit to users;
- base disposal schedules for website material on business need. There may be special considerations arising from the points flagged up in Section 5 of this Toolkit (*Management control mechanisms* pp. 13–20);
- there will be instances when administrative retention by the creating department is required for a period after a version has been sent for permanent preservation to PRO. This should be discussed with the PRO client manager and factored into the retention schedule. Annex 2 contains a sample of how this might be expressed on a Departmental retention schedule;
- the disposal schedule needs to specify two disposal events: removal from the website and **final** disposal (i.e. destruction of the material or transfer to the PRO);
- collaborate with the web publishers and IT function to ensure the retention schedule is understood and implemented consistently;
- some internet users may be able to view non-current website content that has had its links from home and other pages removed. It is better to move it to another place if it is still required.

Appraisal and scheduling of compound documents

- the levels of granularity involved are, theoretically, limitless. Some implications of the suggested approach to management of compound documents on websites are as follows:
- the purpose of retention scheduling is to produce class rules for retention of records of a similar nature. On websites, you may need to approach scheduling from both the browser end (as normal and familiar – what is seen?) and from the server end (i.e. what objects are there present?) and resolve any retention inconsistencies by retaining for the longer of the two periods;
- it may be possible to simplify administration of this sort of issue in a fully functioning EDRMS compliant with the PRO's functional requirements. Some of these products resolve similar issues with non-web records by inhibiting disposal of an object apparently covered by differing retention periods until the *later* of them is reached).

Implementation of disposal schedules

- final disposal should mean just that. Normal deletion of an electronic document merely removes the pointer to it: it will remain there (and could potentially be recovered) until it is written over. If you do not have an EDRMS in place, you will need to ensure the material cannot be viewed any longer (and not just by normal website users).

Sustainability

- feed back to your colleagues the implications of the sustainability issues in the next section of this Toolkit and the toolkit on Sustainable electronic records.

7 Sustaining web resources over time

There will be business needs for sustaining access to some website content over long periods of time. An EDRMS fully compliant with the PRO's Statement of functional requirements will have some exporting functionality to enable migration to another system. This improves, but does not resolve all the issues of sustaining records across time, especially those relating to software obsolescence.

Another toolkit, *Sustainable electronic records: strategies for the maintenance and preservation of electronic records and documents in the transition to 2004*, contains more general guidance on tackling sustainability issues within departments between now and the implementation of full EDRMS. This section should be read in conjunction with it.

File and text formatting

The e-Government Interoperability Framework (e-GIF) stipulates access standards and specifications for government websites to ensure websites are properly inclusive. Many of the file formats used to support videoclips, virtual reality, sound and other interactivity will be proprietory formats that do not conform to any industry-wide standards and may become obsolete very rapidly. This is not only problematic for current accessibility: it could also prevent the future retrieval of their information content and/or presentation. Some organisations may take the decision to avoid this sort of format as a result.

Even with ordinary documents created in office suite type software (such as Microsoft *Office*), there is no guarantee that future versions of the software will be able to handle files held in their native format in a previous version of the same software. Documents in these formats may sometimes form part of website content. Refer to the separate toolkit on sustainability strategies for further discussion of this point.

Certain file formats are emerging as industry standards and these may provide some degree of insurance against this tendency. Examples are:

- **.pdf** (portable document format)
- **HTML** (hypertext mark-up language)[13]

These may not be true 'open' (or completely vendor independent) formats[14]. Merger, insolvency, takeover or just product development by any of the owners of these formats could lead to obsolescence.

Records, IT and information managers are advised to maintain an awareness of the recommendations of the World wide web consortium (W3c), an influential consortium

---

13 Proprietory RTF (rich text format), such as Microsoft's, has the same backwards compatibility issues as other word processing formats. There may also be issues with some of the functionality of proprietory versions of HTML.

14 There are proprietory versions that need to be distinguished from 'clean' HTML (software is available to check the status of HTML content in this respect).

making recommendations for the world wide web (http://www.w3c.org). It is government policy that our websites support the W3c recommendations for such things as the Web Accessibility Initiative. (Refer to the *Guidelines for UK government websites* published by the e-Envoy).

Examples of W3c recommendations that may prove relevant to the management of web based records include:

- **XML**[15] (extensible markup language and the standard for the e-GIF);
- **XHTML** (a version of HTML that complies with XML recommendation published by W3c);

Preservation of website 'snapshots'

Organisations where the website risk exposure (as identified in the risk assessment) is high may need to capture full or partial snapshots of their websites at regular intervals to manage these business risks. For simple websites a program such as Adobe *Webcapture*, *Teleport PRO*, *HTTrack* or *WebCopier* can be used to achieve this.

Taking snapshots for records management purposes is a separate issue from the Office of the e-Envoy's recommendation to keep mirror copies of complete websites for site management (i.e. checking how new content will fit the existing structure, links and general usability) and backup purposes. Any snapshots likely to be required for extended periods should ideally be in a format that is platform independent.

The intervals will shorten according to the level of risk, the public exposure, how controversial the content is likely to be and how often the content changes. Tracking and sustaining these changes comprehensively across time will pose complex issues and involve substantial costs as noted above.

Transactional databases

Management of structured information in databases, including the challenges posed by technical obsolescence and migration, has historically tended to be better achieved than with other electronic information. This is because of the inherent structure of the resources and the frequent congruence between the immediate operational need and the preservation of the data.

Use of the web to transact a far greater proportion of government business will necessitate more comprehensive audit logging in databases accessed through the user's browser. Experience with accessioning databases of historical interest to the National Digital Archive of Datasets (NDAD) has also raised a number of issues relating to the preservation of database content (especially database technical documentation).

It is likely that further guidance from the Public Record Office will be required in this area.

---

15 XML is a subset of SGML (standard generalised markup language). SGML is an international standard [ISO8879] that will be familiar to many records managers and archivists as the parent language of Encoded Archival Description (EAD).

References

Public Record Office publications and guidance

(available at http://www.pro.gov.uk/recordsmanagement)

e-Government policy framework for electronic records management, 2001

Functional requirements for electronic records management systems, (2 vols.) 1999

Sustainable electronic records: strategies for the maintenance and preservation of electronic records and documents in the transition to 2004, 2001

Good Practice in managing electronic documents using Ofice 97 on a local area network, 2001

Guidance for an inventory of electronic record collections, 2000

Evaluating information assets: appraising the inventory of electronic records, 2001

Management, appraisal and preservation of electronic records, (2nd edition, 2 vols., 1999)

Office of the e-Envoy publications and guidance

(available at http://www.e-envoy.gov.uk/publications and http://www.govtalk.gov.uk)

Guidelines for UK government websites (2nd edition, 2001)

e-government interoperability framework (e-GIF)

e-government metadata framework (e-EMF)

e-government metadata standard, (e-GMS) Version 2, 2001

Other resources

British Standards Institution DISC PD0008: Code of Practice for Legal Admissibility and Evidential Weight of Information Stored Electronically, 2nd Edition, 1999)

National Archives of Australia: Archiving web resources: Guidelines for keeping records of web-based activity in the Commonwealth government, 2001

National Archives of Canada, Office of Government records: Managing internet and intranet information for long term access and accountability: implementation guide. IM forum, 1999

LeFurgy, W. G. *Records and archival management of world wide web sites*. Society of American Archivists, Government Record News, April 2001. http://www.governmentrecordssection.org

McClure, C. R. and Sprehe, J.T.: *Guidelines for electronic records management on state and federal agency websites* at http://istweb.syr.edu/~mcclure/nhprc

World wide web consortium website: http://www.W3c/org

Annex 1: Checklist for implementation of model action plan

Action	Responsibility	Checked	Other issues/references	Suggested time interval (for action)
Initial website risk assessment				
Category 1. site used for information provision only. Any subsequent transactions follow other exchanges of information captured elsewhere (not on website)	Web project manager / webmaster in collaboration with content providers		—	Very frequent (at least bi-monthly) as e-Government agenda is taken forward
Category 2. Information provision is a key corporate function (e.g. public advisory functions)	Operational managers, web project manager / webmaster and records managers		May need internal legal advice on status of disclaimers (may be ineffective)	Very frequent (at least monthly) as e-Government agenda is taken forward
Category 3. Site used to host transactional applications	Operational managers, database administrators, records managers		Extent of audit trails in application itself	Very frequent (at least monthly) as e-Government agenda is taken forward
Existing content audit completed	Records managers		—	—
Addressing issues of non-capture of records	Records managers in consultation with business managers		—	As use of website develops, arising from repeats of risk assessment
Appraisal of content from both browser and directory levels	Record managers		—	As use of website develops, arising from repeats of risk assessment

continues overleaf

Metadata				
1. e-GMS compliant metadata standards agreed for documents and records	Record managers		—	On extension of the e-GMF/ PRO standards
2. e-GMF compliant metadata standards introduced for new content	Records managers in collaboration with web project managers, webmasters, content providers and IT managers		—	—
3. e-GMF compliant metadata standards applied to legacy content	Records managers in collaboration with web project managers, content providers and IT / web managers		—	—
Examine publishing processes				
1. Consider desirability of document management for directory used to publish website documents	Records managers / IT managers / web project managers		Depends on whether records can be captured adequately elsewhere; needs of publishing function / contingency arrangements	—
2. Flag up sustainability issues relating to file formats	Records managers		Particularly important if web content to be sustained for across several platform migrations for business or historical reasons	—
Audit compliance with standards and implementation of scheduling	Records managers, (also webmasters following user feedback)		—	Six-monthly
Repeat risk assessment	—		—	As above

Annex 2: Sample entry on a Departmental retention schedule expressing disposal arrangements for website records

Function	Series	Domain	Trigger 1 ('T1') (removal from site)	Trigger 2 ('T2') (disposal rule)
Operations	Operational guidance / Staff instruction	Internet & intranet	On becoming non-current *(i.e. superseded or withdrawn)*	5 years from becoming non-current (T1 +5 years)
			Disposal action 1	Disposal action 2
			To offline storage	Final deletion
Policy	Major policy consultation	Internet	On change of administration *(in line with OeE web guidelines)*	Transfer to Public Record Office within 1 migration cycle (T1 + max 5 years)
			Disposal action 1	Disposal action 2
			To offline storage	Final deletion 15 years from creation / posting ***if transfer confirmed***
		Intranet	12 months from end of consultation period	12 months from end of consultation period (internet version retained as above)
			Disposal action 1	Disposal action 2
			To offline storage (content will be the same, RM metadata different)	As above *(if selected)*
Facilities	Health and Safety arrangements	Intranet only	On becoming non-current	T1 + 75 years
			Disposal action 1	Disposal action 2
			To offline storage	Final deletion

- *Italicised text is for explanation only.*
- *Please note any retention periods on this page are for illustration only, they should not be construed as disposal scheduling guidance from PRO.*

Frequently asked questions

Q: We've decided we don't have any original records on our website – it's all either copies of something else or purely ephemeral. Is that OK?

A: Early government websites were typically used merely to make available official publications that continued to be produced and distributed in hard copy. Later, even the use of websites for publication purposes developed so it is unusual for a department not to have at least some publications that have only been published on the internet.

Since then, a great deal has changed and there are central guidelines published by the Office of the e-Envoy on (among other things) what minimum content a government website must contain.

The e-government agenda also means that the web is becoming the principal method for delivering electronic services. Records are a necessary by-product of business activity. They capture information about the transaction for the benefit of both parties, confirming understanding and promoting trust. This is particularly important in the electronic environment. This makes it highly unlikely that you do not have records management issues to consider with your website.

Q: Can we manage the records on our website by just printing to paper?

A: Probably not, or at least not for much longer. Please refer to the answer to the previous question. In addition, there are types of electronic content that cannot be accurately represented by a paper surrogate (e.g. interactive features). You should also be preparing to manage your electronic records electronically by 2004 to meet the cross government target.

Q: So this means we're all right with our intranet then?

A: We have scoped this guidance to include intranets and extranets (including the GSI) as well as public websites.

Civil servants within your organisation or otherwise with access to your organisation's information through the GSI will be making business decisions on the basis of the information. For example, they may be viewing staff instructions to assist them with casework. Refer to Section 5 of the toolkit for further discussion of this issue.

In addition, the Freedom of Information Act 2000 and other developments will greatly increase the potential for public access to your information resources previously thought of as purely 'internal'.

Then, many departments make information available to their staff about their employment, health and safety etc. using an intranet. Some of these records may be required in evidence and some have to be retained for a very long time.

Q: What about dynamic content and compound documents?

A: This is one of the principal differences between web based resources and more conventional material. The presentation of content from a variety of different source files in the user's browser is part of the communication revolution currently underway and a vital part of providing services tailored to individual customers.

Managing evidential records in this context is highly complex. In fact, for areas of high risk we are suggesting that this type of presentation should be avoided. Some principles are suggested in this guidance, relating to metadata capture and appraisal issues (sections 4 and 6).

Software obsolescence in formats used for web resources is even more of an issue than with ordinary office software. There are very substantial cost implications in trying to maintain the user's experience across long time periods.

This guidance is intended to help departments with the current situation. A great deal will change over the next few years and it is to be hoped that some of the issues will become a little clearer and some consensus emerge about possible solutions.

Q: We don't have an electronic records management system in place yet – what can we do?

A: Departments are at different stages in their implementation plans, although they have been set common milestones along the route to full EDRM. Many of the issues with managing record on websites can be made a little simpler by the implementation of EDRMS, though only some of the software products currently available have much functionality in this area. Most 'Content management' solutions actually do very little to encourage or enhance the management of records.

If you do not have an EDRMS implemented, you have a problem and one that may be exposing your organisation to an unmanaged risk, depending on what your website is used for. Tighter document management procedures may provide some improvement. Refer to Section 5 on risk assessment in the guidance.

If your organisation has a high risk exposure owing to its web based activity you may well have specific business requirements to consider when you do come to drawing up your EDRMS requirements.

Q: Our website is hosted by an external contractor. How do we manage the records?

A: Many organisations have contracted out their IT services in recent years, including web hosting. You need to ensure that the same controls are in place over your information as would be there if it was still in-house. This may require negotiating more stringent controls over your electronic information and records.

There are other areas of electronic information management where this is equally important: e.g. compliance with the Data Protection Act 1998. This too may have implications if personal data is being manipulated (processed) using a website.

Q: Should we use snapshots to keep a complete record of what our website said and when?

A: Snapshots of websites can indeed be captured – some proprietory software is available that is capable of capturing the number of levels removed from a particular URL and maintaining the linkage between the pages involved. Unfortunately, this is only really suitable for simple sites and begs the question: 'What use are the snapshots are likely to be?'

One purpose is the maintenance of a fully functional site for back-up purposes, as recommended by the Office of the e-Envoy (also to aid the publishing process by providing a 'test environment' for new content). This is entirely separate from the use of snapshots for records management purposes.

Considering that over time the format(s) will become obsolete, migration will be difficult and costly, it is questionable whether snapshots will tell a great deal about the appearance of the site to any one individual user in the past. We would not know, for example, which options they selected or which links they clicked. Additionally, some sites change many times a day and a vast number of snapshots would be required.

Snapshots are likely to have some application in high risk areas where retention for a relatively short period is useful in managing the business risks.

Q: Does the Public Record Office want a snapshot of our website?

A: As they contain public records, web resources need to be considered for selection in the same way that other public records are.

The principles governing selection are outlines in the PRO's Acquisition Policy. This is being implemented thorough Operational Selection Policies (OSPs) covering particular themes or departments. OSPs will indicate which records, including those in websites, we will select for preservation in the PRO.

These will be the subject of consultation with the department concerned and other stakeholders.