

Managing URLs

Version:	1.0
Date created:	01/08/2008
Policy official:	David Pullinger
Date last updated:	25/03/2009
Date issued:	March 2009
Lead official:	Amanda Spencer
Guidance number:	TG125

This document sets out best practice for managing online content to ensure that website links are used effectively and that broken links are minimized on UK public sector websites. This document also provides technical guidance on the installation of redirection software provided by The National Archives for use on UK central government websites.

Table of contents

Introduction	2
Purpose	2
A solution to broken links on UK central government websites.....	2
Audience	2
Overview	3
Definition of terms	3
The Standard	5
Maintaining domains	6
Persistent URLs	7
Meaningful URLs.....	8
Managing URLs through change.....	9
Technical change	9
Machinery of Government change	9
Website closure.....	10
Website Convergence/Website Rationalisation	11
Communicating URLs.....	12
Implementing redirection behaviour	13
Displaying a standard departmental 404 error page	14
Displaying a customised error page.....	14
Redirect to another URL	14
Redirect to the UK Government Web Archive.....	15
Redirection Flowchart	17

Introduction

Purpose

Users find broken links on websites incredibly frustrating. URL links that once worked and no longer do so result in 'page not found' errors (Error 404). Whilst it may be unrealistic to maintain all links, good website management can do much to reduce the occurrence of broken links. The purpose of this guidance is to prevent users of government websites from experiencing broken links. This issue was raised by ministers in 2007 and Government is committed to implementing the solution.

A solution to broken links on UK central government websites

By taking four actions, almost all broken links on UK central government websites can be removed:

- i. Maintain all domains
- ii. Good URL management
- iii. Enable The National Archives (TNA) to archive all old content
- iv. Put in place redirection to content that has been moved or archived.

This Guidance explains what needs to be done for i, ii and iv. Archiving is covered in [Archiving websites \(TG105\)](#).

Audience

This document is primarily aimed at UK central government **Website Managers**, **Content Providers** and **Publishing Managers** who have responsibility for managing online content.

It should also provide informative reading for UK central government **Web Convergence Managers** who have responsibility for managing changes to the location of online content through the Web Convergence programme.

It may also be of interest to UK central government **Heads of e-Communications** and **IT Managers** in charge of strategy and deployment, who should note [the standard](#) and ensure that it is adhered to within their organisations.

UK central government **Departmental Records Officers** should find the document informative from an Information Management perspective.

Overview

The guidance on [communicating URLs](#) is relevant to all UK central government employees and civil servants, but will be of particular interest to **Government Communicators, Publishing Managers** and **Parliamentary Offices**.

Web Convergence Managers will find the guidance on [managing URLs through change](#) of particular interest.

UK Central Government **Website Managers, Content Providers** and **technical experts** should read the entire document.

Definition of terms

The following terms are used throughout this document:

- **URL (Uniform Resource Locator):** specifies where an identified resource is available and the mechanism (protocol) for retrieving it - popularly referred to as a web address
- **Persistent URL:** a URL that continues to lead to content over time; often does not directly describe the location of the content but instead an intermediate (more persistent) location which results in redirection to the current location of the content
- **Meaningful URL:** a typically short URL with a "human-readable" format which is easier to understand or remember than one made up of "machine-readable" digits or random characters
- **Domain:** a typically "human-readable" name for an IP address (see below), e.g. [www.domain.gov.uk](#). When translated to the IP address, it enables a computer or website to be found on the Internet
- **Redirect(ion):** a server telling a client (e.g. a browser) to take additional action to complete the request
- **301 redirect(ion):** "Object has moved permanently" - this and all future requests should be directed to the given URL; search engines should transfer ranking to the new location
- **302 redirect(ion):** "Object has (usually temporarily) moved" - the response to the request can be found under another URL; search engines continue to rank the existing location
- **Subdomain:** a subsidiary domain (e.g. [www.subdomain.domain.gov.uk](#)) that is relatively (but not absolutely) dependent on a larger domain; commonly used to assign a unique name to a particular department, function, or service related to a large organisation
- **IP address:** an Internet Protocol (IP) address is a numerical identification (a logical address) or the "machine-readable" name for a computer (e.g. 123.456.789.0) which allows that computer to be found on the internet
- **URL rewriting:** modification of a URL's appearance (usually to a shorter or more memorable form) which "separates" the underlying technology used to generate a web page from the URL that is presented to the world

- **Website closure:** the point at which the content on a website is no longer actively maintained, either because the content is no longer relevant or because it has been transferred elsewhere - but this does not necessarily mean the website is no longer accessible on the internet
- **DNS:** the Domain Name System which gives all computers on the internet a hierarchically positioned name and translates domains and URLs to IP addresses to allow those computers to be found
- **Resolution:** the process of referring a request for website content over the Internet to the location where that content can be found and returned - may involve one or more of the techniques described above, eg. DNS, URL rewriting, redirection or persistent links

The Standard

1. Every central government organisation must maintain all their registered web domains in perpetuity. This is essential to ensure that users can be redirected to content if it is moved or archived. Given the benefit of ensuring the long-term findability of the information, the cost of persisting domains is relatively low.
2. Every central government organisation must [use persistent URLs](#).
3. Every central government organisation must [use meaningful URLs](#).
4. Every central government organisation must [manage URLs through change](#):
 - technical change
 - Machinery of Government change
 - Website Convergence
 - website closure
5. Every central government organisation must [communicate their URLs](#) effectively.
6. In addition, every central government organisation must implement [redirection behaviour](#) on their websites.
7. Central Government departments must ensure that redirection is fully implemented by 1 October 2009.
8. Central government executive agencies and non-departmental public bodies must ensure that redirection is fully implemented by 1 October 2010.

Maintaining domains

9. Maintain all registered domains in perpetuity. Maintaining domains supports persistence of links to content which have been moved or archived (see [Managing URLs through change](#)). The cost of doing this is minimal.
10. Maintaining domains
 - allows users to be redirected to alternative content in an archived or live site;
 - prevents domain names that were previously in official use from being purchased and registered by someone else;
 - prevents new and questionable, or potentially defamatory, content from being published on these domains by the new owner. (Once this happens, the original domain owner has no control over new content.)
11. Maintaining domains is not necessary for campaign websites, where it is acceptable to close the domain. A campaign website or microsite is a site developed as part of a comprehensive advertising plan centred on a specific idea or theme of promotional activity to raise awareness of an issue or highlight an initiative or service.
12. It may be appropriate to capture campaign sites in the UK Government Web Archive, so that a record of such sites exists. For further advice, see [Archiving websites \(TG105\)](#).
13. Government organisations should use freely available webmaster tools to identify the external links to content on their websites. This is a good way of identifying those areas of persistent interest which, in turn, may require special handling to ensure that the content in question is persistently found (for example, via redirection or URL rewriting).
14. Website owners should ensure that if a website is referred to in an official publication (e.g. Hansard) the web domain is maintained so that these links persist. To identify websites referenced in Hansard try searching for the website domain using the “site” function in Google. In the Google search box type:

Site:http://www.publications.parliament.uk/pa/cm “department.gov.uk”

(Note the space between cm and “.)

Persistent URLs

15. Use persistent URLs. Using persistent URLs ensures that documents and pages cited or referred to by users of the site remain accessible. Resources and documents on websites are often linked or referenced from external sites. Persistent links assist in providing continuous access to these resources. For example, <http://www.department.gov.uk/documents/annual-report-2008.pdf> should persist.
16. Where locations of documents change, Website Managers should provide a redirect to the new location of the content, which would allow links to legacy information to remain, rather than leading to an error page. Depending on the circumstances, such redirection might be to another area on the site, another site, or the web archive. In most situations where material is moved *en masse*, it is not necessary to provide redirection rules for individual URLs, since modern software for URL rewriting provides rich pattern matching capabilities.
17. Do not change the location of documents unnecessarily, because you will then need to ensure that the content in question can be persistently found (for example, via redirection or URL rewriting).
18. A persistent URL may describe an intermediate location which results in redirection to the current location of the resource. For example, <http://www.coi.gov.uk/webguidelines> redirects to <http://www.coi.gov.uk/guidance.php?page=188>

Meaningful URLs

19. Use meaningful URLs. URLs should be meaningful so that they are easily understood by a user. Meaningful or 'human-readable' URLs are good practice for a number of reasons, including usability, security, and search engine optimisation.
20. URLs should be unrelated to any machine location. This means that the URL should be based on understandable text, rather than numbers (such as an IP address) or machine-readable information. For example, '<http://www.department.gov.uk/about-us.html>' is human-readable. We can understand how the address is put together and that this would lead us to the 'About us' section of a website. However, 'http://www.department.gov.uk/1/2/index.jsp?nc=123456&refPg=%2fhome.jsp&hp=-789&CM_REF=12345' is not human-readable as it consists of a string of text and numbers rather than a clear reference to the relevant section of the website.
21. URLs should not be dependant on their underlying technology. URL rewriting can be used to achieve meaningful URLs. Meaningful URLs support changes in technology, directory location and format. By abstracting the information resource from the technology that is used to provide it, the technology can be changed as required without breaking any links. For example
 - the underlying technology may change from ASP.NET to PHP;
 - the document could move to another directory on the server;
 - the format of a document could change from Portable Document Format (PDF) to OpenDocument Format (ODF).
22. Website applications should support the rewriting of URLs to human-readable configuration. Many Content Management Systems (CMSs) support URL rewriting. URL rewriting engines are available for all major web servers. When procuring CMSs, website managers should specify support for publication of static versions of URLs, or URL rewriting, to ensure meaningful URLs.
23. URLs should be kept as short as possible. However, link shortening services such as Tiny URL¹ should be avoided as these URLs are too short to provide a useful description of the content to the user. It is preferable to rewrite URLs to achieve a more meaningful user-friendly version.

¹ Tiny URL <http://tinyurl.com/>

Managing URLs through change

Technical change

24. Manage URLs through technical change. As explained in [Meaningful URLs](#), URLs should not be dependant on their underlying technology. If they are dependent on underlying technology, steps should be taken to ensure that the original links persist through any technical change.
25. In the event that a new website is procured, steps should be taken to ensure that the links on the original website persist.

Machinery of Government change

26. Manage URLs though Machinery of Government change. A Machinery of Government change is where the functions of a government organisation move either to another existing organisation or a newly created organisation, or cease altogether. The new owner of the function is responsible for the maintenance of any legacy web domains which relate to that function, (see [Maintaining Domains](#)) and for the appropriate management of any content served from that domain.
27. If the content hosted on the legacy domain is to be moved to the new domain, appropriate redirects should be put in place. If there are no plans to move the content elsewhere (e.g. it is out of date or no longer relevant), the content can be removed from the website, provided that an appropriate redirect to an archived copy is in place.
28. For UK central government organisations, TNA retain copies in the [UK Government Web Archive](#)², and provide guidance and software to ensure redirection to the Web Archive when no other redirects are in place. This ensures the best possible user experience. More information on [implementing redirection behaviour](#) is available in this guidance.

² www.nationalarchives.gov.uk/webarchive

Website closure

29. Manage URLs through website closure. A website may be closed and all of its content removed but the domain should be left running to resolve incoming requests. Wherever possible, such requests should be forwarded to the relevant content URLs on the new site, and not to general landing or "catch all" pages.
30. If some of the content is to be moved elsewhere, consider putting appropriate redirects in place. The [Redirection flowchart](#) may help.
31. If the content will not be moved elsewhere, it should be archived. If an archive exists, appropriate redirects should be put in place.
32. As an interim measure, before appropriate redirection is in place website managers should ensure that a 'closed' site is clearly signposted as such. This should take the form of a message on the home page in a prominent place with wording such as, 'This website is no longer being updated and the URL is maintained purely for archive purposes....'
33. Once appropriate redirection is in place the domain still needs to be maintained. The cost of this is minimal.
34. Keep old domains registered with the relevant certifying authority (JANET in the case of .gov.uk domains). This does not include DNS services which resolve the domain name to an IP address. In order to serve content or redirect users elsewhere, it is necessary maintain the appropriate DNS settings so that users can be directed to the correct IP address for the domain. The IP address is assigned to the department (or their supplier) by the hosting provider.
35. Where multiple sites are involved, the same IP address may be used for several different websites or there may be more than one IP address to resolve different domain names. The web environment and redirection tools can then be configured to allow requests for resources on the old domain to be redirected to the appropriate location on the new one, without the need to maintain any content on the old domain.
36. There should also be no need to maintain additional hardware for old domains because virtual hosting can be used. Whilst the initial setup may be complex, once this is in place, the ongoing overheads of maintaining this should be minimal.

37. The National Archives retains archival copies in the [UK Government Web Archive](#)³, and provides guidance and software to ensure redirection to the Web Archive when no other redirects are in place. This ensures the best possible user experience. For more information see [Implementing redirection behaviour](#).

Website Convergence/Website Rationalisation

38. Websites are being restructured because of the Website Convergence programme. You will need to consider how to manage the URLs of the affected content. The [Redirection flowchart](#) is designed to help you work out your content management requirements.
39. Highly linked-to pages with high page rankings, should be valued. If the content on those pages moves, organisations should put in place appropriate page level re-directs to the right new home for the content (e.g. to DirectGov). For example, we strongly encourage Departments to consider using 301 redirection at page level for key transactional services and for content being moved to DirectGov and BusinessLink as part of Website Convergence.

³ www.nationalarchives.gov.uk/webarchive

Communicating URLs

40. The preferred case is always to use the **domain/top-level-directory** for example www.direct.gov.uk/actonCO2
41. It is acceptable and some organisations may find it preferable to drop the www for marketing purposes (e.g. direct.gov.uk). If so, this should work when entered into a browser address box ('resolve').
42. However, the domain name with www in front (e.g. www.direct.gov.uk) must work as an address. The form with www should be used as the preferred Unique Resource Identifier and cited as links in official documents.
43. When citing a URL for inclusion in any document or publication, you should check that it is
 - a valid URL
 - still working
44. Sub-domain URLs should never be marketed or advertised.

See [Naming and registering website \(TG101\)](#)⁴ for more information on communicating URLs.

⁴ Naming and registering websites <http://www.coi.gov.uk/guidance.php?page=191>

Implementing redirection behaviour

45. The National Archives is making available components and documentation which enable redirection to the web archive to prevent a user receiving a 404 “page not found” error message. The redirection components and documentation to support installation can be obtained by [contacting The National Archives](#).⁵
46. The redirection software offered by The National Archives to UK Central Government allows Website Managers considerable control over how redirection is handled. These components are based around a web programming technique known as URL rewriting, and use the Perl Compatible Regular Expression (PCRE) syntax. The components can handle not only redirection to the Web Archive, but also other redirections and general rewriting. The PCRE syntax offers considerable flexibility and control, including complex pattern matching capabilities that make it possible to redirect large numbers of URLs with a single redirection rule. Therefore, it is not generally necessary to provide rewrite rules on an individual URL basis.
47. There are a number of components that can be used to achieve similar results, and Website Managers do not have to use the ones that The National Archives supplies. However, Website Managers do need to implement functionality that will reduce the occurrence of broken links, and redirect users where appropriate to the Web Archive.
48. Before looking into the technical issues surrounding implementation, Website Managers should take a strategic view as to how they intend to handle unresolved links across the whole of the website. The technical implementation can be delivered with a combination of the native functionality offered by web server software, and the redirection software. To follow the process through:

The user requests a URL e.g. <http://www.mydepartment.gov.uk/default.html>

- If the URL is resolved, it is served back to the user in the normal way. Wherever possible, Website Managers should specify default pages for sites and directories. E.g. a user who types in <http://www.mydepartment.gov.uk> would also arrive at the above page.
- If the URL is not resolved, then a process needs to be in place to handle the request appropriately. This process should assess the URL and decide upon the most appropriate course of action. This could include:
 - i. displaying a standard departmental 404 error page
 - ii. displaying a customised error page

⁵ Email: webcontinuity@nationalarchives.gov.uk

- iii. redirecting to another URL
- iv. redirecting to the UK Government Web Archive

We consider each in turn.

Displaying a standard departmental 404 error page

49. In many cases displaying a standard departmental error page is the most appropriate course of action. This includes cases where the user has clearly mistyped, or followed an incorrectly coded hyperlink. For example, <http://www.mydepartment.gov.uk/<> is obviously incorrect, and a standard error page would be appropriate since there is no obvious pointer as to the site content that the user was seeking. Websites should have one or more standard error pages.

Displaying a customised error page

50. Website Managers may decide to display customised error information to users for specific areas of the site, in order to include relevant information, links and advice about what action the user may then take. For example, you may want to display different error information for the following:

www.mydepartment.gov.uk/about-us/no-such-page.htm

www.mydepartment.gov.uk/publications/no-such-publication.pdf

51. Customised error information can be implemented by the use of separate error pages, or a dynamic page that modifies the content displayed depending on the requested URL.

Redirect to another URL

52. Redirecting to another URL may be the most appropriate course of action where the website has been reorganized, or responsibilities have changed due to departmental reorganization or Machinery of Government changes. Wherever possible, the redirection should be to a specific URL for the original content, and not to a general “landing page”. E.g. an original URL like:

www.mydepartment.gov.uk/about-us/visting-times.htm could be redirected to:

www.mydepartment.gov.uk/info/visting-times.htm or

www.mynewdepartment.gov.uk/about-us/opening-hours.htm

53. In general, most such redirections are likely to be specified at the directory level.

Redirect to the UK Government Web Archive

54. In some cases TNA redirection will be the first resort, not the last. If content has been deleted and is no longer current then redirection to the archive is the correct response. However, this guidance is intended to address issues that have been raised as a consequence of convergence/rationalisation. Where sites are closing and significant amounts of content are being moved (e.g. to Directgov), it may not be appropriate for users to be redirected to the web archive. It may be more appropriate to redirect users to a different website where the content has been moved or made available in an updated form. Depending on what the information is, web managers should manage this content appropriately through change.
55. Redirection software such as the components issues by TNA can be programmed such that the web server will issue a “moved permanently” instruction to the user’s browser (HTTP code 301). The browser will then automatically redirect to the Web Archive and attempt to find the latest version of the page. The intention is that, for example <http://www.mydepartment.gov.uk/about-us/may-2008-info.htm> resolves to [http://webarchive.nationalarchives.gov.uk/+
http://www.mydepartment.gov.uk/about-us/may-2008-info.htm](http://webarchive.nationalarchives.gov.uk/+http://www.mydepartment.gov.uk/about-us/may-2008-info.htm)
56. Alternatively, users can be directed to an intermediate or “bridging page”, from where they can choose whether or not to look for the page in the web archive. This approach is more complex to implement but may be less disconcerting for users. Also, the absence of a direct 301 redirection means that any search engine ranking possessed by the original page is not transferred to the archived version.
57. If the page is found in the Web Archive, it will be served back to the user. This page will effectively be the original page that existed on the site, but with the addition of a banner at the top to indicate that it is archived content. The page title will also be prefaced by the words [ARCHIVED CONTENT], for clarity of context especially when the page is retrieved via search engines. The links away from the page will point back to the original live destination.
58. If the page is not found in the web archive, a further redirection back to the original site will be issued. This redirection is to a dummy “archive checked not found” page in the format:

<http://www.mydepartment.gov.uk/ukgwacnf.htm>

This page need not actually exist, since the redirection software can be programmed in such away to allow you to handle these cases as you decide.

59. In such cases you should program the redirection behaviour in such a way as to inform the user that the page requested could not be found in either the original site or the Web Archive.

Redirection Flowchart

