

Identifying Roadkill on the Information Superhighway: A Website Appraisal Case Study

Catherine Nicholls and Jon-Paul Williams

Catherine Nicholls is the Archivist, Records Management Program, University of Melbourne. Between 1995 and 2001, Catherine was the Archives Systems Officer for Records and Archives Services, Monash University. Catherine has a Bachelor of Arts in History and a Masters of Information Management and Systems (Honours) degree from Monash University.

Jon-Paul Williams has worked in a variety of roles in Records Services, University of Melbourne since 1998. He is also an experienced video production artist and along with his business partners, runs a small multimedia company. Jon-Paul is currently completing video production and multimedia studies at the Royal Melbourne Institute of Technology.

The University of Melbourne Web Archiving Working Group (WAWG) is the driving force behind developing a web archiving strategy for the University. The case study presented below outlines the initial steps and processes that have been undertaken in this project. Records Management Program (RMP) staff have taken the front seat in driving the development of selection criteria for web archiving. This article shares the key outcomes and findings of this process and as well as discussing future directions for the WAWG.

If advanced technology is truly going to work, then it must bring benefits, without driver or occupants, being conscious of anything unusual.

Peter Robinson, discusses the Mercedes-Benz SL500
in *Wheels Magazine*, November 2001

As we accelerate down the Information Superhighway, there is still a danger that many of the webpages and records created now are not going to be available for accountability and historical purposes in the future. Although many of Australia's government archives programs (including the National Archives of Australia and the Public Record Office Victoria, PROV) have put forward strategies for dealing with electronic records,¹ the actual implementation of these strategies can seem daunting.

The concept of web archiving presents so many challenges; people are often deterred from investigating the issues in too much detail because finding a single effective solution to the problem appears almost impossible. When it comes to web archiving, decisions are influenced by multiple drivers including: what, when, where, how and why?

On top of this, whatever the solution, it needs to be easy to use, and should be integrated seamlessly into the current business practices of the organisation. If the web archiving solutions are too complicated, and the 'drivers' (web creators) and 'passengers' (web users) are required to do too much, then the strategy may not be adopted and the records won't be adequately captured for future evidentiary and accountability purposes.

The purpose of this article is to present a case study on the work of the University of Melbourne Web Archiving Working Group, the current driving force behind the University's web archiving strategy. This article will begin by introducing WAWG and outlining its role and purpose. It will then focus on the process Records Management Program (RMP) staff have been through in developing web archiving selection criteria.

In some respects, selection performs a similar function to a car's engine. Without an engine, the car can't go. Without selection criteria about what to archive, the web archiving strategy could lose direction, purpose and control. The focus of this article will be the construction of the WAWG's engine. The article will then conclude by discussing some of the major outcomes of this process and some possible future directions for WAWG.

WAWG: The driving force for the University of Melbourne's web archiving strategy

During 2002, the Information Division convened a Web Archiving Working Group (WAWG). The purpose of this group is to formalise a website archiving strategy for the University and to produce its final recommendations in a report due at the end of 2002. The group also aims to advise the Web Advisory Group

and the Information Division Executive on a number of issues including policies and other relevant issues, and measures for web archiving now and in the future. The WAWG draws on a number of different areas of expertise across the University, including records managers and archivists (from Records Services and University Archives), librarians (including the Special Collections Librarian), information technology (IT) professionals, web specialists (including the Web Centre Manager), and the University's Metadata Co-ordinator.

The 'what' and 'why' of the engine design

During initial discussions of the WAWG, the main question to arise was: What should we web archive and why? The group was keen to preserve material for cultural and historical purposes, other concerns related to the fate of academic staff pages and the pages of agencies affiliated to the University. The representatives from RMP were also interested in all of the above. However, the key focus of web archiving from the recordkeeping perspective was to safeguard all administrative, legal and archival requirements with respect to University records through efficient selection, retrieval and storage policies and procedures. The recordkeeping perspective of web archiving is also a focus of this article.

The view that webpages should be preserved for evidential purposes was supported during a web archiving literature review.² For example, the Department of Industry and Technology Government of Western Australia states:

Failure to capture full and accurate records of web-based activity will leave agencies exposed to considerable legal, financial and political risk.³

There was also a general consensus among the Australian government archives programs that:

webpages and websites created and published by public authorities, and the electronic transactions carried out by them, are public records ...⁴

The literature also acknowledged that in most cases, webpages and websites containing 'public records' should be treated like any other type of record. The National Archives of Australia (NAA), for example, states that:

... [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.⁵

Although the NAA acknowledges the diversity of web-based resources and addresses the various different challenges they present,⁶ its guidelines also state that the most effective way to capture web-based records is to pursue a 'systematic

approach that is generally applicable to all records, regardless of format'.⁷ In developing draft criteria for the selection of webpages, RMP staff consulted a number of policies and procedures from the NAA and the PROV – including the Victorian Electronic Records Strategy (VERS). However, in the end, the RMP chose to modify its own engine design (ie alter a series of questions used by clients in assessing electronic data for recordkeeping purposes).

Establishing specifications for the engine design

The University of Melbourne RMP combines a number of strategies into policies and procedures⁸ to ensure that 'adequate records are created, captured and managed'⁹ throughout the University. Although the RMP is moving towards being fully compliant with the Australian Standard,¹⁰ a combination of factors, including client needs, the internal administrative climate of the institution and resources, all play a role in determining how and when RMP reviews, redevelops and implements its policies and procedures.

The RMP has a long history of working with its clients to ensure that they are able to meet the recordkeeping requirements of the University environment. RMP is aware that some clients are time poor and may not have a strong recordkeeping background, so a series of questions were developed in the 1990s to assist them in identifying emails and word documents with no long-term evidentiary value. These questions represent the first stage for clients in assessing their records, and are designed to work with other appraisal methodologies employed by RMP staff to manage University records. NAA policies for web archiving state that all:

Commonwealth records, including the web-based records described [in the policy] be captured into recordkeeping systems that comply with the Australian Standard AS 4390-1996, *Records Management*.¹¹

In simplistic terms, the NAA policies, AS 4390 and supporting documentation, all insist on a function-based approach towards recordkeeping. The RMP staff were also aware that the series of questions represented a records-based approach, not a function-based approach towards recordkeeping. As a first stage tool, the questions have always been effective in assisting clients to determine the evidentiary value of their emails and word documents. Also, when working with other types of professionals on WAWG (including IT experts and librarians), it was important to keep recordkeeping theory accessible to non-recordkeeping professionals. This did not imply that the RMP needed to 'dumb down' its methodologies, but it did mean that RMP staff needed to communicate the recordkeeping needs of the project in a way that enabled others in the group to understand the processes RMP staff were presenting and exploring.

Although the NAA (for example) stated that web records were the same as any other type of record (despite their media) and should be treated as such, there were not enough practical examples or case studies available to demonstrate this. If a simple records-based set of questions was useful and accessible for clients to use on emails and word documents, there was no real evidence from a University environment to suggest that a similar approach to web records wouldn't at least be worth trying.

The modified questions were therefore used to establish criteria for selecting webpages that contained records with long and short-term evidentiary value. It should be noted that at this point in time, RMP staff were only assessing static webpages.¹²

The draft criteria for selecting webpages for long and short-term retention were as follows:

Essential Recordkeeping criteria (long-term preservation)

1. *Is the content of the webpage published in another format (eg paper)?*

• *If answer is YES - then check to see if the content between the two is significantly different. If the content is the same, then retain the paper copy only. If the content is different then check against the criteria listed below.*

• *If the answer is NO - then check it against the criteria listed below.*

Does the webpage:

2. *Publish a change in policy?*

3. *Create a publishing precedent for the University (ie is this the first time the material on the site has been published on the web)?*

4. *Represent a substantive business of the work unit, section or University?*

5. *Publish legal advice?*

6. *Publish information involving negotiations on behalf of the University?*

7. *Does the webpage transmit formal communication(s) between officers?*

8. *Has the webpage been used to initiate, continue or complete a departmental activity/transaction?*

9. *Does the webpage have continuing value for others in the work unit?*
10. *Does anyone external to the work unit need to be aware of, or refer to, this webpage for evidentiary purposes now or in the future?*
- *If answer is YES to any of the above, retain for long-term preservation*

Essential Recordkeeping criteria (short-term preservation)

Does the webpage:

11. *Duplicate other webpages (either internal or external to the University) which carry information on the matter presented in the webpage (in the same form and content)?*
12. *Contain informational value only, eg an information copy?*
13. *Comprise of draft information leading to a final version?*
14. *Contain a message of short-term facilitative value (eg arranging a meeting)?*
- *If answer is YES to any of the above, then the page may not be suitable for web archiving or the page may need to be archived for a short-term period.*

It should be pointed out that within the University environment, most departmental websites follow a similar format. They all contain an index or home page which often contains (or links to) information about what the department is, where it is located and who to contact. The remaining content of the site depends on the function of the department, but can include policy information, links to other sites, course handouts, and so on. Therefore, it was anticipated that the application of the above selection criteria against a small percentage of websites would provide some indication as to whether the criteria would be effective in identifying records contained within the University website.

A number of pages from the University's Communications and Marketing Office, and a small percentage of pages from the International Programs Office were selected to test the criteria. These websites were selected on the basis that the Development Division (which these departments belong to) recently underwent a restructure. This restructure may result in a future web redevelopment project and consequently the Division's websites were flagged by senior University administrators for web archiving purposes.

RMP staff then bolted the test engine into the WAWG chassis and headed off towards the Information Superhighway to give the engine (ie the test criteria) an

intensive test drive. A total of 21 html static pages were assessed in the case study. The following analysis presents the outcome of that test drive.¹³

Too much torque not enough action

In order to understand the results of the test drive, each of the criteria questions and the outcomes obtained from using them will be discussed.

1. *Is the content of the webpage published in another format (eg paper)?²*

In instances where the webpage existed in another format (eg paper) and there was no substantial difference between the two versions, the webpage version would be labelled a 'copy' and the paper version would be the 'master' copy. Although this was a useful first step, this question didn't identify whether a webpage had evidentiary value or not. It was anticipated that the evidentiary value of the webpages would be determined once they had been checked against the remaining criteria.

2. *Does the webpage publish a change in policy?²*

Unless the information on the page indicated that it was 'publishing a change in policy' it was difficult to know whether a change had occurred or not. Although a change in policy on a website could have long-term evidentiary value, the focus of this question should have been broader. For example, it may have been more relevant to know whether the webpage content was representative of a 'unique instance' and if it was, its importance should have been identified. Or if it weren't representative of a 'unique instance' then it would have been necessary to identify whether the webpage had business importance in its own right.¹⁴

3. *Does the webpage create a publishing precedent for the University (ie is this the first time the material on the site has been published on the web)?*

This question was too ambiguous to be used in the web environment, and did not help to identify webpages that may have had evidentiary value (short or long-term). It implied that *all* pages, when loaded for the first time, had value because they had been published. The question was poorly worded and may have worked better had it focused on whether this was the first time the material had been published *in any format* (rather than focusing on whether this was the first time it was published on the web).

Question 3 may have provided more relevant information if it had focused on finding out whether the webpage was dynamic or static, whether it is being published on the Internet or the intranet and who the target audience was.

4. *Does the webpage represent a substantive business of the work unit, section or University?*

The webpages surveyed in the test analysis were all related to the business of their respective work units. It is probable, that in regard to emails and word documents, Question 4 may have been useful in distinguishing between a 'personal' email and an email that was related to the business of the work unit. Although some sectors of the University website (eg student and staff webpages) may contain 'personal' (or non-work related) material, it would be assumed by RMP staff that the vast majority of webpages produced by the University would be related to the business of the University.

5. *Does the webpage publish legal advice?*
6. *Does the webpage publish information involving negotiations on behalf of the University?*
7. *Does the webpage transmit formal communication(s) between officers?*
8. *Has the webpage been used to initiate, continue or complete a departmental activity/transaction?*

Questions 5 to 8 were not particularly useful in identifying webpages that contained long-term evidentiary value either. Although these questions worked well in appraising records created in emails and word processing documents, they became almost redundant when applied against the University's webpages. These questions are not representative of the types of webpages produced by the University.

9. *Does the webpage have continuing value for others in the work unit?*
10. *Does anyone external to the work unit need to be aware of, or refer to, this webpage for evidentiary purposes now or in the future?*

Questions 9 and 10 alluded to risk analysis. However, the wording of these questions made the analysis subjective. Ideally, a 'risk analysis' type of appraisal would need to be executed in a more formal, systematic way if the results gathered from it were to be more consistent.

11. *Does the webpage duplicate other webpages (either internal or external to the University) which carry information on the matter presented in the webpage (in the same form and content)?*
12. *Does the webpage contain informational value only, eg an information copy?*

Question 11 was not useful in the web environment. In the test case, it was more likely that webpages would be hyperlinked, rather than page content being repeated throughout various webpages. However, if a webpage duplicated the content of another one, then the duplicated page would have no long-term evidentiary value. Question 12 asked if the webpage was an information copy. In most instances, pages that contained 'about us' type of information and contact details (eg the Communications and Marketing Office - Web Centre index page) may have been categorised as 'information value' only. Although the information presented on the page may not exist in the same format anywhere else, it could be argued that the content does not carry any long-term evidentiary value. Much of the content simply repeats information (ie what the Web Centre does and who the staff are) that could be found elsewhere in other formats, eg email and phone contact details about staff can be found on the 'University of Melbourne - Contacting People' website.

The International Programs Office home page initially appeared to meet the criteria of short-term 'informational' value. The content of the site (which includes details about what the office does, contact details, etc) is available in other formats as mentioned above. However, when the target audience (mainly overseas students) and core functions of the office (which includes supporting operations to recruit international students) are taken into consideration, the page's long-term evidentiary value becomes apparent. As noted by the Public Record Office, United Kingdom:

if webpages are what the public, user etc will refer to when making decisions or deciding to travel abroad, then both they and the organization making the information available have a continuing stake in it and there is a business record present.¹⁵

As demonstrated in the case study, a webpage may initially meet the criteria of short-term informational value but further investigation may reveal that it also has long-term evidentiary value. Therefore, a question that only focuses on the 'informational' value of the webpage may produce misleading results.

13. *Does the webpage comprise draft information leading to a final version?*
14. *Does the webpage contain a message of short-term facilitative value (eg arranging a meeting)?*

Questions 13 and 14 were not relevant to the webpages being assessed against the criteria because they related to drafts and records of short-term facilitative value. The terms used in these questions were not well suited to the web environment. However, these questions are more useful in identifying records of short-term evidentiary value in email and word processing documents.

Overall, the case study demonstrated that modifying a series of questions used by clients to assess their emails and word documents did not work effectively when applied against webpages. In other words, the reconditioned RMP engine did not survive its test drive on the Information Superhighway. However, the exercise was not a waste of time, as it reinforced the need for RMP staff to focus on the issues involved in appraising all of its records. In particular, it demonstrated that in some instances a records-based approach to appraising webpages was relevant (in particular, identifying whether the webpage was a copy of a document that exists in another format).

Yet it also highlighted the point that a total records-based approach was not useful in identifying the evidentiary value of webpages either, because overall, the questions were too specific and rarely relevant to the type of information presented in the webpages. The case study results were presented to the WAWG to demonstrate that a simple records-based approach was not going to provide any useful results in the appraisal process. This opened the way for RMP to investigate other approaches with the full support and interest of WAWG.

Once it was established that the draft selection criteria could not be taken forward, further research was undertaken into the issue of appraising webpages for web archiving purposes. A document produced by the Public Record Office, UK, made some interesting distinctions between webpages and other electronic records and noted that a 'website could include a combination of':

- static pages
- compound pages formed by displaying together content from a variety of sources
- dynamic pages formed according to the user's expressed preference from a variety of sources
- active server pages
- web forms capturing information for processing in a separate database application
- graphics, audio files, video clips and virtual reality
- linked documents such as publications.¹⁶

The document also acknowledged that many of the above components were:

very different in nature from the traditional image of a record; in so much so that it can tend to give the impression that no records are present. This can be highly misleading.¹⁷

Therefore, the RMP slammed the WAWG vehicle into reverse and took it off the Information Superhighway. The draft selection criteria was abandoned, and RMP staff decided that although the concept of appraising websites based on current recordkeeping practices was relevant, the chosen method for selecting University of Melbourne webpages for web archiving would need to be reviewed.

Maintain the chassis but overhaul the engine: Applying functional analysis

Although RMP staff had decided to haul out the WAWG engine #1 (ie the draft selection criteria), there were plenty of other possibilities available to investigate. For example, RMP staff had also been exploring other parts of the web archiving process, including metadata and the ability to capture webpages using electronic recordkeeping systems already in use within the University. It was anticipated that it could be possible to track a webpage (both the ones in current use and the ones eventually selected for web archiving) through a specific functional metatag which would be linked to a disposal sentence. In an approach that is yet to be tested, a webpage could be assigned a specific functional metatag, which would then be used as part of the trigger to capture the pages for web archiving purposes. Within the context of the University of Melbourne,¹⁸ metatags are:

pieces of information that are stored inside [the] HTML document. They provide useful information about the document which will aid The University of Melbourne server to correctly index documents. This will make the documents easier to find when people search for information using the search engine.¹⁹

The University's Information Strategy Advisory Committee (ISAC) has also set up an expert committee, the Metadata Working Group (MWG), to advise on the implementation of a uniform approach to the creation of metadata across the University's web sites. The University has appointed a Metadata Coordinator to oversee the implementation of the project. At present the committee has proposed a metadata standard consisting of 14 Dublin Core (DC) fields²⁰ with some University of Melbourne administrative fields.

DC fields were chosen because they followed an international standard and were easy to use. The metadata standard is only being devised for webpages at the present time. The University Metadata Co-ordinator has invited input from RMP staff who have reviewed the proposed metadata elements. It was during the process of reviewing the proposed University metadata standards that RMP staff decided to investigate the concept of applying a functional descriptor to the webpage at its creation point. Ideally, if a disposal sentence were attached to a functional descriptor

which was included in the metadata at the creation stage, this would streamline the appraisal process and maybe even open up the possibility of automating some of the web archiving processes. For example, if webpages were assigned a 'trigger' point at the creation stage (via a functional descriptor in the metadata tags), then pages suitable for short- or long-term web archiving could be searched for and selected automatically. However, this idea needs further thought and investigation before it can be implemented.

It was the concept of applying a functional descriptor at the creation stage of the webpage process that led RMP staff to consider a functional analysis approach to webpage appraisal. RMP staff acknowledge that the concept of applying disposal sentences at the creation stage of a record, and the idea of utilising a function-based analysis are not new. According to DIRKS, the basic premise of functional analysis is to:

identify and document the role of [the organisation], its structure, the business, regulatory and sociopolitical environments in which it operates, and major factors affecting its recordkeeping practices ...²¹

Implementing a functional approach for the webpage appraisal seemed a logical approach for a number of reasons. The University of Melbourne's Central Records Department already has a functional thesaurus in place.²² It is anticipated that this thesaurus could also be utilised in the appraisal of University of Melbourne websites. This process would involve assigning appropriate functional terms (as determined by the thesaurus) to the metatags of webpages, at the point of creation.

However, it is more likely that webpages initially identified for archiving will be assigned a functional descriptor at the appraisal stage. This method would perhaps be used in combination with some records-based appraisal. For example, identifying whether the content of the webpage has been presented in another format may be the first step in the appraisal process.

RMP staff have already discussed the possibility of including some AGLS²³ descriptive elements (in particular, 'function') into the University of Melbourne's metadata schema. Another feature of developing and using a functional approach towards records appraisal and classification is that it can include a risk analysis component. According to the DIRKS manual the 'risk connected to a function or activity is the risk of events occurring that expose the organisation to adverse consequences ...'.²⁴ DIRKS therefore recommends that:

Assessing risk at the functional level assists in prioritising areas for future analysis and identifying organisational areas that perform the function as requiring more stringent recordkeeping practices and training.²⁵

According to the Public Record Office, UK, conducting a risk assessment of the website and its function is very important and should take into account the business transactions carried out on the site and the interests of all 'stakeholders' who have information presented on the site.²⁶

The next phase of the RMP contribution to the construction of the WAWG vehicle will be to fully investigate and later report on the suitability of using the Central Records function-based thesaurus for appraising University of Melbourne webpages. It is anticipated that the thesaurus will provide a set of functional descriptors that will be assigned to the webpages (via the metadata) during the webpages' construction.

Each functional descriptor will then be linked to a 'risk analysis' rating.²⁷ The functional descriptor could then be used as a 'trigger' point for identifying pages that needed to be retained for short or long-term evidentiary purposes. The functional descriptor could also be used as a means of identifying and locating the webpages once they have reached the 'web archive'. Overall, the functional approach should provide the WAWG vehicle with a more focused, effective and efficient web archiving solution. In motoring terms, a functional approach should give the WAWG vehicle more 'grunt', with enhanced control, handling and reliability.

The final lap

The work that has been completed on developing web archiving selection criteria for the University of Melbourne indicates that the structure and content of webpages can present challenges that may not necessarily be covered by an organisation's current recordkeeping policies and practices.

RMP has learnt that when it comes to appraising webpages, there are a number of key issues to address, including finding out whether the website content is representative of a 'unique instance' and if it is, identifying its importance. If it isn't representative of a 'unique instance', it will be necessary to identify whether the webpage has business importance in its own right. The business importance may be uncovered through an understanding of the function of the agency producing or authoring the page. Other important factors to consider when appraising webpages may include the target audience, whether the page is static or dynamic, available on the Internet or intranet and whether there are long- and short-term risks associated with not 'web archiving' the page for a set term. Initial investigations seem to imply that a combination of function and records-based appraisal methods may be a practical solution for the current University environment.

Another outcome of the test case was the confirmation that a successful web archiving strategy would only be possible if the organisation implements sound recordkeeping procedures and web publishing policies. The recordkeeping and web publishing policies need to work together with the technology in order to produce a useful web archiving solution. For example, there is already a strong indication that correctly completed metadata tags could streamline (or even automate) some aspects of the web archiving process.

However, the process is not just about finding a technological solution that automatically 'archives' everything. The solution is more about finding a balance between the needs of the various stakeholders involved in the project (be they recordkeepers, librarians, historians or information technologists), and then assessing these needs against what the technology can actually provide. If the needs of the stakeholders can best be met (initially at least) by manually appraising the webpages, then an automated approach may not be essential in the early stages of the project.

Therefore, whatever the web archiving solution is, it should be integrated as seamlessly as possible into the current business practices of the organisation. The web archiving solutions must be relevant to the organisation and the staff who are going to have to implement them. The web archiving solution must also ensure that webpages of the past and future can be accessed for accountability and evidentiary purposes. To return to the car analogy, if the WAWG vehicle is going to be able to cruise through fast corners and handle the hairpin turns of the Information Superhighway, then its design and structure must be efficient, effective and supremely easy. The journey continues ...

ENDNOTES

¹ For example, the National Archives of Australia's '*DIRKS methodology is an eight-step process which agencies can use to improve recordkeeping and information management practices ...*' at www.naa.gov.au/recordkeeping/dirks/summary.html. The Public Record Office Victoria's Victorian Electronic Records Strategy (VERS) is 'a framework of standards, guidance and implementation projects which is centred around the goal of reliably and authentically archiving electronic records created or managed by the Victorian government at www.prov.vic.gov.au/vers/welcome.htm'.

² The literature review involved analysing a number of government (local and one international) web archiving policies and procedural documentation including the National Archives of Australia's 'Archiving Web Resources: Guidelines for Keeping Records of Web-based Activity in the Commonwealth Government', March 2001, at www.naa.gov.au/recordkeeping/er/web_records/guide_contents.html, and 'Archiving Web Resources: A Policy for Keeping Records of Web-based Activity in the Commonwealth Government', at www.naa.gov.au/recordkeeping/er/web_records/policy_contents.html (revised January

2001); Queensland State Archives, 'Managing Records of Webpages and Websites: Policy Statement, Principles and Guidelines', version V1.00.00, April 2002, at www.archives.qld.gov.au/index_govserv.html; Department of Industry and Technology, Government of Western Australia, 'Guidelines for State Government Websites: Draft for Comment - Information and Communications Policy Directorate', 18/02/2002, version 3, doc. no. 120605, at www.indtech.wa.gov.au/govt/polguides/websites; and the Public Record Office, United Kingdom, 'Managing Web Resources: Management of Electronic Records on Websites and Intranets: An ERM Toolkit', version 1.0, December 2001, at www.pro.gov.uk/recordsmanagement/eros/website_toolkit.pdf. Other sources included the Australian Standard AS 4390-1996, *Records Management*, and International Standard ISO 15489, *Records Management*.

³ Department of Industry and Technology, Government of Western Australia, 'Guidelines for State Government Websites', p. 37.

⁴ Queensland State Archives, 'Managing Records of Webpages and Websites', p. 3.

⁵ National Archives of Australia, 'Archiving Web Resources: Guidelines', p. 13.

⁶ *ibid.*, p. 11.

⁷ *ibid.*, p. 13.

⁸ Many of these guidelines can be found online at www.unimelb.edu.au/unisec/recordsm.htm.

⁹ Terminology used here draws on Standards Australia, AS ISO 154489.1-2002 Australian Standard, *Records Management*, Part 1: General [ISO title: Information and Documentation - Records Management - Part 1: General], Standards Australia International Ltd, Sydney, 2002, p. 1.

¹⁰ *ibid.*

¹¹ National Archives of Australia, 'Archiving Web Resources: Policy', p. 5.

¹² The WAWG decided to initially focus the strategy on static webpages because they represented the majority of webpages currently online in the University environment. According to the National Archives of Australia's 'Web Resources Guidelines', static websites 'may be nothing more than a collection of static documents sitting in folders on a server and tied together with hyperlinks'. Dynamic websites are generally built 'on the fly' and the 'component parts of each individual page - its content, structure and presentation - are generated dynamically using a combination of databases and style sheets', pp. 11-12.

¹³ RMP staff acknowledge that this does not represent a substantial amount of testing, but early results indicated that it was pointless going any further with the initial selection criteria.

¹⁴ Public Record Office, UK, 'Managing Web Resources', p. 13.

¹⁵ *ibid.*, p. 8.

¹⁶ *ibid.*, pp. 6-7.

¹⁷ *ibid.*, p. 7.

¹⁸ Communications and Marketing Office, University of Melbourne, 'World Wide Web Publishing Policies and Guidelines', 2001, at www.unimelb.edu.au/guidelines/web-guidelines.pdf.

¹⁹ *ibid.*, p. 29.

²⁰ The Dublin Core Metadata Initiative is an open forum engaged in the development of inter-operable online metadata standards that support a broad range of purposes and business models, at dublincore.org.

²¹ National Archives of Australia, 'DIRKS: A Strategic Approach to Managing Business Information' at www.naa.gov.au/recordkeeping/dirks/summary.html

²² A system of keyword titling was introduced to Central Records in 1993. Keyword titling is designed to facilitate subject information retrieval through the use of keywords and descriptors in file titles. The file titling has three components:

1. Primary keyword representing the major functions and activities of the university, eg 'students'.
2. Descriptor further describing the primary keyword, ie leading the searcher to a more specific aspect of the subject which the keyword represents, eg 'enrolment' is a descriptor for the keyword 'students'.
3. Free text for the remainder of the file title, although terms used are to a certain extent controlled by the Thesaurus. We are guided by the Thesaurus to use 'preferred terms' in preference to 'non-preferred' terms. At www.unimelb.edu.au/unisec/keyintro.htm.

²³ National Archives of Australia, AGLS website, at www.naa.gov.au/recordkeeping/gov_online/agls/summary.html, explains that the NAA is the maintenance agency for the Australian Government Locator Service (AGLS) metadata standard.

²⁴ National Archives of Australia, 'DIRKS: A Strategic Approach to Managing Business Information'.

²⁵ *ibid.*

²⁶ Public Record Office, UK, 'Managing Web Resources', p. 21.

²⁷ Which may, for example, be based on the risk register outlined in the DIRKS Manual at www.naa.gov.au/recordkeeping/dirks/dirksman/dirks_A11_risk.html