Uncovering Hidden Government Documents

BY ANDREA L. HAMILTON

S hifts in presidential administrations, agency policies and procedures, and internet standards often result in changes to the content and organization of federal websites. This can make it difficult for researchers to locate documentation—such as consent decrees and guidance memoranda—that may still be in force. The hyperlink or title of such guidance may appear in the Federal Register or within an article or court filing, or it may be referred to by the agency itself, but the link may be broken or the document may be otherwise difficult to locate. The following tips can be used to find elusive government records.

If the URL is Known

If the URL is known, there are several ways to search for the information. First, the Internet Archive[1] provides access to 20-plus years of web history through the Wayback Machine.[2] After entering a URL, you'll be taken to a calendar showing when and how successfully the site was saved. Larger circles indicate multiple captures on a particular day, making more of the site's content available for viewing; blue circles indicate a more successful crawl of the site.[3] Click on a date and provided snapshot link to view the desired site or document. (See Fig. 1.)

Another option is to perform a Google search. Google's cached version of a website shows what the web page looked like the last time it was visited by one of its crawlers. Enter the **cache:** search operator in front of a URL into the Google.com search field.[4] For example, to review a cached version of the CBA website, enter **cache:cobar.org** into the search field.
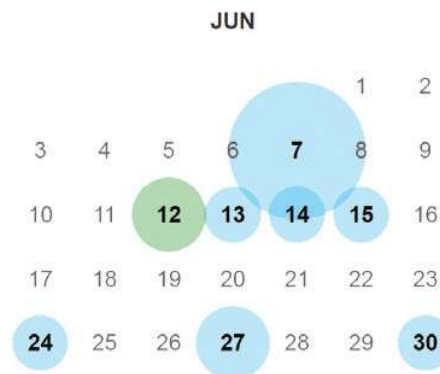

Figure 1. The 2007 Wayback Machine calendar page for www.whitehouse.gov.

While these backup snapshots are generally used to provide Google users with quick access to slow or nonresponsive web pages and are typically not very old, they may provide the needed entry to a site or document.

Finally, if navigating directly to a known web address is not successful, the URL might provide insight as to where a document used to "live" on a website before its restructuring. For example, the URL www.epa.gov/newsreleases/epa-declares-outdoor-burn-ban-tulalip-reservation.html demonstrates the needed document is titled "EPA Declares Outdoor Burn Ban for Tulalip Reservation." Simply remove the end of a link—anything following the last forward slash, repeating as necessary—until a valid site is reached. In the case of a major website reorganization, you may need to remove everything but the root of a site's URL (e.g., www. epa.gov). Then, search the site for the desired guidance using the keywords provided in the URL. This strategy is particularly useful when trying to locate a PDF that's not text-searchable (i.e., it's been saved as an image), as keyword searches would not be effective.

If the Document Title is Known

Just as a URL can provide insight as to where a document used to reside, it can also indicate the title of the needed guidance. Searching for a known document title in quotation marks ensures the search engine will keep all terms together as a phrase.[5] You can use the **site:** search operator to further refine the search to a specific website.[6] Using the previous EPA example, you would type **"epa declares outdoor burn ban for tulalip reservation" site:epa.gov** into the Google.com search field. This leads to the archived article on the EPA website.

This strategy is often more effective than using a website's embedded search tool. But because each search engine employs proprietary bots that crawl and index the web and use unique algorithms to rank websites, you may need to use more than one search engine and review more than the first page of search results.

Another helpful resource is OCLC, a global cooperative of libraries that "collectively steward a vast quantity of knowledge"[7] through World-Cat, which bills itself as "the world's largest network of library content and services."[8] With WorldCat's advanced search, users can search by keyword, title, and author.[9] The full record of a search result often contains "links to this item" and a unique identifier called a PURL, or persistent uniform resource locator.[10] Even if a known URL is broken, a PURL may still be a successful means of access. "Links to this item" may also list a previously unknown website; use the "Known URL" tips above if this address no longer works.

Finally, HathiTrust is a nonprofit collaborative of academic and research libraries with more than 17 million digitized items.[11] The mission of its US Federal Government Documents Program is to enhance digital access to US federal publications, including those issued by the US Government Publishing Office (GPO) and other federal agencies.[12] Use the advanced catalog search to input information about the needed

document, including author, title, and subject.[13] More refined searches may be conducted within the following US Federal Documents Collections: US Federal Documents, US Congressional Serial Set, Bureau of Indian Affairs publications, US Environmental Protection Agency publications, Foreign Relations of the United States, Statistical Abstract of the United States, and US Civil Rights Commission.[14] Nonmembers can search across all collections, but viewing and downloading privileges may be restricted.

Additional Search Options

If you have a quote or passage from the desired document or site, try searching for that excerpt within quotation marks. The guidance you seek may have been archived on a website other than the issuing agency's. If searching for the excerpt generates a lot of results but not the full text of the document, try using the **filetype:pdf** search operator for a more targeted search by document format. For example, entering **"epa wastewise" filetype:pdf** into the Google.com search field returns only PDF results. Additionally, both the Wayback Machine[15] and HathiTrust[16] allow for full-text searching.

If you don't have a URL, the document title, or a direct quote, conduct a broader internet search using any known document details, such as author, recipient, parties, agency, document, case or other identifying number, and date. You can also try to identify online repositories or special collections that typically contain the type of document you need. Examples include the US Department of the Interior's Office of Hearing and Appeals database[17] and the EPA Web Archive.[18] As stated above, it may be necessary to use more than one search engine and review multiple pages of search results.

Binding agreements such as consent decrees might be on file with a federal court. If you know the related case number and court, use the Administrative Office of the US Courts Public Access to Court Electronic Records (PACER)[19] to locate the case docket. If the agreement was filed in the late 1990s or later, the full text may be available for

download. Otherwise, contact the appropriate Federal Records Center to request a copy of the physical file.[20]

While PACER does not allow for full-text searches of federal court dockets, this functionality is available via subscription databases such as Westlaw[21] and Lexis.[22] Many paid databases also allow subscribers to search and browse various specialized documents; an example is Lexis's EPA Consent Decrees.[23]

Agency guidance may be submitted to accompany Congressional testimony and can usually be found by searching the US Congressional Serials Set,[24] which contains nearly all reports and documents of the US Congress. In addition, the regulations.gov repository site[25] contains agency guidance used as "supporting and related material" in the course of federal rulemaking.

Not Everything is Online

Not every document is electronically accessible, even if it has been cited recently in the Federal Register or by an issuing agency. The document may have never been made public in an electronic format; this can apply to older as well as recent guidance. Sometimes agencies purposely purge documentation from the internet. In addition, not every website or posted document is indexed by a search engine, and not all sites have been archived and made accessible via the Wayback Machine.

In these instances, it may be necessary to contact a special or repository library. The collection of the National Archives and Records Administration (NARA) ("the nation's record keeper"[26]) can be searched online.[27] It should be noted, however, that only 1% to 3% of all documents and materials created in the course of business conducted by the US federal government are "so important for legal or historical reasons that they are kept [forever]."[28] WorldCat may identify alternative owning institutions. Some libraries listed in WorldCat will lend items or provide copies directly to a requestor; others require that interlibrary loan (ILL) requests be placed through a public or academic OCLC member library. In addition, there are regional archives that specialize in maintaining collections specific to a company

or an industry. Finally, it may be necessary to contact an agency directly, through its own library (e.g., US EPA),[29] or via a Freedom of Information Act (FOIA) request.[30] Any of these options may take time and require payment.

Conclusion

Locating hard-to-find government documents can be challenging, but the content you uncover may be persuasive, informative, and legally binding. While there isn't a uniform method for conducting such research, and more than one technique may need to be employed, the time taken may be time well spent. CL

Andrea L. Hamilton is the knowledge management and research manager for Davis Graham & Stubbs LLP— andrea.hamilton@dgslaw.com.

Coordinating Editor: Robert Linz, robert.linz@ colorado.edu

NOTES

1. https://archive.org/web.

2. https://archive.org/about.

3. Using the Wayback Machine, https://help.archive.org/hc/en-us/articles/360004651732-Using-The-Wayback-Machine.

4. Refine web searches, https://support.google.com/websearch/answer/2466433?hl=en.

5. *Id.*; Advanced search options, https://help.bing.microsoft.com/#apex/18/en-us/10002/-1.

6. *Id.*

7. https://www.oclc.org/en/about.html.

8. https://www.worldcat.org/whatis/default.jsp.

9. https://www.worldcat.org/advancedsearch.

10. PURL, https://www.oclc.org/research/areas/data-science/purl.html. Introduced by OCLC, PURLs are "[w]eb addresses that act as permanent identifiers in the face of a dynamic and changing Web infrastructure." The Internet Archive has managed the PURL service since 2016.

11. https://www.hathitrust.org/about.

12. HathiTrust U.S. Federal Government Documents Program, https://www.hathitrust.org/usgovdocs.

13. https://catalog.hathitrust.org/Search/Advanced.

14. https://www.hathitrust.org/u-s-federal-documents-collections.

15. Search—A Basic Guide, Full-Text Search, https://help.archive.org/hc/en-us/articles/360018359991-Search-A-Basic-Guide#FTS.

16. Advanced Full-Text Search, https://babel.hathitrust.org/cgi/ls?a=page;page=advanced.

17. https://www.oha.doi.gov:8080/index.html.

18. https://archive.epa.gov.

19. https://www.pacer.gov/findcase.html.

20. https://www.archives.gov/frc/locations.

21. https://www.westlaw.com.

22. https://advance.lexis.com.

23. Finding EPA Consent Decrees on Lexis Advance, http://lexisnexis.custhelp.com/app/answers/answer_view/a_id/1086658/~/finding-epa-consent-decrees-on-lexis-advance.

24. US Congressional Serial Set, https://www.govinfo.gov/help/serial-set; searchable via HathiTrust (https://babel.hathitrust.org/cgi/ls?a=srchls;c=148631352;q1=*), HeinOnline (https://home.heinonline.org/content/u-s-congressional-serial-set), or Lexis (http://lexisnexis.custhelp.com/app/answers/answer_view/a_id/1088011/~/finding-congressional-documents-in-the-u.s.-serial-set-on-lexis-advance).

25. https://www.regulations.gov.

26. https://www.archives.gov/about.

27. https://catalog.archives.gov.

28. https://www.archives.gov/about.

29. EPA National Library Network, https://www.epa.gov/libraries.

30. https://www.foia.gov.