

How to classify changes to government websites

A classification of Web content alterations and changes in access to Web resources

Toly Rinberg (trinberg@sunlightfoundation.com), Andrew Bergman (abergman@sunlightfoundation.com)
Sunlight Foundation

October 4, 2017

Version 1.1

Updated January 22, 2018

Executive Summary:

This document serves to classify the different ways in which government agencies can alter their Web content and change access to that Web content. This summary contains the overarching classes of changes, including the ways in which government agencies approach communicating about changes and archiving their Web resources. The rest of the document dives more deeply into these classes of changes, detailing sub-classes that often depend on the mechanisms that need to be implemented to make the changes.

This is a **living document** and will be continuously updated as the classifications below are refined and updated and as additional examples are added. Please contact us with any information about applicable examples or input about the classifications and how they might be better differentiated. For more context and insight into the motivation for the different classifications, read our [blog post](#) on the topic.

The following are the various classifications of changes to government websites:

[Classification of Non-Maintenance Web Content Alterations and Access Reductions to Web Resources](#)

1. [Altering or removing text and non-text content](#)
2. [Altering or removing links](#)
3. [Moving an entire webpage or collection of webpages or establishing redirects](#)
4. [Altering or removing an entire pertinent section of a webpage or collection of webpages](#)
5. [Removing an entire webpage or document](#)
6. [Overhauling or removing an entire website](#)
7. [Altering or removing search engines and open data platforms](#)
8. [Altering, removing, or deleting datasets](#)

Classifications of Approaches to Changes and Access Reductions

- Assessment of continued availability of affected Web resources on agency websites
- Classes of storage of affected Web resources in agency Web archives
- Assessment of notice provided or explanation of change by agency

Classification of Web Content Alterations Relating to Web Maintenance and Expansion of Access to Web Resources

1. Correcting typographic or grammatical errors
2. Adding new, pertinent text and non-text content
3. Updating or adding links
4. Updating resources from classes 4-8 above
5. Adding pertinent information and resources from classes 4-8 above

For each class of change that we have observed in the list above, we've included an accompanying description of an example of the change. Examples below are drawn largely from the [website monitoring reports](#) produced by the [Environmental Data & Governance Initiative](#) (EDGI), in addition to other instances that have been described by the news media and the Sunlight Foundation. Both types of EDGI reports, Content Change Reports (CCRs) and Access Assessment Reports (AARs), are included when applicable.

Acknowledgements: We'd like to recognize [EDGI's Website Monitoring Team](#), made up of volunteer website monitoring analysts and software developers, for contributing to the body of work that helped inform the production of this document and for valuable discussions. In particular, we'd like to recognize contributions by our EDGI Website Monitoring Team co-founders, Maya Anjur-Dietrich, Gretchen Gehrke, and Rebecca Lave, for initial input.

Classification of Non-Maintenance Web Content Alterations and Access Reductions to Web Resources

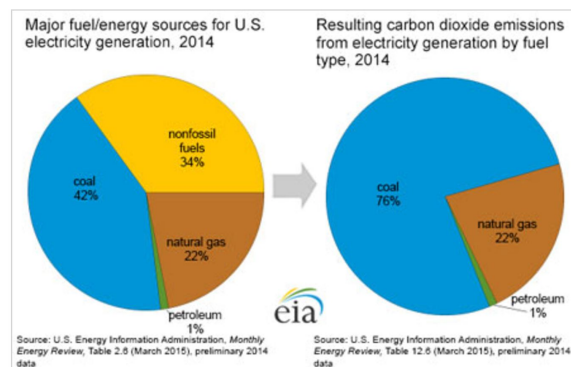
1. Altering or removing text and non-text content

a. Change or removal of text snippets (word(s), sentence(s), title(s), link text)

Example: Text was changed from “U.S. clean tech investment” to “U.S. energy technologies investment” on a Department of Energy webpage. [EDGI Report [CCR-16](#): Change #2.1d]

b. Change or removal of image

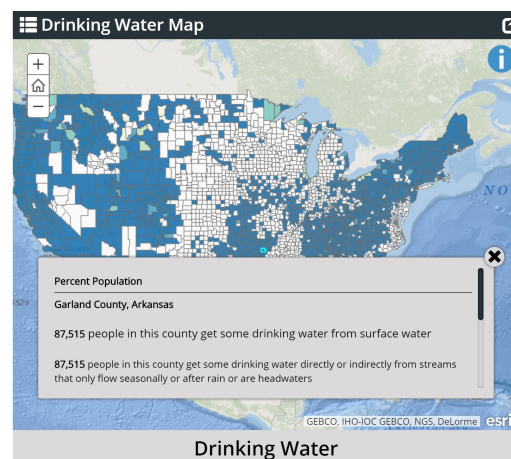
Example: A graph showing major energy sources for US electricity generation and resulting carbon dioxide emissions was removed from the Energy Information Administration’s Kid’s educational website. [EDGI Report [CCR-5](#): Change #1.2b]



c. Change or removal of interactive Web element

Example: The interactive map, “Populations that get Drinking Water from Streams,” was removed as part of the removal of the Clean Water Rule website. [EDGI Report [AAR-2](#): Screenshot #1.1]

Populations that Get Drinking Water from Streams



View a live version of the interactive map in the [EPA.gov January 19 Snapshot](#)

2. Altering or removing links

a. Not updating a link to a page that has been moved to a new URL

Example: A URL that had linked the Spanish language webpage about the Executive Order to evaluate the Clean Water Rule was not updated on the “Waters of the United States (WOTUS) Rulemaking” page for two weeks. [EDGI Report [AAR-2](#): Screenshot #2.1]

b. Improperly updating a link to a page that has been moved to a new URL

Example: A link on an Environmental Protection Agency webpage was improperly updated, replacing the “/climatechange/” path segment with “/climate-change/” instead of “/ghgemissions/,” such that it continued not to direct to the live version of the page. [EDGI Report [AAR-3](#): #1, 4]

c. Removing a link to a page that has not been permanently removed or has not been removed at all

Example: The link, [Climate Change and Human Health](#), which directs to a live PDF of a brochure with that name, was removed from the “Climate Change” webpage on the National Institute for Environmental Health Science website. [EDGI Report [CCR-17](#): Change #4.3a]

3. Moving an entire webpage or collection of webpages or establishing redirects

a. Moving a page or collection of pages to a new URL or URLs without a redirect from the previous URL or URLs

i. The previous URL or URLs lead to a “dead” page or a 404 error.

Example: Many webpages, like the [main climate change page](#), were moved from whitehouse.gov to the “obamawhitehouse.archives.gov/” domain, without establishing redirects and thus leaving old URLs that result in 404 errors. [[The Washington Post](#)]

ii. The previous URL or URLs lead to a page or pages that are related yet substantively different

iii. The previous URL or URLs lead to a page or pages that are completely unrelated or may be an organization-wide homepage

b. Moving a collection of pages to new URLs without a redirect from the previous URLs and without updating links, leaving pages disconnected from the rest of their collection.

c. Moving a page or collection of pages to a new URL or URLs with a redirect from the previous URL or URLs

- i. **The redirect is established from each individual previous URL to a separate page**
Example: A series of webpages on the Federal Highways Administration website were moved from the “fhwa.dot.gov/environment/climate_change/” subdomain to the “fhwa.dot.gov/environment/sustainability/” subdomain, with redirects established between the previous and new URLs. [EDGI Report [CCR-11](#): Pages #1-6]
- ii. **The redirect is established from each individual previous URL to just the new website’s homepage**

4. Altering or removing an entire pertinent section of a webpage or collection of webpages

- a. **A section of a single webpage is altered or removed**
Example: Substantial sections of the Department of Interior’s climate change webpage, including descriptions of Interior’s work on climate issues, were removed. [EDGI Report [CCR-15](#): Page #1]
- b. **The same section is removed across a collection of webpages**

5. Removing an entire webpage or document

Pertains to removal of webpages or documents from a particular URL or domain. The resources may still be available on the Web but may be at another domain or stored in a Web archive.

- a. **The previous URL leads to a “dead” page or a 404 error**
Example: The State Department removed its Climate Action Report webpage and the page’s URL began leading to a “We’re sorry, that page can’t be found” [notice](#). [EDGI Report [CCR-1](#): Page #3]
- b. **The previous URL redirects to a page that contains a statement that the previous page and its information has been removed**
Example: The Environmental Protection Agency removed webpages from its climate change subdomains, including “epa.gov/climatechange” and “epa.gov/climate-impacts”. The previous URLs began redirecting to a [notice page](#) stating “This page is being updated”. [[EDGI blog post](#)]
- c. **The previous URL redirects to an existing or new URL for another page**
 - i. **The page is related yet substantively different**
 - ii. **The page is completely unrelated or may be an organization-wide homepage**

6. Overhauling or removing an entire website

Pertains to substantial rewriting and restructuring of text and non-text content. A “website” may be a collection of webpages hosted on one Web domain or may be collection of webpages, with related content and significant cross-linking, hosted on multiple subdomains.

a. An entire website’s previous URLs lead to a “dead” page or a 404 error

Example: The Obama Administration’s [“The Record” website](#), was moved to the “obamawhitehouse.archives.gov/” domain, without establishing redirects and thus leaving old URLs that result in 404 errors. [[The Washington Post](#)]

b. An entire website’s previous URLs redirect to a page that contains a statement that the previous page and its information has been removed

Example: The Environmental Protection Agency removed its climate change website, which was hosted on subdomains such as “epa.gov/climatechange” and “epa.gov/climate-impacts.” Previous URLs on these subdomains began redirecting to a [notice page](#) stating “This page is being updated”. [[EDGI blog post](#)]

c. An entire website’s previous URLs redirect to existing URLs for another page

- i. The pages are related yet substantively different
- ii. The pages are completely unrelated or may be an organization-wide splash page

d. An entire website’s previous URLs redirect to a URL or URLs for a replacement website

- i. The redirect is established from each individual previous URL to a separate page
- ii. The redirect is established from each individual previous URL to just the new website’s homepage

Example: The Environmental Protection Agency removed its Clean Water Rule website, hosted at “epa.gov/cleanwaterrule”, and launched a new website called “Waters of the United States (WOTUS) Rulemaking” at a new subdomain “epa.gov/wotus-rule”. All previous website URLs began redirecting to the new website’s homepage. [[EDGI Report AAR-2: Page #1-2](#)]

e. At the same URLs, the webpages for an entire website are replaced with new content

- i. The pages are related yet substantively different

- ii. **The pages are completely unrelated**

7. Altering or removing search engines and open data platforms

Search engines and open data platforms allow for searching Web resources, often accessing a corresponding database of records. Records may still be available on the Web without the engine or platform, but access to them may be significantly reduced.

- a. **Altering a search engines or open data platform that provides access to documents, datasets, or information that are accessible elsewhere on the website**

- i. **Altering the search function and output such that the same search results are presented or prioritized differently**
- ii. **Altering the search function and output such that fewer search results are available, though the content that the results linked to may not itself be altered in any way**

Example: The U.S. Geological Survey (USGS) changed its [search engine](#), reducing the number of results for the same search term. The search results link to content hosted elsewhere on the USGS website. [[Sunlight Foundation blog post](#)]

- b. **Altering a search engines or open data platform that provides access to documents, datasets, or information that are only accessible there**

- i. **Altering the search function and output such that the same search results are presented or prioritized differently**
- ii. **Altering the search function and output such that fewer search results are available, thereby rendering that content inaccessible**

- c. **Removing a search engines or open data platform that provides access to documents, datasets, or information that are accessible elsewhere**

- d. **Removing a search engines or open data platform that provides access to documents, datasets, or information that are only accessible there**

Example: The Department of Energy removed its [phone directory](#), a platform that provided access to staff contact information only accessible through that platform. [[E&E News](#)]

8. Altering, removing, or deleting datasets

This includes the removal of the primary webpage that provides the URL for the server data request or the removal of the data from the server itself.

- a. **Removing the primary webpage or website that hosts a dataset, contains links to a dataset, or contains a search engines or open data platform that provides sole access to a dataset**

Example: The Department of Agriculture removed [animal welfare data sets](#), removing [pages linking to reports](#) and the reports themselves. [[The Memory Hole blog post](#)]

- b. **Removing or altering entries from a dataset**
- c. **Removing or altering fields from a dataset**
- d. **Deleting a dataset from government servers**

Classifications of Approaches to Changes and Access Reductions

Note: This section will be substantially expanded in future versions of this document and examples for individual classes will be added. An applicable example with accompanying suggested best practices can be found [here](#).

Assessment of continued availability of affected agency Web resources

1. Resources **are** available on agency webpages or websites other than those affected (which means those webpages or websites host, contain links to, or contain a search engine or open data platform that provides access to the resource)
2. Resources **are not** available on agency webpages or websites other than those affected, but they are available on other government websites
3. Resources **are not** available on any government webpages or websites other than those affected
4. Resources **are not** available elsewhere on the Internet other than the affected webpages or websites

Classes of storage of affected Web resources in agency Web archives

1. Archive **is** searchable
 - a. Archived content is discoverable through a public search engine, like Google
 - b. Archived content is **not** discoverable through a public search engine, like Google - the URL for the archived page and the archived page itself may or may not be discoverable using a search engine
2. Archive **is not** searchable
 - a. Archived content discoverable through a public search engine, like Google
 - b. Archived content **not** discoverable through a public search engine, like Google

Assessment of notice provided or explanation of change by agency

1. Agency notified the public in advance of the change
2. Agency responded to or notified the public after the change
3. When altering an entire webpage, collection of webpages, or websites:
 - a. New page or collection of pages **do not** mention the previous version or versions and how to find an archive
 - b. New page or collection of pages **do** mention the previous version or versions and how to find an archive
4. When moving or removing an entire webpage, collection of webpages, or websites:
 - a. The previous URL or URLs lead to a page or pages that **do not** mention the previous version or versions and how to find an archive
 - b. The previous URL or URLs lead to a page or pages that **do** mention the previous version or versions and how to find an archive
5. Agency proactively established a Web archive in advance of the change
6. Agency established a Web archive after the change

Classification of Web Content Alterations Relating to Web Maintenance and Expansion of Access to Web Resources

Note: This section will be substantially expanded in future versions of this document and examples for individual classes will be added.

1. Correcting typographic or grammatical errors

2. Adding new, pertinent text and non-text content

3. Updating or adding links:

- a. Removing a link to a page that has been permanently removed
- b. Updating a link to a page that has been moved
- c. Adding new links to pages

4. Updating resources from classes 4-8 above

5. Adding pertinent information and resources from classes 4-8 above