

How To Make Search Console Work Harder For You

TL;DR

1. Search Console has some pretty severe limitations when it comes to storage, anonymized and incomplete data, and API limits.
2. You can bypass a lot of these limitations and make GSC work much harder for you but setting up far more properties at a subfolder level.
3. You can have up to 1,000 properties in your Search Console account. Don't stop with one domain-level property.
4. All of this allows for far richer indexation, query, and page-level analysis. All for free. Particularly if you make use of the 2,000 per property API URL indexing cap.

Image Credit: Harry Clarkson-Bennett

Now, this is mainly applicable to enterprise sites. Sites with a deep subfolder structure and a rich history of publishing a lot of content. Technically, this isn't publisher-specific. If you work for an ecommerce brand, this should be incredibly useful, too.

I and it love all big and clunky sites equally.

What Is A Search Console Property?

A Search Console Property is a domain, subfolder, or subdomain variation of a website you can prove that you own.

You can set up domain-level or URL-prefix-level properties (Image Credit: Harry Clarkson-Bennett)

If you just set up a domain-level property, you still get access to all the good stuff GSC offers. Click and impression data, indexation analysis, and the crawl stats report (only available in domain-level properties), to name a few. But you're hampered by some pretty severe limitations:

- 1,000 rows of query and page-level data.
- 2,000 URL API limit for indexation level analysis each day.
- Sampled keyword data (and privacy masking).
- Missing data (in some cases, 70% or more).
- 16 months of data.

While the 16-month limit and sampled keyword data require you

to export your data to BigQuery (or use one of the tools below), you can massively improve your GSC experience by making better use of properties.

There are a number of verification methods available – DNS verification, HTML tag or file upload, Google Analytics tracking code. Once you have set up and verified a domain-level property, you're free to add any child-level property. Subdomains or subfolders alike.

The crawl stats report can be an absolute goldmine, particularly for large sites (not this one!) (Image Credit: Harry Clarkson-Bennett)

The crawl stats report can be extremely useful for debugging issues like spikes in parameter URLs or from naughty subdomains. Particularly on large sites where departments do

things you and I don't find out about until it's too late.

But by breaking down changes at a host, file type, and response code level, you can stop things at the source. Easily identify issues affecting your crawl budget before you want to hit someone over the head with their approach to internal linking and parameter URLs.

Usually, anyway. Sometimes people just need a good clump. Metaphorically speaking, of course.

Subdomains are usually seen as separate entities with their own crawl budget. However, this isn't always the case. According to John Mueller, it is possible that Google may group your subdomains together for crawl budget purposes.

According to Gary Illyes, crawl budget is typically set by host name. So subdomains should have their own crawl budget if the host name is separate from the main domain.

How Can I Identify The Right Properties?

As an SEO, it's your job to know the website better than anybody else. In most cases, that isn't too hard because you work with digital ignoramuses. Usually, you can just find this data in GSC. But larger sites need a little more love.

Crawl your site using Screaming Frog, Sitebulb, the artist formerly known as Deepcrawl, and build out a picture of your site structure if you don't already know. Add the most valuable properties first (revenue first, traffic second) and work from there.

Some Alternatives To GSC

Before going any further, it would be remiss of me not to mention some excellent alternatives to GSC. Alternatives that completely remove these limitations for you.

SEO Stack

SEO Stack is a fantastic tool that removes all query limits, has an in-built MCP-style setup where you can really talk to your data. For example, show me content that has always performed well in September or identify pages with a health query counting profile.

[embedded content]

Daniel has been very vocal about query counting, and it's a fantastic way to understand the direction of travel your site or content is taking in search. Going up in the top 3 or 10 positions – good. Going down there and up further down – bad.

SEO Gets

SEO Gets is a more budget-friendly alternative to SEO Stack (which in itself isn't that expensive). SEO Gets also removes the standard row limitations associated with Search Console and makes content analysis much more efficient.

Growing and decaying pages and queries in SEO Gets are super useful (Image Credit: Harry Clarkson-Bennett)

Create keyword and page groups for query counting and click and impression analysis at a content cluster level. SEO Gets has arguably the best free version of any tool on the market.

Indexing Insight

Indexing Insight – Adam Gent’s ultra-detailed indexation analysis tool – is a lifesaver for large, sprawling websites. 2,000 URLs per day just doesn’t cut the mustard for enterprise sites. But by cleverly taking the multi-property approach, you can leverage 2,000 URLs *per property*.

With some excellent visualizations and datapoints (did you know if a URL hasn’t been crawled for 130 days, it drops out

of the index?), you need a solution like this. Particularly on legacy and enterprise sites.

Remove the indexation limits of 2,000 URLs per day with the API and the 1,000 row URL limit (Image Credit: Harry Clarkson-Bennett)

All of these tools instantly improve your Search Console experience.

Benefits Of A Multi-Property Approach

Arguably, the most effective way of getting around some of the aforementioned issues is to scale the number of properties you own. For two main reasons – it's free and it gets around core

API limitations.

Everyone likes free stuff. I once walked past a newsagent doing an opening day promotion where they were giving away tins of chopped tomatoes. Which was bizarre. What was more bizarre was that there was a queue. A queue I ended up joining.

Spaghetti Bolognese has never tasted so sweet.

Granular Indexation Tracking

Arguably, one of Search Console's best but most limiting features is its indexation analysis. Understanding the differences between Crawled – Currently Not Indexed and Discovered – Currently Not Indexed can help you make smart decisions that improve the efficiency of your site. Significantly improving your crawl budget and internal linking strategies.

Image Credit: Harry Clarkson-Bennett

Pages that sit in the **Crawled – Currently Not Indexed** pipeline may not require any immediate action. The page has been crawled, but hasn't been deemed fit for Google's index. This could signify page quality issues, so worth ensuring your content is adding value and your internal linking prioritizes important pages.

Discovered – Currently Not Indexed is slightly different. It means that Google has found the URL, but hasn't yet crawled it. It could be that your content output isn't quite on par with Google's perceived value of your site. Or that your internal linking structure needs to better prioritize important content. Or some kind of server or technical issue.

All of this requires at least a rudimentary understanding of

how Google's indexation pipeline works. It is not a binary approach. Gary Illyes said Google has a tiered system of indexation. Content that needs to be served more frequently is stored in a better-quality, more expensive system. Less valuable content is stored in a less expensive system.

How Google crawling and rendering system works (Image Credit: Harry Clarkson-Bennett)

Less monkey see, monkey do; more monkey see, monkey make decision based on the site's value, crawl budget, efficiency, server load, and use of JavaScript.

The tiered approach to indexation prioritizes the perceived value and raw HTML of a page. JavaScript is queued because it is so much more resource-intensive. Hence why SEOs bang on about having your content rendered on the server side.

Adam has a very good guide to the types of not indexed pages in GSC and what they mean here.

Worth noting the page indexation tool isn't completely up to date. I believe it's updated a couple of times a week. But I can't remember where I got that information, so don't hold me to that...

If you're a big news publisher you'll see lots of your newsier content in the Crawled – Currently Not Indexed category. But when you inspect the URL (which you absolutely should do) it might be indexed. There is a delay.

Indexing API Scalability

When you start working on larger websites – and I am talking about websites where subfolders have well over 500,000 pages – the API's 2,000 URL limitation becomes apparent. You just cannot effectively identify pages that drop in and out of the “Why Pages Aren't Indexed?” section.

Not great, have seen worse (Image Credit: Harry Clarkson-Bennett)

But when you set up multiple properties, you can scale effectively.

The 2,000 limit only applies at a property level. So if you set up a domain-level property alongside 20 other properties (at the subfolder level), you can leverage up to 42,000 URLs per day. The more you do, the better.

And the API does have some other benefits:

But it doesn't guarantee indexing. It is a request, not a command.

To set it up, you need to enable the API in Google Cloud

Console. You can follow this semi-helpful quickstart guide. It is not fun. It is a pain in the arse. But it is worth it. Then you'll need a Python script to send API requests and to monitor API quotas and responses (2xx, 3xx, 4xx, etc.).

If you want to get fancy, you can combine it with your publishing data to figure out exactly how long pages in specific sections take to get indexed. And you should always want to get fancy.

This is a really good signal as to what your most important subfolders are in Google's eyes, too. Performant vs. under-performing categories.

Granular Click And Impression Data

An essential for large sites. Not only does the default Search Console only store 1,000 rows of query and URL data, but it only stores it for 16 months. While that sounds like a long time, fast forward a year or two, and you will wish you had started storing the data in BigQuery.

Particularly when it comes to looking at YoY click behavior and event planning. The teeth grinding alone will pay for your dentist's annual trip to Aruba.

But by far and away the easiest way to see search data at a more granular level is to create more GSC properties. While you still have the same query and URL limits, because you have multiple properties instead of one, the data limits become far less limiting.

What About Sitemaps?

Not directly related to GSC indexation, but a point of note. Sitemaps are not a particularly strong tool in your arsenal when it comes to encouraging indexing of content. The indexation of content is driven by how "helpful" it is to

users.

Now, it would be remiss of me not to highlight that news sitemaps are slightly different. When speed to publish and indexation are so important, you want to highlight your freshest articles in a ratified place.

Ultimately, it comes down to Navboost. Good vs. bad clicks and the last longest click. Or in more of a news sense, **Glue** – a huge table of user interactions, designed to rank fresh content in real-time and keep the index *dynamic*. Indexation is driven by your content being valuable enough to users for Google to continue to store in its index.

Glue emphasizes immediate interaction signals like hovers and swipes for more instant feedback (Image Credit: Harry

Clarkson-Bennett)

Thanks to decades of experience (and confirmation via the DoJ trial and the Google Leak), we know that your site's authority (Q*), impact over time, and internal linking structure all play a key role. But once it's indexed, it's all about user engagement. Sitemap or no sitemap, you can't force people to love your beige, miserable content.

And Sitemap Indexes?

Most larger sites use sitemap indexes. Essentially, a sitemap of sitemaps to manage larger websites that exceed the 50,000 row limit. When you upload the sitemap index to Search Console, don't stop there. Upload every individual sitemap in your sitemap index.

This gives you access to indexation at a sitemap level in the page indexing or sitemaps report. Something that is much harder to manage when you have millions of pages in a sitemap index.

Seeing data at a sitemap level gives more granular indexation data in GSC (Image Credit: Harry Clarkson-Bennett)

Take the same approach with sitemaps as we have discussed with properties. More is generally better.

*Worth knowing that each document is also given **a DocID**. The DocID stores signals to score the page's popularity: user clicks, its quality and authoritativeness, crawl data, and a spam score among others.*

Anything classified as crucial to ranking a page is stored and used for indexation and ranking purposes.

What Should I Do Next?

1. Assess your current GSC setup – is it working hard enough for you?
2. Do you have access to a domain-level property and a crawl stats report?
3. Have you already broken your site down into “properties” in GSC?
4. If not, crawl your site and establish the subfolders you want to add.
5. Review your sitemap setup. Do you just have a sitemap index? Have you added the individual sitemaps to GSC, too?
6. Consider connecting your data to BigQuery and storing more than 16 months of it.
7. Consider connecting to the API via Google Cloud Console.
8. Review the above tools and see if they’d add value.

Ultimately, Search Console is very useful. But it has significant limitations, and to be fair, it is free. Other tools have surpassed it in many ways. But if nothing else, you should make it work as hard as possible.

More Resources:

This post was originally published on Leadership in SEO.

Featured Image: N Universe/Shutterstock

<https://www.searchenginejournal.com/how-to-make-search-console-work-harder-for-you/560697/>