

How to classify changes to government websites

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

Contributors: Toly Rinberg, Andrew Bergman, Rachel Bergman, Sarah John, Aaron Lemelin, Jon Campbell
Web Integrity Project, Sunlight Foundation

Version 1.2

Updated March 6, 2019

(First published: October 4, 2017)

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. Gov404: The Web Integrity Project's Censorship Tracker uses this classification in detailing cases of removals of, and reductions in access to, content.

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](#)

Acknowledgements: We'd like to thank Alex Howard for conversations and input in the initial development of this classification. We'd also 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).
- b. Change or removal of image, graphic, video, or other visual element.
- c. Change or removal of interactive Web element.

2. Altering or removing links

- a. Not updating a link or links to a page that has been moved to a new URL.
- b. Improperly updating a link or links to a page that has been moved to a new URL.
- c. Removing a link or links to a page that has not been permanently removed or has not been removed at all.
- d. Removing a link or links to a page that has not been permanently removed so as to “orphan” that page, making it inaccessible by navigating through the website within which the page is located.

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, [a 404 error, or a 403 error](#).
 - ii. The previous URL or URLs lead to a statement that the previous page(s) and its information has been removed.
 - iii. The previous URL or URLs lead to a page or pages that are related yet substantively different.
 - iv. 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.
 - 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.
- 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, a [404 error](#), or a [403 error](#).
- b. The previous URL redirects to a page that contains a statement that the previous page and its information has been removed.
- 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.
- d. The previous URL leads to a page that requests authorization credentials to access the content or a 401 error.

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, similar menus, or significant cross-linking, hosted on multiple subdomains.

- a. An entire website is removed or overhauled, and a significant portion or all of the website’s previous URLs lead to “dead” pages, 404 errors, or 403 errors:
 - i. A significant portion of the previous website is replaced with new content at different URLs.
 - ii. The previous website is not replaced with new content.
- b. An entire website is removed or overhauled, and a significant portion or all of the website’s previous URLs redirect to a page that contains a statement that the previous information has been removed:
 - i. A significant portion of the previous website is replaced with new content at different URLs.
 - ii. The previous website is not replaced with new content.
- c. An entire website is removed or overhauled, and a significant portion or all of the website’s previous URLs redirect to existing URLs for another page or pages:
 - i. The pages are related yet substantively different.
 - ii. The pages are completely unrelated or may be an organization-wide homepage.
- d. An entire website is removed or overhauled, and a significant portion or all of the website’s previous URLs redirect to a URL or URLs for a replacement website:
 - i. The redirects are established from each individual previous URL to separate pages, which are related yet substantively different.
 - ii. The redirects are established from each individual previous URL to separate pages, which are completely unrelated.
 - iii. The redirects are established from each individual previous URL to the new website’s homepage.
- e. An entire website is overhauled, and a significant portion or all of the website’s webpages are replaced with new content at the same URLs:
 - i. The pages are related yet substantively different.

- ii. The pages are completely unrelated.
- f. An entire website is overhauled, and a significant portion or all of the website's URLs lead to a page that requests authorization credentials to access the content or a 401 error.

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 engine 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.
- b. Altering a search engine or open data platform that provides access to documents, datasets, or information that are only accessible through the search engine or open data platform:
 - 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 engine or open data platform that provides access to documents, datasets, or information that are accessible elsewhere on the website.
- d. Removing a search engine or open data platform that provides access to documents, datasets, or information that are only accessible through the search engine or open data platform.

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 or contains links to a dataset, or contains a search engine or open data platform that provides sole access to a dataset (as in 7.d). The dataset is "islanded;" it is still available at some public address, but it is inaccessible from the website within which it is located.
- b. Removing or altering entries from a dataset.
- c. Removing or altering fields from a dataset.
- d. Adding a request for authorization credentials to access the dataset.
- e. Removing an entire dataset.
- f. 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
 - i. 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 **is** discoverable through a public search engine, like Google
 - b. Archived content **is 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**