Introduction May 10, 2024:

My name is Carolyn Holzman and I've been a search professional since 2009. In addition to being a working seo, I have developed an expertise in Google indexing and forensic SEO and in April 2024 released a data-proven hypothesis on the Helpful Content System and what specifically it targets. My analysis establishes through testing and field data evidence that the Helpful Content targets websites based on a topical nucleus by marking multiple types of repeated content as unhelpful. Data provides no evidence that the system looks for or marks any content specifically as "helpful."

"*Decoding Google's Helpful Content System*" is available for purchase on AmericanWayMedia.com.

What my research uncovered is that through no fault of website owners, HCS proves to be an existential threat due to a basic and fundamental feature of the Google system being broken.

It is likely that by implementing what HC updates did learn in its "crawl and learn method" during updates run from August 2022 - January 2023, is the reason why the negative impact on websites was exponentially compounded and felt the most strongly during the last standalone HCU in September 2023 and the March 2024 Core Update with HCU migrated into the core algorithm.

What is broken?

Google's canonical system, specifically its previous ability to attribute and consolidate ranking signals to the canonical version of a page, is no longer providing the ability to correctly attribute jump links to the canonical.

Below is my walk-through on how data supports this conclusion, what the implications are for a broken canonical system and why it is my professional opinion that the decimation of sites due to this issue can not be laid at the feet of site owners or content writers.

Hypothesis:

Google's historical treatment of the canonical is no longer in play, it is broken. If there are jump links in content, they now are treated as individual URLs resulting in each URL individually being indexed, and the creation of each with a separate protocol buffer within the Google indexation system with its own page history within Google's protocol buffer system.

Implications:

If the canonical system is broken, it means that the system can no longer correctly attribute content to the canonical URL. The inability to attribute to the canonical has multiple implications:

Duplicate Content: The same content may be indexed under multiple URLs instead of being consolidated under the canonical URL. This results in duplicate entries in Google's index.

Diluted Ranking Signals: Without proper canonical attribution, ranking signals such as backlinks, user engagement, and other SEO metrics are spread across multiple URLs. This dilution can weaken the overall ranking strength of the content if search engines like Google might not be able to consolidate these signals to the canonical version.

Potential SEO Penalties: Search engines may penalize sites for having duplicate content, which can negatively impact the search rankings of the affected pages.

Helpful Content System: If true, this would create a devastating impact for sites in SERPs ruled by a Helpful Content System that sees duplicate content as unhelpful and tags each of those individual URLS as unhelpful content.

Background:

Google has provided information on how they handle and determine canonicals.
> Often called deduplication, this process helps Google show only one version of the otherwise duplicate content in its search results.
>
> ...indicating a canonical preference is a hint, not a rule. Source - [Google](#)

Google also provides detailed recommendations to webmasters to avoid duplicate content.

> To consolidate signals for similar or duplicate pages. It helps search engines to be able to consolidate the signals they have for the individual URLs (such as links to them) into a single, preferred URL. This means that signals from other sites to https://example.com/dresses/cocktail?gclid=ABCD get consolidated with links to https://www.example.com/dresses/green/greendress.html if the latter becomes canonical. Source - [Google](#)

Jump links, or anchor links, refer to links that direct to specific parts of a page, rather than the entire page itself. This includes a standard practice of adding Table of Contents (TOC) to enhance user experience.

Jump links are also referred to as fragment links, text fragments or anchor links. And may include any link fragments such as FAQ jump links.

Protocol Buffer Canonical URL Handling Process

[Google Protocol Buffer,](#) commonly referred to as Protobuf, is a language-neutral, platform-neutral, extensible mechanism developed by Google for serializing structured data. It's used extensively by Google for various internal systems and applications, including those involved in content indexing and management.

If a canonical URL is present, the system understands that the content should be attributed to the canonical URL rather than the current URL. If working it would prevent duplicate content from appearing in the search index and consolidate ranking signals to the canonical version.

Conventional Treatment When Using Jump Links

For websites to communicate to Google to treat the canonical and the jumps as the same page, it is customary to set the rel canonical. Like so.

```
<link rel="canonical" href="https://www.domain.com/url-slug/" />
```

Jump Links Are Generally Considered Good for SEO These Links Lead To Reader Satisfaction

I did find one published test by One Up Web published a test that they had run to answer the question - [Are Anchor Links Good for SEO](). The general consensus though is not based on actual data from specific sites but the general understanding if someone is looking for something specific, providing a shortcut to what they are looking for must be good since it satisfies the reader.

Examples of a jump link

```
https://www.americanwaymedia.com/#indexation
https://developers.google.com/search/docs/crawling-indexing/canonicalizati
on#canonical-how
```

Jump links in Search Console under normal conditions

1. If canonical attribution is working, they should not be getting impressions in SERPs
2. If canonical attribution is working,they should not be ranking in SERPs
3. If canonical attribution is working,they should not be getting clicks from inside SERPs
4. Search Console is reveals jump link URLs are receiving impressions, rankings and clicks

Random observation

Jump links as described in #4 above showed up in December 2022 - given that by the time this was noticed, the 16-month window closed and thus the opportunity to gather screenshots had passed.

Notes:

This write up will refer to every type of link that links to a section of a page as jump links.

The most likely source of validation of the hypothesis is to look at sites that have Table of Content links on most of their pages.

If this hypothesis is true the only way we can determine that is by the data and behavior in the SERPs. It is possible that Google is unaware of this condition. It is also possible they are aware of it. Determining which is beyond the scope of this write up.

Google Search Console goes back as far as 16 months. Every day we move forward that time window shifts the same. Any data not gathered prior to that last available date is "lost" to us to use it.

The concepts I refer to in this write up regarding the HCS are based on my analysis of the HCS which is available in the guide - "*Decoding Google's Helpful Content System*" AmericanWayMedia.com. It provides additional context but this report is a standalone analysis of the current canonical issue.

The Helpful Content system encompasses a larger set of content issues that are beyond the scope of this report.

Executive Summation

- If the canonical on a site-level is working, jump link URLs would not compete against the canonical URL. Data reveals they are competing.
- If Google now attributes jump link URLS as separate pages, fighting for their own position for terms, you could see a growth in impressions without a significant growth in click rate. This occurred approximately from March - end of July 2023. (see composite image below)
- In a Helpful Content environment with the canonical not able to provide the historical "protection" of each of these URLS against becoming duplicate content, the canonical and each jump link related to the canonical would be considered duplicate pages.
- When Google sees duplicate content on the domain, multiple URLs that lead to the same content, if the canonical is not attributed all the page data, then the benefits are split between all the URLs for the same page. Instead of the canonical winning, the "juice" is split and the winner is determined by which URL or jump link is most relevant to the query.
- The harm of Google seeing each jump link as a separate page with a separate history is that Google's Helpful Content Classifier would "tag" each duplicate page with an "unhelpful" tag. Given that if the majority of pages on a site contain varying numbers of jump links, the math would very quickly determine that a site consists with a majority of the pages of unhelpful content. For example, 700 canonical pages along with 3500+ jump link pages tagged as unhelpful content, would overwhelm the site as a majority of unhelpful content and effectively ensures virtual extinction of the site in SERPs.
- Those pages with the unhelpful tag would start to drop out of the serps and drag the canonical URL down with them. It would appear the HCS would see a site 100% of "unhelpful" content.
- The site within GSC would appear in the aggregate view to be flat lining on impressions, clicks, traffic and ultimately revenue.

This is what it looks like to have this issue. Isolating impressions makes it easier to see. This site suffered catastrophic loss of impression, rankings, clicks and ultimately revenue.

1. Inexplicable increase in Impressions between spring and summer 2023 -
2. Hard drop after Sept 2023 HCU
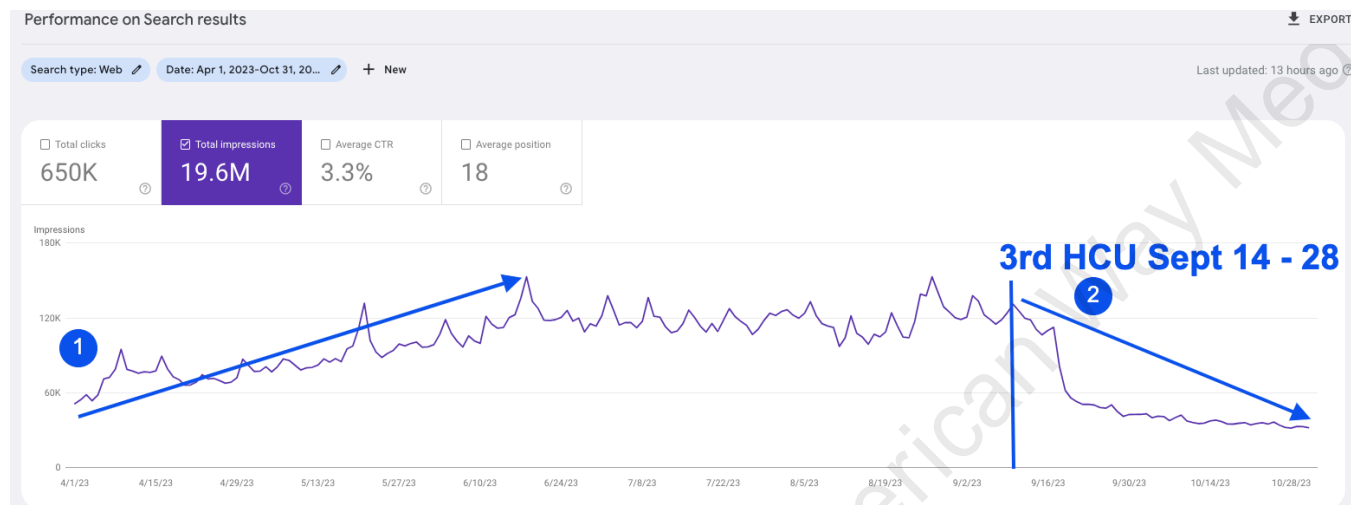
1. Pages have canonical set
2. Jump links are getting impressions by ranking as individual pages within index
3. Site has suffered catastrophic loss of traffic and revenue since September HCU and continues to flat line further since March Core update
4. Publisher has done exactly as Google Developer documents recommend by setting rel="canonical" in the source code of the header on every page

Field Evidence:

The most reliable source for site evidence is within search console data. Outside search documented incidents provide additional context.

Search Console Data:

Site: Ecommerce Sports Education Niche
Uses Table of Contents - sampled number of jump links in TOC ranged from 7 to as high as 25
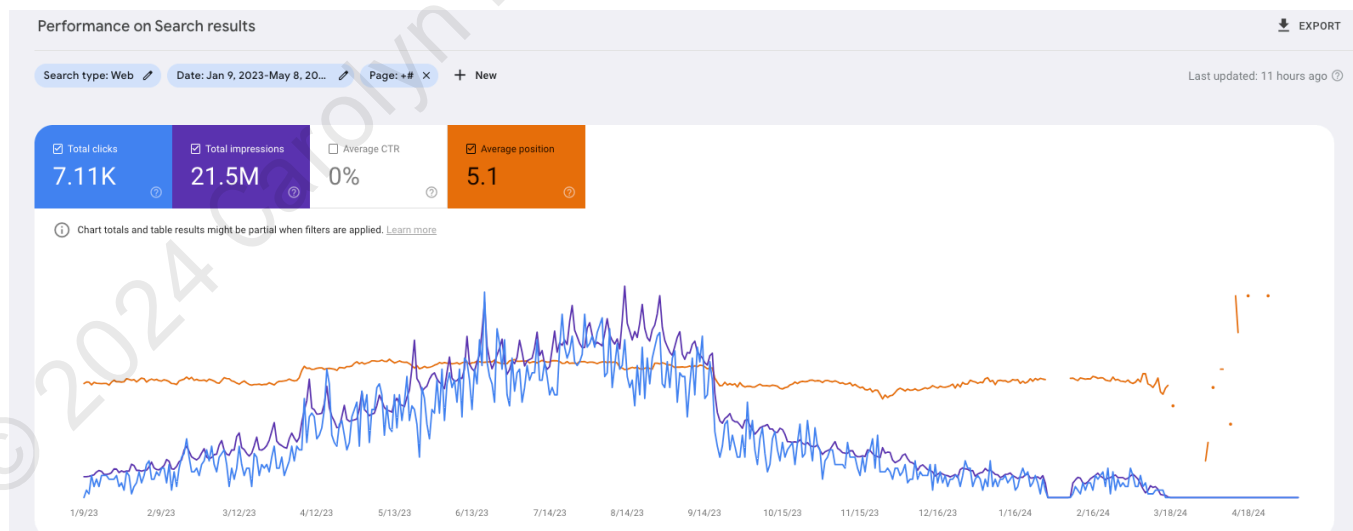All URLs have rel="canonical" set on each page.
Approx 700 pages
All composite screen shots are from Jan 9, 2023 - May 8, 2024 - full 16 months

Composite of all pages on site.
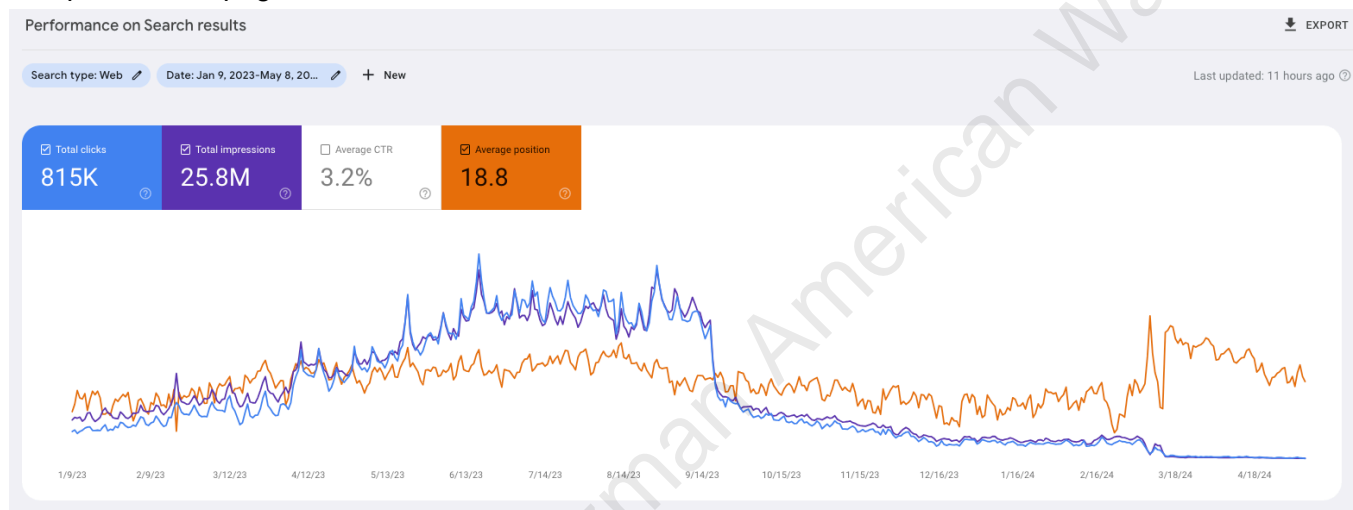
Same timeframe only for pages that have "#" in the URL that are getting impressions, clicks and rankings.

It makes sense that less than 1% of overall clicks were driven from jump links.

Selecting 3 individual pages from the collection of pages with jump links, searching for all URLs with the same URL across the same timeframe.
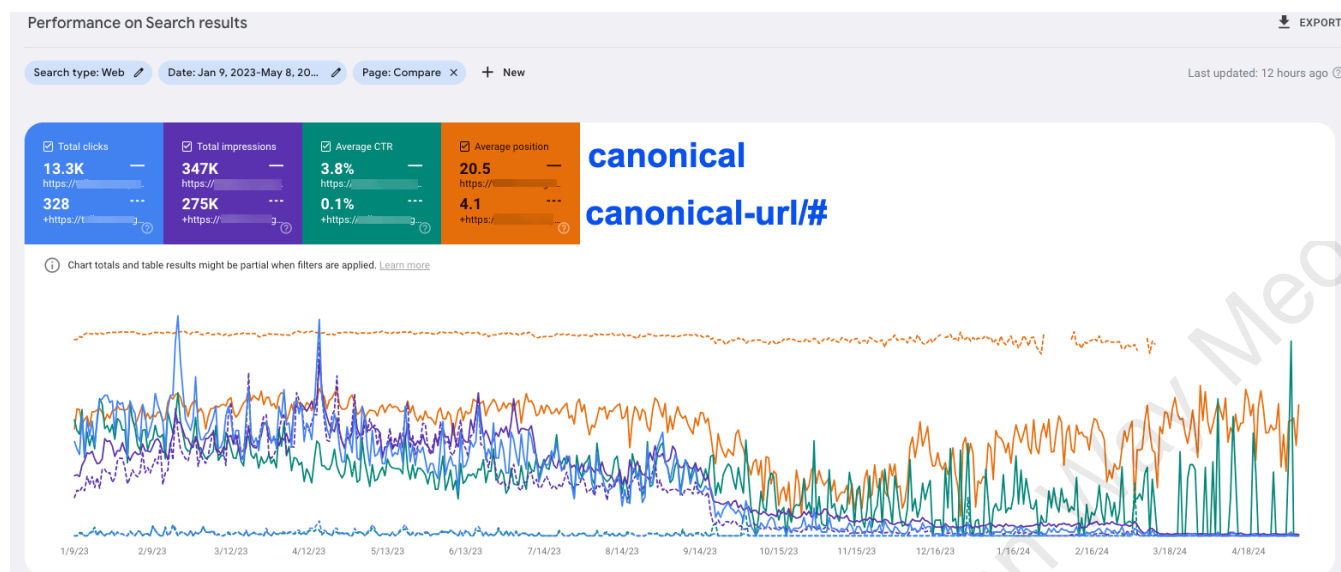
URL #1 - canonical page plus 14 jump links

Data visualization based on all URL #1 urls - canonical and 14 jump links = 15 URLS to this page
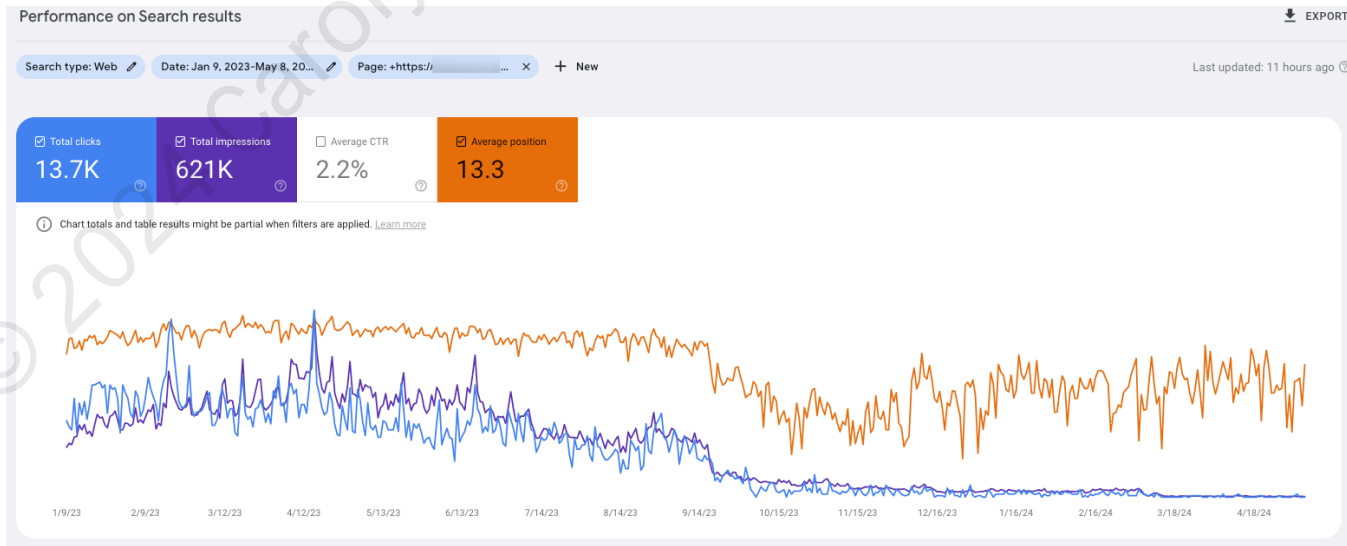
All 14 jump links for URL #1 were ranked higher for position than canonical. Can confirm it visually in the average position data.

URL #1 after Sept 2023 Helpful Content Update - October 1 - Dec 31, 2023

After March Core Update with HCU now part of Core March 5, 2023 - April 30, 2024

URL #2 - canonical page plus 6 jump links = 7 total URLS

6 jump links outranked canonical - added in positions
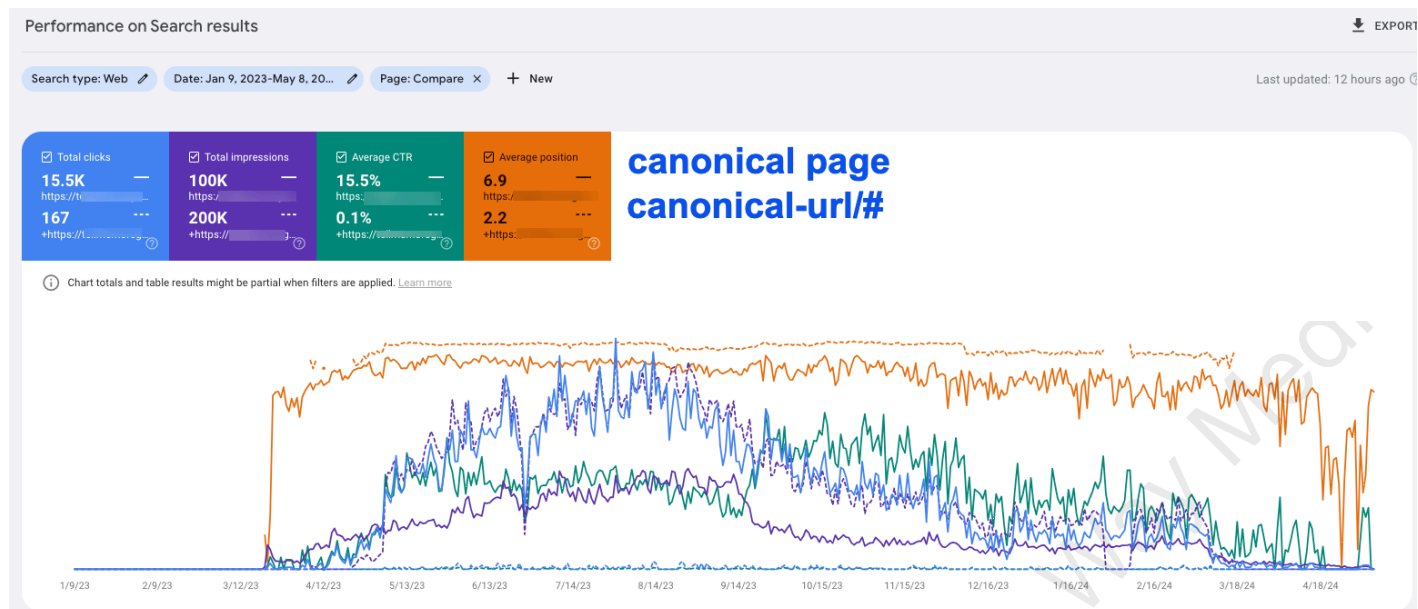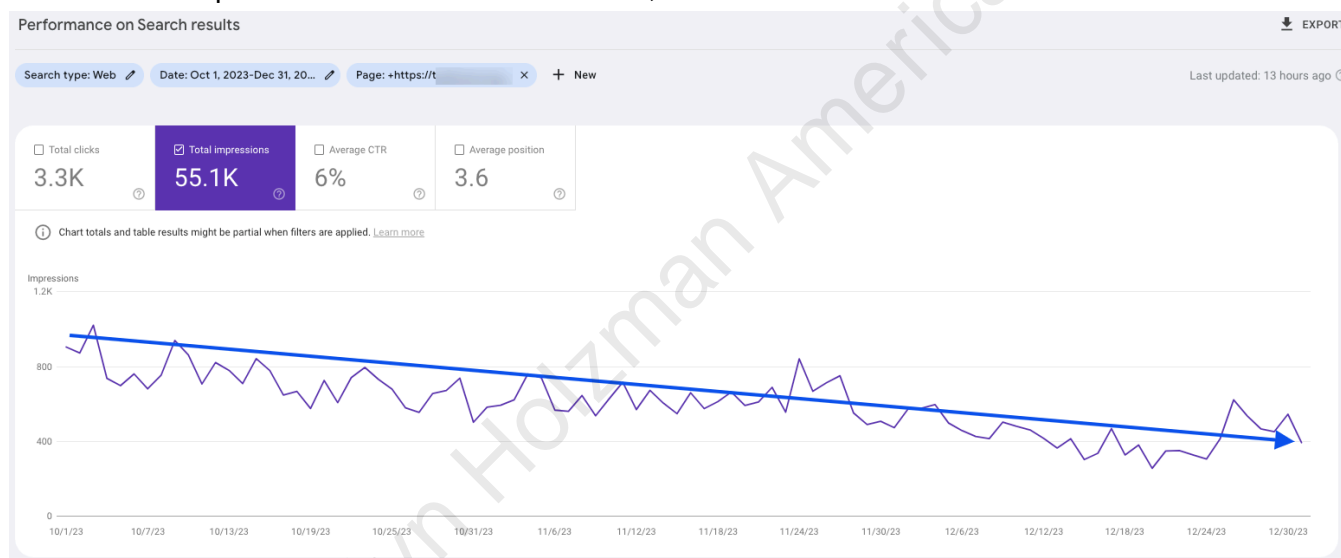

Data for all URL #2 URLs together - canonical and 6 jump links of URL #2 = 7 URLS to this page

Comparing Canonical data to Jump Link Data for URL #2 - again jump links are out ranking canonical

canonical page
canonical-url/#

URL #2 after Sept 2023 HCU October 1 - Dec 31, 2023

URL 2 after March Core Update with HCU now in core algorithm - March 1 - April 30, 2024

Performance on Search results

Search type: Web ✏ | Date: Mar 1, 2024–Apr 30, 2... ✏ | Page: +https://▓▓▓▓▓▓ : | + New

Last updated: 13 hours ago ⍰

☐ Total clicks	☑ Total impressions	☐ Average CTR	☐ Average position
235	6.29K	3.7%	7.8

ⓘ Chart totals and table results might be partial when filters are applied. Learn more
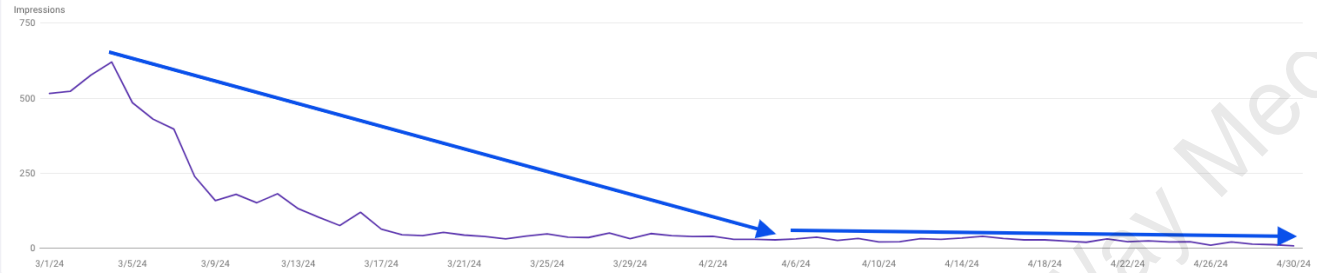
URL 3 -Canonical plus 7 URLS
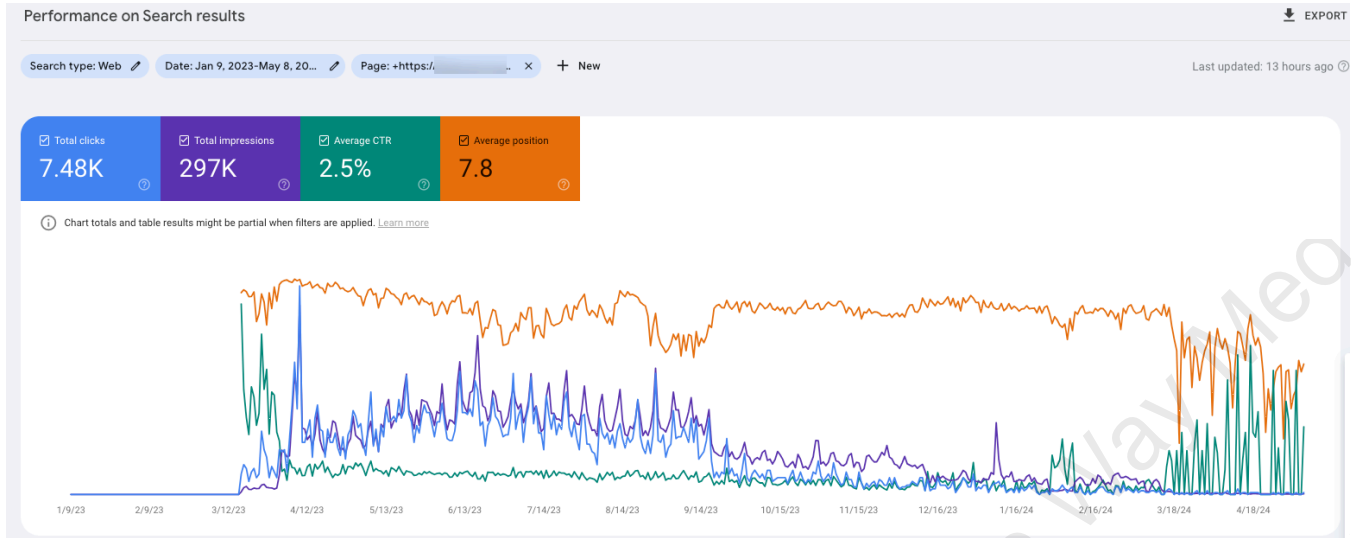
URL #3 comparing canonical to jump links - again jump links outrank canonical

URL 3 after Sept HCU - HCU October 1 - Dec 31, 2023

Performance on Search results

Search type: Web · Date: Oct 1, 2023-Dec 31, 20... · Page: +https://t... ×

Total clicks	Total impressions	Average CTR	Average position
3.3K	55.1K	6%	3.6

URL 3 after March Core Update with HCU now in core algorithm - March 1 - April 30, 2024

Performance on Search results

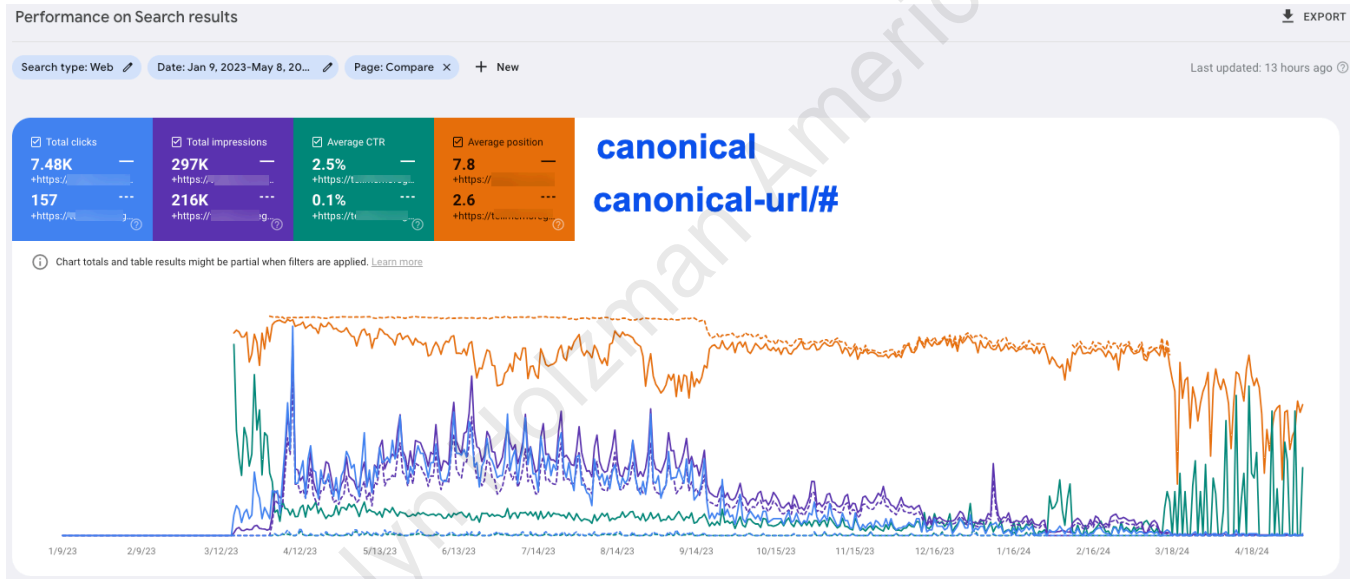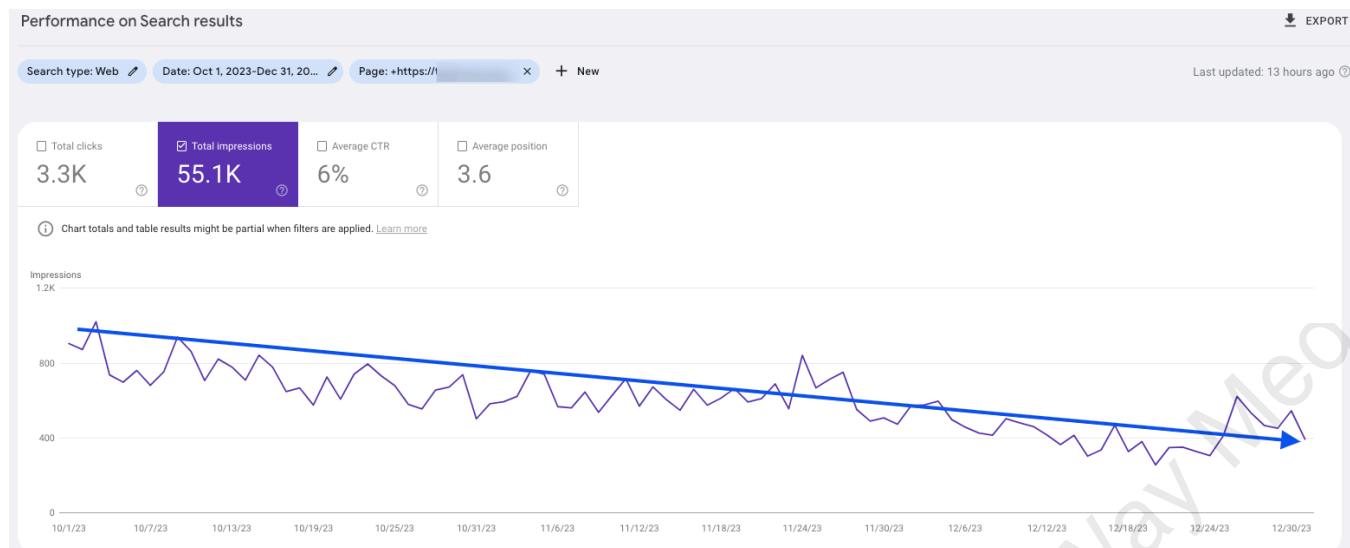Search type: Web · Date: Mar 1, 2024-Apr 30, 2... · Page: +https:/... ×

Total clicks	Total impressions	Average CTR	Average position
235	6.29K	3.7%	7.8
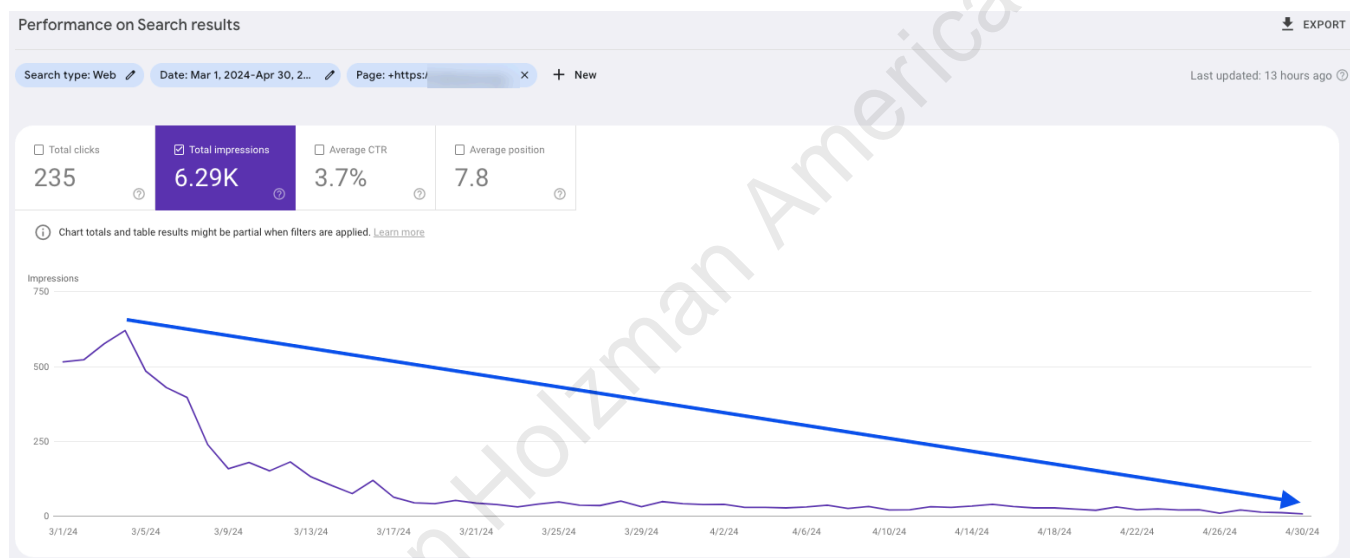
What outside of SERP incidents could implicate an issue with the canonical apparatus?

Three events or incidents provide some additional context - retirement of the parameter tool in Google webmaster tools, updating canonicalization documentation in early 2023 and the Google News Indexing incident in June 2023. I'll walk through each of them.

Google URL parameter tool retired

The largest suspect in my estimation is the [retirement of Google URL Parameter Tool](#) in April 2022. They did indeed remove it from Google webmaster tools on April 26 2022. It was referred to as "spring cleaning" and that previously it was needed but now Googlebots were more sophisticated and didn't require those instructions.

> Going forward you don't need to do anything to specify the function of URL parameters on your site, Google's crawlers will learn how to deal with URL parameters automatically. Source
> –[Google](#)

The possible scenario of what might happen was discussed on [Confessions of an SEO](#) when I invited Ted Kubaitis to talk with me about possible outcomes. At the time it was thought that this would impact ecommerce sites using parameters on product URLs.

Google updating canonicalization recommendations in early 2023

It might be a complete coincidence that after sunsetting the parameter tool they started to update the documentation for canonicalization. Given that canonicalization is a deduplication method, it makes sense with the introduction of HCS.

February 2023
https://developers.google.com/search/updates#2023

- **February 2**: Refreshed our documentation about canonicalization. To better help site owners, the original documentation is split in three distinct sections:
    1. [What is URL canonicalization](#).
    2. [How to specify a canonical with `rel="canonical"` and other methods](#).
    3. [Fix canonicalization issues](#).

May 2023
https://developers.google.com/search/updates#2023

- **May 2**: Updated our [documentation about canonicalization](#) with explicit recommendations for syndicated content.

Repeatedly, Google confirms in their documentation that canonicals are "suggestions" to Google, not commands and Google will choose the page it FEELS is the correct canonical.

Google News Indexing Incident June 21 - July 14, 2023

During early summer there was a significant issue regarding syndicated content from enterprise news sites. This incident was not officially recognized by Google when it began to impact enterprise news sites. The issue concerned news articles originally published on the source news site and then syndicated the same article in the rss feed to their partners.

In the SERPs, once the content was syndicated, the partner site was attributed with the canonical (not the one stipulated within the rss from the original publisher). This [issue was on the search status dashboard](#) for 23 days, 14 hours.

At the time, in my conversations with Charles Taylor (Foxnews Director of SEO) about the issue, he confirmed that the final recommendation from Google was news sites syndicating their content were requested to no longer include their canonical in their feeds - to instead leave them out and request partners using their content to set the syndicated to noindex on the partner's site.

Twitter Conversations from SearchLiaison Danny Sullivan

https://twitter.com/searchliaison/status/1770867218059800685

[Danny Sullivan] Stop trying to "show Google" things. I have been through so many sites at this point (and I appreciate the feedback), and the patterns are often like this:

> - Something saying an "expert" reviewed the content because someone mistakenly believes that ranks them better
>
> - Weird table-of-content things shoved at the top because who knows, along the way, somehow that became a thing I'm guessing people assume ranks you better

Which Incident Most Likely Confirms State of Canonical System

When evaluating evidence from one side, meaning none of us can see into Google we only witness items publicly, it's still important to note what could be seen which is why I included all three. But if I had to pick the most likely to confirm the hypothesis from these three, it would have to be the news indexing incident from the summer.

1. It happens around the time when impressions for sites with many jump links started to increase
2. It directly involved the canonical question

3. The resulting "solution" from Google didn't solve the source issue - the question remains open, why did the canonical rel tag become moot on the rss feeds?

Conclusion

1. Evidence confirms that pages with multiple URLS to the same content are retaining their individual attributions even though canonical tags are used.
2. Evidence shows steep loss of impressions, rankings and serve rates within the searchable index after each Helpful Content Update and the March Core Update w/HC migrated over.
3. The webmasters followed the instructions from the Google Developer Documents by placing the rel="canonical".
4. Webmasters and content writers have no direct control over the internal systems that govern the attributions of the canonical system and can not be responsible to fix it.
5. Google developers have direct control over internal systems that govern the attributions of the canonical system and thus are responsible to fix it.

Takeaway

Until Google fixes what is broken, sites will continue to be decimated within the searchable index and with their content, their revenue in this Helpful Content System-run SERPs.

This also contributes to why there have been no recoveries to these sites after updates.

Site owners until this report, even if this data can be seen in their search console data,continue to assume that the canonical feature could never be the issue because Google Developer documents say that this is how the canonical works and Google would not permit something so fundamental to remain broken within its systems.

The nature of the harm to site owners and their businesses indicates that for short term protection of their site and revenue may require complete removal of jump links from independently hosted sites until Google's processing systems are able to again attribute all jump link URLS to the canonical.