A Content Manager’s Guide to [redacted]

Last updated [redacted]

This guide is intended for [redacted] staff as well as anyone who uses more specific or complicated functionality and tools within [redacted]’s [redacted] build. It is **not** a comprehensive “how to write for web” or “how to use a CMS” guide.

To understand more about [redacted]’s core functionality, you can check out the official user guide. Note that [redacted]’s core uses can differ greatly from how [redacted]’s [redacted] build works.

For more on [redacted]’s visual patterns and design language, check out the Design System.

For editorial style guidance or other content-specific questions, please refer to the [redacted] Voice and Style Guide or the Content Types Guide on the intranet.

Otherwise, use the table of contents below to make your way around.

Table of Contents

Something broke!

If you run into what appears to be either a bug with the [redacted] software, an error or bug in how [redacted] looks or functions, or any other kind of issue related to either, you have two places to reach out:

Talk to Chuck

Email Chuck Sebian-Lander, or reach out to him on [redacted] or [redacted]; he should have the most knowledge about intended functionality, possible bugs introduced by ongoing work, and context for any problems you may run into.

Submit a bug report to [redacted]

If Chuck can't help or is unavailable, you can also submit a bug report to the [redacted] team. That team is responsible for the [redacted] platform as well as [redacted] and should be able to help with troubleshooting and correcting any urgent problems, or else assessing and prioritizing less urgent work to make sure it's eventually picked up and fixed.

I just need to publish a blog/press release...

We have specific documentation designed for [redacted] users who exclusively use the system for off-hours, emergency publication of short items such as blogs or press releases. This can be good also as a broad reference for how to stage and publish items in the CMS.

You can find that documentation here:
[redacted] **Onboarding and Training (Simple, Urgent/Off-Hours Requests)**

General Page Editing

These rules apply to **any content entered into [redacted],** regardless of page type or context. They're most useful as a way to think about and process content that's been copy-pasted into [redacted] content fields from another source, typically one of our Word-based content templates.

Spacing

Remove any extra spaces between paragraphs or between sentences. Our page templates should ignore the extra white space when rendering the page, but it can lead to inconsistency or lack of clarity when you're editing content.

Headings

Make sure to format provided text with appropriate headings. If you get text with **bold lines** that are not explicitly headings, change them to headings! Headings should always be H2 unless they are sub-headings, in which case they should always be H3s (etc). Headings are important for accessibility. Screen reader users rely on headings to accurately read the content.

Links

Ensure that all links to pages within https://[redacted]are relative links **unless** they use anchor links (i.e. have an # in the URL). This means that the URL should omit the domain name, i.e. start with / instead of https://[redacted]/.

Ensure that all external links are formatted properly (look out for the safelinks URL strings that come when someone copy-pasted a link from an internal email; you need to click through on these hyperlinks and take the URL from the page that ultimately resolves).

Anchor links

Custom anchor links can be set using the "anchor" option in the rich text field menu (choose the anchor icon), or by using the "content with anchor" block. Headings are automatically slugified; for example, an H2 titled "My heading" will auto-generate the slug `my-heading`. Note that this functionality is limited to lowercase alphanumeric characters only, and spaces are replaced with dashes, so the heading "Here's an Info Unit Group!" will render as the slug "heres-an-info-unit-group". (For longer headings, you can use Google Chrome to inspect the heading and paste the id from the code.)

Once anchor links are set, you can link to them on the same page by using the "Anchor link" option in the link chooser, then copying-and-pasting the slug into the first field shown. Note that you can also anchor link to expandable group headings, expandable child item headings, and info unit group headings.

Note that auto-generated anchors can sometimes be faulty or fail to render; if you need to use an anchor that isn't generating in this way, you can overwrite the automatic anchor using the "anchor" rich text field option.

SVG Icons

[redacted] has its own iconography set defined as part of the Design System. You can add these anywhere there is an "Icon" field **or** into any rich text editor using the "Icon" option:

[image]

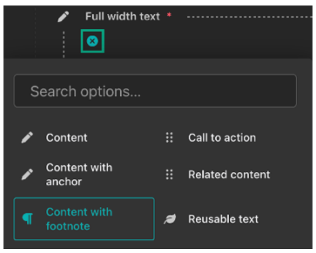
Clicking that button will bring up a pop-up window where you will add in the **canonical name** of the icon you want. You can find the list of icons and their corresponding canonical names in the Design System. Note that if you want e.g. an icon within a circle, just append "-round" to the icon's canonical name (e.g. "warning-round" to see what's been added in the screenshot above).

Schema

If content is written in the form of an FAQ or a how-to guide, consider using the built-in “FAQ schema” or “HowTo” schema blocks. These blocks help Google read our content more easily for better SEO and provide some consistent pre-baked formatting.

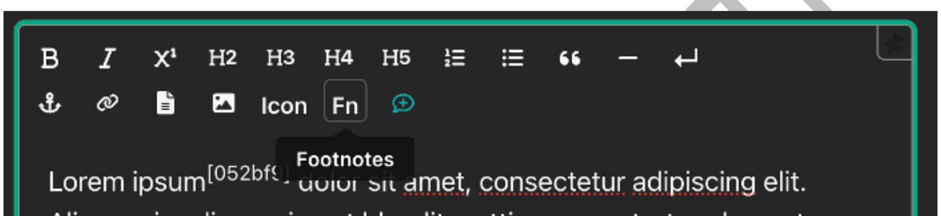
Please review the fields on these blocks to better understand their intended formatting and limitations.

Footnotes

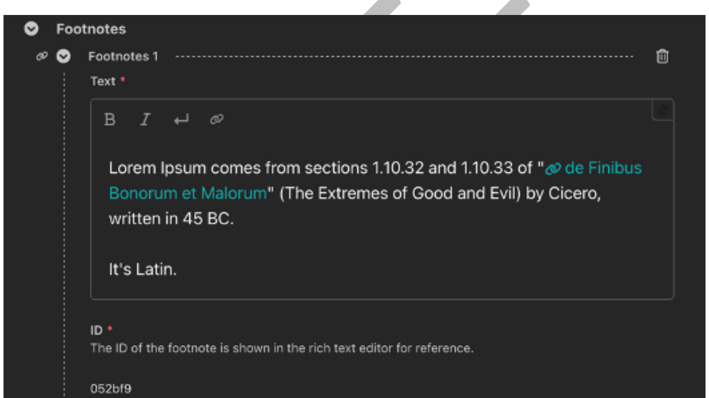


We now have dedicated footnote-adding functionality in [redacted]! This will automatically set up your “Footnotes” section at the bottom of the page, number things appropriately, and add “back to top” links and other accessibility and user-experience niceties.

You can add footnotes to a page that supports them by first selecting the “Content with footnotes” block. Within this block, in the rich text formatting options, you’ll see a Fn button:



Using this button won’t do anything until you first add your footnotes. At the bottom of any page that supports footnotes, you’ll find a “Footnotes” section:



You can add and remove footnotes as needed and format them accordingly.

Once you’ve added footnotes, using the Fn button within the rich text editor will display a list of all existing footnotes, from which you can choose the appropriate item. These receive a unique ID string within [redacted] but will be rendered with standard numbering on the page.

Some other notes for using this functionality:

- A footnote added to the "Footnotes" section will only actually appear on the staged/live page if it's being referenced somewhere in the content.
- You can have two consecutive footnotes (e.g. two notes at the end of this line[2, 3]) with proper formatting by placing a comma between them and then using superscript formatting on the whole chunk (both footnotes and the comma) to ensure all the spacing is correct.
- Do not reference the same footnote more than once—instead create a new/duplicate footnote object and link to the second one. This ensures ordering and linking order work properly on the final page.
- If you **copy** a page with existing footnotes, you're going to create some weird issues re: duplicate footnote items. Get around this by deleting the footnotes from the newly copied page and recreating them as needed.

Tables

We use [redacted]'s built-in block for tables, which is relatively intuitive but has a few issues that are worth noting:

- Header cells must be **plain text only**
- Tables must be square (i.e. number of rows = number of columns)
- Cells cannot be blank; if you must add a cell without content, instead add "-"
- It can be hard to see content for multi-column content because of how limited the editor space is; you can use keyboard tabbing to navigate cells or the horizontal and vertical scroll bars available at the edges of the block

One custom function we've added to tables is the ability to **copy-paste content** from Word or Excel documents, which is hugely important for content management since we aren't typically building tables from scratch. Look for the "Paste from clipboard" option within a Table block to use this feature.

Some functionality notes about copy-pasting tables:

- The pasting option works in Chrome only (not Firefox or Safari).
- Pushing the button for the first time will trigger a Chrome "allow/deny" prompt for access to your clipboard. Make sure to select "allow"!
- Pasting is plain text, i.e. content will be stripped of formatting, so you'll need to go through and re-add any links or other formats to cells once the pasting is done.
- There are two paste options: plain text and rich text. **Only use plaintext if you know you won't need formatting on cells in the table**. Otherwise, though the initial paste will still be text only, rich text will give you all the normal flexibility you may need.
- If the paste fails, there is no prompt or notice; your table simply won't be generated. Usually you can solve for this by first taking your copied table content, pasting it into Excel, and standardizing it as described above (i.e. make sure header cells are plain text, that tables are square, and that all cells have some kind of content).

Translation and Internationalization

Internationalization on [redacted] is a work in progress. However, we have some functionality to support the automation of cross-linking and reporting on translations of pages into various languages.

At the very bottom of the Configuration tab is a section titled TRANSLATION with two fields: Language and English Page. The first should be set to whatever the native language is of the current page. If the page is a non-English translation of an English page on the site, then English Page should be set to the corresponding English page.

When page(s) are published with this field filled out, it will automatically generate a set of links on each page to each translation of that page. As a content manager, you shouldn't need to take any other steps!

Documents and Files

Filenaming conventions

Broadly speaking, files should follow this convention for naming (parts in bold are placeholders depending on the specifics of your file):

- Lowercase only
- No spaces – use dashes and underscores
- General filename structure: [redacted]**_general-topic _specific-file-name_20YY-MM**

Having both the "general" and "specific" parts of the filename is optional depending on whether it makes sense for the file, e.g. enforcement actions will always have this, by virtue of the case name and the specific file type.

There are instances where files should not include dates because they are evergreen, e.g. [redacted] files or images.

When in doubt, use recent precedent (i.e. similar documents uploaded in the past 2-3 years).

Optimization and accessibility

General design needs

If you have any questions, issues, or needs regarding design assets (images, documents, etc) you can reach out [redacted] for support. Their team is in part tasked with supplying fast-track and ad hoc support for design-related requests that come through the [redacted].

Image optimization

Try not to upload images that are more than 1200px in width. Use a program or website like TinyPNG to ensure that the image has been optimized as much as possible. Broadly speaking, an image uploaded to [redacted] should never need to be more than ~500KB in size, depending somewhat on its complexity.

Accessibility

PDFs need to be made 508-compliant before uploading. Checking whether a document is accessible should be up to a combination of the content owner, the Section 508 office, and the Office of Public Affairs; when in doubt, copy in the Public Affairs inbox and ask whether the document is compliant.

Images need alternative text if they provide any visual information or have text embedded within the image. Again, content editors are responsible for providing this alternative text; please ask for it if you don't receive it. If content editor has provided it, alternative text is usually in the Word doc under the image or right click the image and click "View alt text."

Our web components and patterns should already be 508-compliant by virtue of being derived from our Design System. If any issues with this are flagged or noticed by you or a stakeholder, please reach out to [redacted].

Page Editing and Building

This guide includes information about how to build certain page types or how to edit certain specific existing pages. Each section has its own "last updated" date, so make sure to check that against the timeline for [redacted] upgrades and let your friendly neighborhood content management team know if anything needs to be updated!

Page Types

Blog

Content

Make sure to format provided text with appropriate headings. If you get text with **bold lines** that are not explicitly headings, change them to headings! Headings should always be H2 unless they are sub-headings, in which case they should always be H3s (etc).

Ensure that all [redacted] cross-links are relative links **unless** they use anchor links (i.e. have a # in the URL).

Ensure that all external links are formatted properly (look out for the safelinks URL strings that come when someone copy-pasted a link from an internal email; you need to click through on these hyperlinks and take the URL from the page that ultimately resolves).

Sidebar

Always include, in order:

1. Related posts
    a. Limit: 3

    b. Show Heading and Icon? Checked
    c. **Slug Title:** Further reading
    d. Blog Posts: Checked
    e. Newsroom: Checked
    f. **Events:** Checked
    g. **Specific categories:** Post category **or** Unselected
    h. **Tag filtering:** ANY topic tags
    i. Alternate "View more" URL: blank
2. Email signup
    a. Choose email sign up: Blog ([redacted])

Configuration

Authors will always be listed in alphabetical order **by last name**. This cannot be modified.

Newsroom

Content

Read our general page editor content guidance!

Press releases typically begin with [redacted]. Make sure that this text is:

- Boldface
- All-caps
- [redacted]

Just copy the text above and paste in place of existing text if you're unsure.

Always include, beneath any provided content:

- Reusable text
    - **Choose reusable text:** Press release disclaimer

Sidebar

Always include, below any provided sidebar content:

- Email signup
    a. **Choose email sign up:** Newsroom ([redacted])

Event

Event pages are a little weird. The key thing to remember is that they will display three distinct content sections based on the timing relative to the event's start and end dates:

- **Future:** Only shown before the event's start date
- **Live:** Only shown during the event
- **Archive:** Only shown after the event's end date
- **Persistent:** Always shown

You may not receive information for all three; if not, you can request it, or you can just use the "persistent body" field, or you can copy what you have across them accordingly. (I'd recommend modifying the text for the Archive body to be in past tense; otherwise make no changes.)

When an Event ends, it is automatically archived, which in this case means the page itself is moved to the Archived Events section of the site and [redacted] automatically generates a redirect for the past URL.

Document Detail Page

When you're dealing with any kind of page that uses the "Document detail" page type, there will be a variety of possible sidebar items and subheadings to add. The best way to ensure consistency is to **use precedent**.

Use site search, [redacted]'s internal search, relevant listing pages, your own memory, or any other methods to find either a previous iteration of your page or an equivalent style of content. Either copy the page entirely as a basis for your new page, or use it as a guide for how to phrase and style the sidebar, subheadings, and so on.

Over time, the intention is that more of these pages will become more heavily structured within [redacted], requiring less recreation of elements such as the above. In the meantime, though, just rely on what exists on the site as your best resource for how things should be structured.

Outside of this, **in general**:

- Use H5s to set off lists of downloads or resources
- Make sure the category is clearly set and use "Item introduction" in the header block so that the category appears above the H1

[redacted]

URL slug

[redacted] pages are automatically created with a special bit of text at the end of the URL, e.g. **-en-1234**. Do not ever change or remove this! It ensures that Ask pages and their corresponding translations are handled properly by the system.

Spanish pages

When you create a new [redacted] question, a Spanish-language blank page draft can be edited and published using the "EDIT SPANISH PAGE" button at the bottom of the English page's editor. **Do not create the Spanish page independently;** you won't be able to cross-link it to the English page properly.

Setting breadcrumbs

The breadcrumb at the top of an [redacted] page is determined by its [redacted] tags. Specifically:

1. If a single [redacted] tag is selected, that topic's landing page will be the breadcrumb.
2. If a **primary** [redacted] tag has been selected, that topic's landing page will be the breadcrumb.
3. If multiple [redacted] tags have been selected and no primary tag is selected, then the "[redacted]" landing page will be the breadcrumb.
4. If no tags are selected, then "[redacted]" will be the breadcrumb.

Never attempt to move an [redacted] page as a way to change the breadcrumb; this will break its visibility and functionality across the site. (If you forget and do move it, things will be restored as long as you move it back.)

Sunsetting or redirecting pages

When you need to remove an [redacted] question, you need to go through some extra steps to ensure the page is no longer appearing in various automated listings:

- **Unpublish** the page.
- **Move** the page to Trash.
- **Within the Trash, re-publish and then re-unpublish the page.**
- **Clear the site cache.**

Various bits of functionality across [redacted] can break when [redacted] pages are removed, including but not limited to the "related questions" section or the search listings for [redacted].

The best way to clean up after sunsetting pages is to use the [redacted] Indexer to track down stray links to the sunset question *or* to links to "None", which will be caused when some of the built-in linking tools break.

Specific Pages

Homepage

In the [redacted] editor, you will see no options for editing except in the Configuration tab. This is not an error; right now, the homepage's content is modified directly within the codebase.

If you need content updates to the homepage, please reach out to the [redacted]

Org chart

Headshots uploaded for use on the org chart should be square and no more than 1000x1000 pixels, optimized for web as either JPG (preferred) or PNG. Please use the filenaming convention [redacted]**_firstname-lastname_headshot** for all such images.

Make sure that the option "Border radius for images?" is set to "Circle" for any 25/75s on the page.

Using Specialized Print CSS

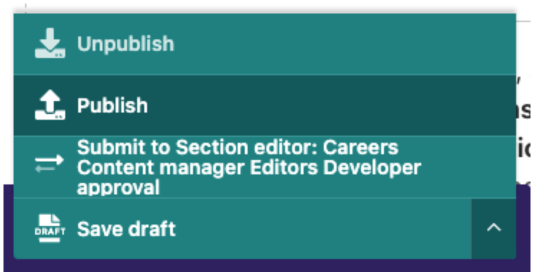
Recently we've added functionality to [redacted] that allows for a more print-friendly version of CSS styling to be added to a page that has been properly prepared for its use. You can find out how to use this styling in our guide here.

Publishing, Un-Publishing, and Deleting

Publishing Drafts

All drafts and draft revisions are visible as though published on [redacted]. In order to publish content and make it live and visible, you must publish that content.

The “Save draft” button at the bottom of the page editor contains extra items when you click the up arrow to the right of the button:



You can ignore any “Submit to Section editor” options; these are part of a workflow we don’t use. Instead, if you’re ready for your content to go live, you can select the **Publish** button. You’ll have to pass through a confirmation screen before your content actually goes live.

Notes about publishing:

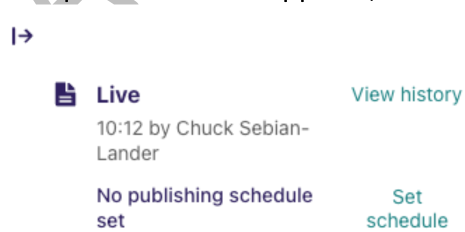
- It is always possible to **un-publish** content, if it was published by mistake or if it has been slated to be removed from the site. If content is meant to be removed permanently from the site, then move it to the trash

Scheduled Publishing

[redacted] has a scheduling tool built in to allow us to set content to go live at a specific time or date. Using this functionality is relatively straightforward. **After you have finished your draft** and are ready to schedule it for publishing, do the following from the page editor for that page:

1. Check the top-right toolbar and select the “Status” icon (it’s a lower-case “i” within a circle): 

2. In the panel that now appears, select “Set schedule”



3. A tooltip will appear titled **Set publishing schedule**. Set your “Go live date/time” as desired in Eastern Time. Unless you want the content to automatically un-publish itself, leave “Expiry date/time” blank.

4. Select Save schedule.
5. Double-check the “Status” panel to confirm that the schedule has been set properly.
6. Select the normal **Publish** button.

Your content will now automatically publish at your selected date/time. You may still need to clear the cache for certain changes to be visible after that point.

Clearing the Cache

Sometimes, pages that have been published will not be immediately visible on [redacted]. More commonly, elements on pages that are updated dynamically—filterable lists or “related posts” sidebar modules, for example—take longer to update even after the new content is live. This is because [redacted] uses a server-side cache that allows for quicker re-loading of pages for users but may be slower to reflect updates site-wide in response to certain changes.

If you want or need this content to be visible immediately, rather than waiting for that cache to clear itself, you can manually purge the previous cache and force a re-build. Note that overusing this may cause some performance issues for end users (nothing major, just don’t push the button 10 times in a minute or anything like that).

To find this functionality, select the **CDN Tools** item in the lefthand sidebar of the [redacted] admin panel:

You’ll be taken to a page that includes a history of cache-clearing actions taken by other [redacted] users, as well as the **Cache Invalidator** that allows you to clear some or all of the existing site cache. If you want to make sure that the cache is properly cleared for all filterable list pages, variations, etc. then just **leave the text field blank** and click on **Invalidate cache**.

It may still take a few moments for this action to propagate. Be patient, wait a few moments, and then click “refresh” in your browser to see if the changes come up. (If the [redacted] page takes longer than usual to load in usual, this typically means the cache was indeed cleared—so if you still have a problem, it’s not a cache issue. Take another look at your published content or feel free to contact your friendly neighborhood content manager.)

Sunsetting and Redirecting

First, set up the redirect (so that it is active as soon as the sunset page has been removed):

- Go to the Redirects panel. (When logged in, click “Settings” on the lefthand navigation, and choose “Redirects”)
- Click the “Add redirect” button.
- For the “From” field, type in the URL of the page to be sunset. Format the URL as a **relative link**:
    - If the URL is https://[redacted]/page-url, the relative URL is **/[redacted]/page-url**

- Under “Redirect to a page”, select “Choose a page” and find the new page to which you are directing this URL.
    - If the redirect includes an anchor tag (#) or if it is to a file or a non-[redacted] page, instead place the **full URL including https** under “Redirect to any URL”
- Click “Save”.

The redirect will not work yet; you now need to **delete the existing page**.

Content Removal and Deletion

If you have permission to edit a page in [redacted], and the page has not been locked, you will also have permission to unpublish or to delete it. You can find the prompt for doing so in a number of places, including the dropdown menu on the editor page: [image]

You can also find this option in the dropdown menu on a parent page’s listing of child pages: [image]

Selecting the deletion option in any of these places will bring up confirmation window(s) to ensure you are deleting what you intend to delete.

Unpublishing vs Deleting Content

It is always harder to quickly restore deleted content vs. restoring unpublished content. If you are not absolutely certain that the content in question will be permanently removed from the site, please unpublish it instead.

In general, you will be safer unpublishing content and then deleting it shortly thereafter once confirming that the live site functions as intended. Unpublishing a page should be sufficient to make the live site function as though the page has been deleted.

Restoring Deleted Content

[redacted] content is continually crawled and archived using the Archive-It service from the Internet Archive. You can visit our site-wide collection to search for pages, including any that were published but have since been deleted, moved, or otherwise altered in ways you might want to revert: [redacted]

(NOTE: If the above Archive-It collection is not working for some reason, you can use the standard Wayback Machine from the Internet Archive. Its collection will not be as consistent but should still generally contain most if not all of [redacted]’s published content.)

As of now, all deleted pages also have their content stored internally as a raw data file. [redacted] users, or other admin-level [redacted] users, have the ability to restore deleted content using these files.

If you do not have access to deleted content files

Contact [redacted] to request the restoration of any inadvertently deleted content. Please specify the URL where the content was located, to the best of your memory, as well as the time at which the deletion occurred.

If you do have access to restore deleted content files

Pages are stored at [redacted] based on their IA placement (i.e. based on the URL path), and specific page files also have a timestamp appended to them corresponding to when they were deleted.

For example, below is a part of the archive storing three different instances of deleted content for the page [redacted]:

[image]

You should be able to use this to find the page you deleted based on its URL and the time of deletion. To download the JSON file, you may have to use the File – Save As… feature in your browser, or else right-click on the link above (i.e. instead of navigating to the json file itself) and clicking “save as” from there.
Once you have downloaded the appropriate JSON file, you can use it to prepopulate a recreation of the page using the new “Import” option in the dropdown menu on parent page listings:

Note that the JSON file may be missing information. You’ll want to review the page generated by the report and compare against what you know or would have expected the page to contain before it was deleted. Feel free to both note where you see this happening and consider if any of the gaps in information feel particularly important or necessary, i.e. you’d be worried about being able to recreate the page properly in the absence of that data.

Content Archival Notices

Archiving in bulk

Bulk archival (as with legacy blog and newsroom items, see example here) isn’t done within [redacted]; we use a banner applied at the level of the HTML template, which vastly reduces the amount of manual labor involved.

This does mean that such efforts need a developer and can’t be done by a content manager. If such an effort needs to happen, get in touch with [redacted] leads on resourcing and prioritizing such a project.

Archiving one-off content

[redacted]’s content management system provides a standard “notification” block, with form fields for title and URLs, that can be used by content managers without developer intervention to provide individual context and links in unique cases of removal or archival of content on the live website.

Example: [redacted]

Design Pattern: [redacted]

How to add the notification

- In the **Header** section of General Content, click the plus icon below any existing introductory block and select **Notification**
- Select of fill in the fields as so:
  - Type: Information
  - Message: Archived content
  - Explanation: Leave blank or fill in depending on what context is being provided for archival
  - Link:
    - Text: “View this page as it was originally published”
    - Aria label: “Archived content of **[page title]**”
    - Url: See “Getting an archived content link” below

Content block when content is removed

If you are removing all current content on the page as part of the archival, make sure to at least leave e.g. a **Content** block (within a **Full width text** block) that reads “Content has been archived” (note the lack of punctuation).

Getting an archived content link

All archived versions of any [redacted] URL can be found using this URL pattern:

https://wayback.archive-it.org/[redacted]

You can fill in a given [redacted] URL from there, and you’ll get an Archive-It page with a calendar showing all the times the weekly crawl has captured the page (if you’ve ever used the Wayback Machine, this will look very familiar). From there you should be able to pick an appropriate page and copy the URL for use in the notification banner

Background on archival using Archive-It

[redacted] maintains multiple internal backups of its content and content database, but also maintains a publicly accessible backup of site content, via **Archive-It**, a tool managed by the Internet Archive.

A link to [redacted]’s Archive-It collections is here: [redacted]

The legacy blog and newsroom archive is a one-time archival of all blogs and newsroom pages pre-2016, which were stripped of formatting in the live content database in order to simplify content and structures that had predated the website’s current content management system. This archive is referenced as part of the URL pattern for these pages’ “archived content” banner as described below.

The ongoing backup utilizes [redacted]’s sitemap file as well as a file list to collect and store up-to-date website contents and PDFs on a weekly basis.

What *can’t* be edited in [redacted]?

Most standard page content and almost all pages are, to some degree, created and editable within [redacted]. However, there are certain pages or aspects of content that cannot be edited within the CMS and require developer assistance.

This is not necessarily a comprehensive list, but should be close.

- Header (i.e. anything above the mega menu)
- Footer
- Page layouts, beyond what is described in the Design System
- Hardcoded pages:
    - Homepage

Site Health Checks

As part of our regular work in the [redacted], we (starting in [redacted]) also regularly audit and remediate typical site errors. Below are some explanations and workflows of how to find and respond to the errors we should be reviewing on a roughly monthly basis, depending on other priorities and general [redacted] bandwidth.

Internal Broken Links

What are they

Internal broken links are hyperlinks on [redacted] pages that attempt to link *to other* [redacted] *pages* but fail. Broken links of this type can take several forms:

- **"None" links:** These are generated when a link that's using [redacted]'s built-in **Internal Link** option breaks because the linked page was moved to Trash or otherwise no longer exists as [redacted] expects it to.
- **Incorrect domains:** Links that try to directly point to "[redacted]", or [redacted], or any of our other non-www domains.
- **404s:** Any other links that end up taking users to a 404 error page.

How to find

You can use the [redacted] web page indexer to gather all live instances of these errors:

- **"None" links:** Search using the "Links" option for "None", or just use this link, and download the resulting report. Note that you may get a few false positives (for URLs that happen to have "none" within the string) so make sure to review results carefully.
- **Incorrect domains:** Search using the "Links" option for "[redacted]", "[redacted]", and "[redacted]", and download the resulting reports.
- **404s:** Scroll down to the "Discover more" well at the bottom of the crawler, and select the "Download website errors (including 404s)" report.

How to fix

Broken link remediation is a somewhat subjective process of determining the best solution for each case. Here are some general guidelines you can follow:

- Broken link fixing **does not require staging or** [redacted] **review/approval**; as long as there are no corresponding significant content changes, you can just make the fixes and publish them immediately.
- Incorrect domains should be straightforward—simply correct the link to the www domain. If the resulting link is still broken, follow steps below.
- "None" links will require some sleuthing, as you often won't have a clear idea of where the link was *intended* to go in the first place. If you can't determine a solution from context, you can try using the Wayback Machine to find a historical version of the page where the link was valid.

- On old time-stamped pages – e.g. older blogs or press releases, meaning 3+ years – you can remove the link entirely rather than tracking down an alternative.
- If a solution is unclear, but you know who the stakeholders would be for the content (e.g. if the content is related to [redacted], the stakeholders should be within [redacted]), you can send the link in question to those stakeholders directly and ask if they have recommendations for fixing.
- If a solution is unclear and you don't know who the stakeholder is, ask Chuck to take a look. He can either determine the proper stakeholders or make an executive decision on how to handle the link.

External Broken Links

What are they

External broken links are hyperlinks on [redacted] pages that attempt to link *to non-*[redacted] *pages* but fail. Broken links of this type can take several forms:

- **Domain resolves to a 404:** You still end up at the right domain, but within that domain you get a 404 error.
- **Link doesn't work at all:** The domain doesn't exist at all; you get a generic "page canno be found" error in your browser.

How to find

Right now the only methods for collecting these errors is external reports, e.g. public reports via the [redacted] email address or stakeholder requests. Once we know that a particular URL leads to a 404, we can use the web page indexer to find all instances on [redacted] that link to that page.

Wyatt developed a script that can run against the indexer's data to produce a report of all external links that produce an error code. We are planning to integrate this script into the indexer itself or otherwise provide a way to automate the generation of this report so you can download it as easily as the internal broken link report that's currently available.

How to fix

Broken link remediation is a somewhat subjective process of determining the best solution for each case. Here are some general guidelines you can follow:

- Broken link fixing **does not require staging or** [redacted] **review/approval**; as long as there are no corresponding significant content changes, you can just make the fixes and publish them immediately.
- If the error report comes from a stakeholder, they should also provide the valid URL we should put in place of the error. Always double-check the broken link using the indexer to make sure there are no missed instances of the error that remain.
- On old time-stamped pages – e.g. older blogs or press releases, meaning 3+ years – you can remove the link entirely rather than tracking down an alternative.
- Google is your friend. Try using context from the link text and the URL to track down possible clear alternatives.

- If a solution is unclear and you don't know who the stakeholder is (e.g. the report came from a public user who didn't provide a solution), ask Chuck to take a look. He can either determine the proper internal stakeholders or make an executive decision on how to handle the link.

Redirects

What are they

We have a lot of redirects in place in [redacted], set up for a variety of reasons. Redirects can cause page-load and SEO problems if they're improperly set up:

- **Unused redirects:** Many redirects get less use over time or are never used to begin with; these take up unnecessary space in our database and can end up accidentally triggering some of the below problems as more redirects get added.
- **Redirect loops:** We can create redirect loops, or chains, if we have one URL redirecting to another URL that redirects to yet another URL from there. This makes pages take longer to load and disrupts Google's understanding of what the "correct" URL is, hurting SEO.
- **Internal redirects:** Sometimes the only place where we're using a redirect is in a link on our own website. We shouldn't do this; it's optimal for page load and user experience to ensure we link directly to the correct, final URL.

How to find

120-day server log: On a daily basis, the server log of redirects is automatically downloaded and added to this section of a Github repository. This is a log of redirect hits over the past 120 days. You'll want to scroll to the very bottom to download the most recent file.

[redacted] report: [redacted] has an option to export all existing redirects within the Redirects admin panel (click the dropdown arrow next to "Import redirects" at the top right, and select export).

How to fix

There are some extra steps you need to take using a tool like Excel to analyze the data from the above sources to find errors.

Unused redirects

(Find a sample of the described Excel work here, for reference. You should be able to take the formula column, highlighted green, and move it into the new reports once set up as described.)

Take the 120-day server log report and place it in an Excel tab. Then add [redacted]'s redirect report as a second tab.

Create a new column in the [redacted] tab. For each cell, use an XLOOKUP function to check whether the "from" URL exists in the server log.

Every item that comes up "false" should be slated for removal; just go right into [redacted]'s redirects panel, find the redirects in question, and delete them.

Redirects deleted in this way should be logged in the tracker for sunset redirects, which you can find here.

Redirect loops

(Find a sample of the described Excel work here, for reference. You should be able to take the formula column, highlighted green, and move it into the new reports once set up as described.)

Take [redacted]'s redirect report and add a new column. For each cell in this column, do an XLOOKUP function checking the "to" column against the "from" column. Essentially, we're checking whether any of these items appear both as a redirected URL *and* as the destination of a redirect.

For any that end up "true," go into [redacted] and modify any "from" columns so they only direct to the valid/final URL.

Internal redirects

(Find a sample of the described Excel work here, for reference. You should be able to take the formula column and move it into the new reports once set up as described.)

Take [redacted]'s redirect report and place it in a tab. Then, take the indexer's report of all hyperlinks on the site and place it in a second tab. (NOTE: The indexer report is going to be very large and slow to work with in Excel. Be patient.)

Add a new column to the [redacted] report tab. For each cell, use an XLOOKUP function to check the "From" column of the [redacted] report against the "link_url" column of the indexer report.

If you use the sample file's formula, the output will be only the **first URL where the internal redirect is used**. Therefore, for each of these, use the indexer to find all page(s) where the internal redirect is being used, and change it so the link is directly to the valid/final URL.

SEO Fields

What are they

Every [redacted] page has fields that allow us to customize the meta title and description field, which are used by Google as the title and short description of the page when it appears in its search results. (These fields are also used for our internal searches and listings, such as the Blog and Newsroom landing pages.)

There are some rules for optimizing the description field in particular, but in general we just want to make sure these fields are present across all our site content.

How to find

[redacted]'s internal Page Metadata report includes the search description field in column I. Note that this report can take a while to generate and download; just be patient after clicking the "Download CSV" button.

How to fix

Just filter column I of the Excel report for blanks, and remediate case-by-case. Keep in mind the following:

- If the page has any kind of short, single-sentence description at the top ahead of other content, just use that to populate the field.
- Search descriptions should generally not exceed **150** characters, but use your best judgment and/or work with what's available on the page.
- If you're unsure how to populate this field for a given page, ask Chuck for support.