

Jade Kastel

**Name:** Library of Congress, Web Archives

**URL:** <https://www.loc.gov/websites>

**Creator:** LOC, US Government

**Name:** The National Library of Australia, Web Archives

**URL:** Trove <https://trove.nla.gov.au/website>

Pandora is

**Creator:** AU Government and The National Library of Australia

Comparing the search functions in the web archives of the Library of Congress and the National Library of Australia was very helpful in learning how a user can navigate, search, limit parameters, learn the policies for privacy & rights, request archive of their own web pages, data mining, update notifications, and access to archived websites.

Beginning with the searching parameters, both allow search by URL and keyword, however, the LOC allowed users to search in the main library catalog search bar to retrieve archived websites, and the NLA did not allow that function in the main library catalog search bar. You needed to go to the NLA web archive Trove, and then you could conduct a search within the web archive.

For a Boolean search in the LOC and the NLA, I feel like this should be possible an advanced search option, however, limiting by a boolean search was not initially intuitive if it is possible. I was not able to limit using not, minus, or symbols like + etc.

Both the LOC and the NLA allow users to narrow down the search results by date. In the LOC, I was not able to use the internal search bar within a specific archived website. I could click around in the Library of Congress and use “Control F” within the archived page I was viewing. However, I could not search within the search bar of the handful of archived websites that I previewed. Perhaps this function is active on some archived websites at the LOC, and maybe this is linked to the depth of desired archival access the creator specified in the initial crawl. In the NLA, I was able to search within some of the archived pages I opened, however, I’m wondering if that function is disabled on some archived web sites depending on the importance of the site and the parameters of the crawls.

In the LOC web archive, I was able to search by web content characteristics including date, subject, contributor, language, original format (webpage or legislation), online format i.e. web page, PDF, image, online text, audio, country location, part of what other website, and if the archive is access available online or only at the LOC. In the NLA, this feature is also possible to limit a search by update frequency, genre, media type, however, the parameters are listed across the left side, and also across the top. In the NLA, there’s also an option to do ADV for advanced search, similar to ones I’ve seen on databases or library catalogs. For the LOC, there was no real distinction from a basic or an advanced search, mainly because the parameters to do an advanced search are already visible on the left side panel, similar to ordering something off Amazon. In the NLA, they have many unique ways to narrow a search which may be ideal for researchers in the end, but the learning curve was higher on the NLA web archive.

I did not see a safe search to filter adult content on either the LOC or the NLA. Maybe this is an add on a user could apply from a browser. I did notice on the NLA that when I accessed an Indigenous archive,

there was a pop up that warned me that there was content of people who had died, and that may be traumatic. I thought perhaps there is a similar pop up warnings for adult content, or violent content, and perhaps I just didn't open an archived site with those metadata flags to give the warning.

The search results are ranked based on relevancy to the query for both the LOC and the NLA. For the LOC you can rank the search by alphabetical title, date, shelf order, and ascending or descending order for any of those parameters.

There is a hierarchical presentation of the results in both the LOC and the NLA. You search for a web site, and then you enter that archive. From there, in the LOC you can generally select dates from a calendar, or in the NLA, you see the main page and the calendar is a toggle plus sign that allows a user to migrate from date to date. Sometimes, the website has multiple snapshots from a given day. For example, I previewed an election campaign website on the LOC, and one day had 25 snapshots from just that single day, where as other websites had one snapshot a month on average.

For the most part, users can see all archived versions of a URL available online, but some were available only by going to the LOC in person or the NLA. This is related to privacy, copyright, and security clearance. On the LOC, there's a rights tab on the bottom of every archived web site.

Both the LOC and the NLA allow users to see the dates and time which each URL is archived, they provide content summaries, users can be informed of when a page has changed and what parts have changed, export a summary of the difference between the two archived pages, and navigate the website as if it were live on the internet. For the most part the web archives appeared as if they were live, however, I was not able to use the search function within the websites I previewed. I could click around the website, and maybe other archived sites have a live search function, but the ones I previewed did not have that activated.

Each LOC and NAL page had a persistent identifier and it's clear users know that they are browsing archived web because of ample marketing saying this is an archived web site, and the web site does not open to take up the whole screen, which may end up causing confusion.

It seems almost 100% of the website's functions and the archived sites were available. However, there was a limiter on the LOC left hand side that denoted which sites can only be accessed by a user onsite at the LOC. Perhaps this is for security clearance reasons.

Both the LOC and NLA have policies related to requesting the institution to archive a website or requesting a website to be taken down. A simple Google search for LOC archival requests pulls up a list of frequently asked questions and links to the official forms to fill out for those requests. This does not guarantee an archive will be removed.

The NLA allow users to create a personal account, however, the LOC does not provide this at the public level, but perhaps this function is available to researchers or government employees. Both archives have an Ask A Librarian tab which is very helpful when first navigating each archive.

The NLA actively notifies users (who have created an account) about updates in the web archive by email and to users whose sites are archived.. The LOC extends this update to sites they are archiving. The LOC and the NLA notify new site of the institutions intention to archive.

For data mining, both the LOC and NLA allow a function to return graphs illustrating how certain archived websites associated with certain events in a period time. When opening an archived site on the LOC, there's a graph showing the time it's been archived, which can be associated with given events in a time period. NLA has an icon to click on to show all the snapshots recorded on a given site.

The web archives on the NLA and the LOC can provide linking information such as incoming, outgoing, and internal links. The LOC web archive allows users to share, save, email, and extract portions of an archive based on limiter of time span, name, file, metadata, etc. NLA also allows users this access, however a user needs to create an account to access these privileges.

After you create an account with the NLA web archive users can extract a subset of the archive or process and analyze data in the web archive. On the LOC, users can extract subsets of data, but if they can process and analyze the data on the web archive, it appears that's only available by special permission.

Reconstruction of a lost website is possible at both the NLA and LOC, by request. Web site owners can petition via government form for the reconstruction of a website, just as they can petition an archived web site can be taken down or accessed only with a security clearance.

The LOC and NLA both provide many of the same features for searching, exporting, policy, and access. The main difference as a new user was getting accustomed to each archive interface and exploring the parameters of how to narrow searches, download information, and create a user account.