

OPERATIONAL SELECTION POLICY OSP27

UK Central Government Web Estate

April 2014



© Crown copyright 2014

You may re-use this information (excluding logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence or email psi@nationalarchives.gsi.gov.uk.

Where we have identified any third-party copyright information, you will need to obtain permission from the copyright holders concerned.

This publication is available for download at nationalarchives.gov.uk.

Contents

1. Authority.....	3
2. About the UK Government Web Archive.....	4
3. Evolution of Scope.....	4
4. Content selected for permanent preservation and made available through the UKGWA.....	5
5. In scope material that is not available within the UKGWA	8
6. Further guidance and contact information.....	9

1. Authority

1.1 The National Archives' [Records Collection Policy](#) was published in November 2012 and replaces the 2007 Acquisition and Disposal Policy. The Records Collection Policy sets out the records that The National Archives will and will not seek to collect for permanent preservation, when it will collect them and where they will be held.

1.2 Operational Selection Policies (OSPs) are intended to be working tools for those involved in the selection of public records. This policy may, therefore, be reviewed and revised in the light of comments received from the users of the records, from archive professionals, the departments' experience of using the policy, or as a result of newly discovered information. There is no formal cycle of review and we welcome comments at any time. The extent of any review and revision exercise will be determined according to the nature of the comments received.

1.3 If you have any comments on this policy, please email:
information.management@nationalarchives.gsi.gov.uk

Or write to:

Information Management and Practice Department
The National Archives
Kew, Richmond
Surrey
TW9 4DU

2. About the UK Government Web Archive

2.1 The [UK Government Web Archive](#) (UKGWA) is one of the largest and most heavily used web archives in the world, containing over three billion URLs and frequently receiving more than ten million page views each month. The National Archives has a remit to preserve government owned web content in all its forms (including web pages, official publications, datasets, tweets and multimedia) and seeks to preserve this part of the record in its original context wherever possible through this archival resource. The UKGWA is free to use and fully accessible via the web.

2.2 The National Archives' approach to web archiving involves harvesting websites remotely and automatically to a schedule by using a crawler as this is the most appropriate method for archiving at scale.

2.3 Originally, The National Archives worked with the Internet Archive to capture and preserve the online government record. The UKGWA is now a service that is provided, under contract to The National Archives, by the Internet Memory Foundation, a not-for-profit organisation.

2.4 Other collecting institutions in the UK undertake archiving of the wider UK web domain. For example, the websites relating to the activity of the devolved administrations are largely captured by the National Library of Scotland and the National Library of Wales. The Legal Deposit Libraries (Non-Print Works) Regulations 2013 enables copyright libraries in the UK to undertake their web archiving activity.

3. Evolution of Scope

3.1 The National Archives collects records from all government departments and bodies whose records are public records under the Public Records Act¹. Public records can exist in any format or any medium. Our role for the 21st century is to collect and secure the future of the public record, both digital and physical, to preserve it for generations to come, and to make it as accessible as possible.

3.2 Government use of the online space is evolving. During the 1990s most central government departments and agencies began to produce websites to make information and online services available to the public. Prior to 2007 The National Archives focussed its web archiving activity on capturing a selection of websites that encompassed departments of state, constitutional arrangements and public inquiries. Themed collections that included other sites were also developed based on important issues or events such as the budget or general election.

3.3 Government continued to expand its use of websites to publish official documents and communicate with the citizen and it was recognised that these websites were of archival value.

¹This can include records that may be held or created on behalf of a government department by another body, for example a body carrying out work for the department under contract, depending on the terms of the contract.

UK Central Government Web Estate

Successive website rationalisation programmes and the transient nature of web publishing were identified as threats that would lead to a loss of this part of the record in the longer term and cause a frustrating user experience for anyone wanting to locate information that had been moved or removed from its original location. From 2008 The National Archives expanded the scope of the UK Government Web Archive and introduced the [web continuity service](#), using redirection technology that works with the UKGWA to create persistent links enabling users to find material removed from its original location online.

3.4 Government now publishes a range of information online in a wide number of formats. Social media, datasets and more conventional web content such as websites and blogs can all be captured and made accessible within the boundaries of web archiving technology. Over time this will mean that we are able to capture more of the government record for permanent preservation via the web archiving programme.

3.5 It is The National Archives' intention to capture datasets published on government websites into the UKGWA where the publication format and data contained within it is amenable.

3.6 However, all digital records produced by government departments should continue to be appraised and, where appropriate, selected for permanent preservation in line with The National Archives' [guidance](#). Underlying data that cannot be published for reasons of sensitivity may still be selected for long term preservation, and managed in the creating department, until the sensitivity falls away and the data can be published.

3.7 Government now also uses social media and other forms of digital media to consult and interact with the citizen online. The way in which the public record is published has changed and The National Archives has adapted its web archiving activity in order to continue capturing, preserving and giving access to the entire UK government web estate. A list of the sites and social media accounts that are currently in scope for the UKGWA is available on The National Archives [website](#).

4. Content selected for permanent preservation and made available through the UKGWA

4.1 As outlined in section three, The National Archives now aims to capture and preserve the whole of the UK central government web estate where said capture is possible within the confines of resource and web archiving technology.

4.2 In practice this means that selected websites are captured at least once and, if closing as part of a wider government programme, within six months of their closure, where possible. Additional crawls are undertaken when events are likely to impact significantly on site content such as the period before and after a general election. The core content of the UKGWA is comprised of material published by departments of state either on their own family of sites or on gov.uk; non-departmental public bodies; agencies; delivery channels; royal commissions; NHS websites that are national in focus. The list of sites selected for permanent preservation

fluctuates in line with government's use of the web. Sites for inclusion in the UKGWA are identified by The National Archives, through successive government web rationalisation programmes and by the individual site owners themselves. If the status of the organisation that owns the web presence is not clear The National Archives may make use of its [determination and change of status guidance](#) to inform a decision on whether or not it is in scope.

4.3 Website owners and / or digital leaders across government should seek to ensure that web archiving is considered when designing a site. They should make sure that The National Archives is informed of the existence of a new web presence at the earliest opportunity so that, if it is within scope and suitable for preservation, preparations to enable its capture can be made at a sufficiently early stage. Site owners are responsible for checking that successful capture of a site has been achieved, ensuring that the content is preserved fully and can remain accessible through the UKGWA and web continuity, prior to closure.

4.4 Websites with a clearly limited lifespan such as public inquiries, inquests and campaigns are captured at least once prior to their closure. Examples of our holdings in this category include The Shipman Inquiry and the Coroner's Inquests into the London Bombings of 7 July 2005. Informing The National Archives of the existence of a web site in this category is particularly important as sometimes the majority of records produced by the inquiry are made available online and facilitating their capture through web archiving may mean that these records do not need to be transferred by alternate means.

4.5 The legal framework² that allows the public record to be collected, preserved and made accessible provides the UKGWA with the boundaries of a collecting remit. Many of the potential information rights issues posed by the archiving of the government web estate are inherent within the creation and active management of other formats of digital, or indeed paper, records.

4.6 Where material that forms part of the public record is published on sites or services outside of government ownership the aim is to retain this part of the record without infringing the rights of the site or service provider or other users of the service so capture and preservation is limited to material firmly in scope. For example, complex technical solutions have been developed in order to limit the risk of our capturing out of scope material when archiving social media.

The National Archives' Twitter archive has been developed to capture material in line with the following rules in order to use resources effectively and limit capture to the public record:

In scope: Twitter feeds of departments of state and non-ministerial departments. Other government organisations and delivery channels can be archived at the request of departments. Tweeted links to published digital material that is accessible on domains that are within scope for the UKGWA will resolve fully.

Out of scope: Re-tweets, responses to tweets and the accounts run by officials in a personal capacity. Tweeted links to out of scope material (e.g. newspaper articles) will be visible but the material will not be available in the UKGWA.

² Including the [Public Records Act 1958](#) and [Section 163 of the Copyright Designs and Patents Act 1988](#)

UK Central Government Web Estate

Below is an illustrative example of a twitter feed that is in scope (@GOVUK) showing what will or won't be captured for permanent preservation:

Replies sent by the account that is in scope **will** be archived

Link to a site in scope **will** resolve in full

The replies, or conversation, encompassing tweets from other accounts **won't** be archived

Re-tweets **won't** be archived

Links to sites out of scope **won't** resolve. Users **will** be given the full URL to use to search the live web

4.7 As demand for web archiving services has grown The National Archives has undertaken pilot studies to capture websites that are outside of the scope of our day to day preservation processes. Such web archiving activity serves to raise awareness of web archiving as a method of capturing the record and is undertaken at the request of the sectors or organisations with a collecting remit that extends to the digital material captured. Information on the websites captured as part of the local NHS and local government pilot studies is available on The National Archives [website](#). The usual quality assurance processes undertaken when archiving central government sites were not applied to these projects and their capture quality may not be to the usual high standard as a result.

4.8 In addition to these pilot studies we have also undertaken limited capture of websites and social media where The National Archives has a clear collecting role that includes capturing material published on sites that are outside of our usual scope. For example, as part of our activities to capture the record of the London 2012 Olympic Games, The National Archives developed a [themed collection](#) that encompasses government sites as well as a wider range of cultural and social media sources. Such projects and pilot studies will continue based on resource and demand from government and archive sector partners.

4.9 Through the UK Government Web Archive, The National Archives aims to preserve as complete a version of the open government record as possible within the boundaries of what is technically achievable. This online record is kept accessible through deployment of the web continuity redirection component, the provision of a full text search across the contents of the UKGWA, [themed collections](#) that highlight particular events or types of record and an [A-Z](#) list that covers the majority of sites being archived. Entries from the A-Z list are available as record series through Discovery, The National Archives online catalogue. The National Archives also invests in discovery tools to enable further exploitation of the data contained in the UKGWA such as implementing [memento timestamps](#) and providing a way for developers to interrogate and build on the 90 terabytes of data it contains through an application programming interface (API).

5. In scope material that is not available within the UKGWA

5.1 In the early days of the web archiving programme The National Archives set out to capture the look and feel of the selected websites, their basic functionality and navigation. The way in which the remote harvesting of sites worked at that time meant that the archiving software often collected the top-level pages of a website only, usually without images. Some of the older archived websites in the collection therefore do not display as expected and have only basic functionality.

5.2 Following improvements in web archiving technology and our commitment to quality assurance The National Archives now captures the government web estate as comprehensively as possible, although the capture and preservation of every level, document or media cannot be guaranteed.

5.3 Further [guidance](#) is available for those managing government websites regarding the web archiving process and how to ensure sites can be captured. This guidance explains the technical limitations further and highlights where capture is not routinely possible. For example, material that is only available behind a log-in screen cannot be captured and the capture of sites that exist solely for the purpose of providing a transactional service may be incomplete. For illustration, the interface DVLA employ at www.gov.uk/tax-disc will be captured but a user will not be able to access the full functionality of that part of the web domain in its archived form.

5.4 Government departments are responsible for maintaining their web presence in accordance with all relevant legislation. Material published should comply with data protection, defamation and copyright law. The National Archives relies on the government organisations originally publishing the material on their sites to ensure that the content is Crown copyright and can be archived without infringing the copyright of non-government organisations or individuals. Should a site contain material that is outside of Crown ownership and / or copyright of a third party the publishing department must notify The National Archives of its presence, provide evidence of its status and ensure that said status is clearly visible where applicable on the archived site. If appropriate permissions have not been granted by the copyright owners then The National Archives may be unable to archive a site.

5.5. The National Archives has a [takedown policy](#) and process that enables the removal of content from the UKGWA to take place as transparently as possible. On receipt of a request from a member of the public or a government department material can be removed temporarily from the web archive. The case will then be considered by a Takedown Panel, composed of members of staff who provide relevant expertise. The panel will approve continued withdrawal of the material only if one of the criteria set out in the policy is met. A list of [removed web pages](#) and the policy that guides their removal is available on The National Archives website.

6. Further guidance and contact information

6.1 Government digital leaders and website owners should contact the Government Digital Service for further advice on the single government domain project. For advice on web content and design you can use the government service [design manual](#).

6.2 If a government owned web presence is to close or change significantly, for example as part of the migration to a single domain (www.gov.uk) or as a result of a machinery of government change, site owners must notify The National Archives at least eight weeks in advance of closure or change to allow sufficient time to capture the site effectively.

6.3 If a government web presence is not being captured by The National Archives but should be (perhaps it is the website of a newly created government body or initiative) the department should contact The National Archives to enable us to decide whether we, or another body, should capture the site and begin preparation to ensure effective capture can take place.

6.4 Please direct all enquiries about the UK Government Web Archive to The National Archives' Web Continuity team via webarchive@nationalarchives.gsi.gov.uk