Google On Search Console Noindex Detected Errors via @sejournal, @martinibuster

Google's John Mueller answered a question on Reddit about a seemingly false 'noindex detected in X-Robots-Tag HTTP header' error reported in Google Search Console for pages that do not have that specific X-Robots-Tag or any other related directive or block. Mueller suggested some possible reasons, and multiple Redditors provided reasonable explanations and solutions.

Noindex Detected

The person who started the Reddit discussion described a scenario that may be familiar to many. Google Search Console reports that it couldn't index a page because it was blocked not from indexing the page (which is different from blocked from crawling). Checking the page reveals no presence of a noindex meta element and there is no robots.txt blocking the crawl.

Here is what the described as their situation:

- *"GSC shows "noindex detected in X-Robots-Tag http header" for a large part of my URLs. However:*
- *Can't find any noindex in HTML source*
- *No noindex in robots.txt*
- *No noindex visible in response headers when testing*
- *Live Test in GSC shows page as indexable*
- *Site is behind Cloudflare (We have checked page rules/WAF etc)"*

They also reported that they tried spoofing Googlebot and tested various IP addresses and request headers and still found no clue for the source of the X-Robots-Tag

Cloudflare Suspected

One of the Redditors commented in that discussion to suggest troubleshooting if the problem was originated from Cloudflare.

They offered a comprehensive step by step instructions on how to diagnose if Cloudflare or anything else was preventing Google from indexing the page:

“First, compare Live Test vs. Crawled Page in GSC to check if Google is seeing an outdated response. Next, inspect Cloudflare’s Transform Rules, Response Headers, and Workers for modifications. Use curl with the Googlebot user-agent and cache bypass (Cache-Control: no-cache) to check server responses. If using WordPress, disable SEO plugins to rule out dynamic headers. Also, log Googlebot requests on the server and check if X-Robots-Tag appears. If all fails, bypass Cloudflare by pointing DNS directly to your server and retest.”

The OP (orginal poster, the one who started the discussion) responded that they had tested all those solutions but were unable to test a cache of the site via GSC, only the live site (from the actual server, not Cloudflare).

How To Test With An Actual Googlebot

Interestingly, the OP stated that they were unable to test their site using Googlebot, but there is actually a way to do that.

Google’s Rich Results Tester uses the Googlebot user agent,

which also originates from a Google IP address. This tool is useful for verifying what Google sees. If an exploit is causing the site to display a cloaked page, the Rich Results Tester will reveal exactly what Google is indexing.

A Google's rich results support page confirms:

> *"This tool accesses the page as Googlebot (that is, not using your credentials, but as Google)."*

401 Error Response?

The following probably wasn't the solution but it's an interesting bit of technical SEO knowledge.

Another user shared the experience of a server responding with a 401 error response. A 401 response means "unauthorized" and it happens when a request for a resource is missing authentication credentials or the provided credentials are not the right ones. Their solution to make the indexing blocked messages in Google Search Console was to add a notation in the robots.txt to block crawling of login page URLs.

Google's John Mueller On GSC Error

John Mueller dropped into the discussion to offer his help diagnosing the issue. He said that he has seen this issue arise in relation to CDNs (Content Delivery Networks). An interesting thing he said was that he's also seen this happen with very old URLs. He didn't elaborate on that last one but it seems to imply some kind of indexing bug related to old indexed URLs.

Here's what he said:

> *"Happy to take a look if you want to ping me some samples. I've seen it with CDNs, I've seen it with really-old crawls*

(when the issue was there long ago and a site just has a lot of ancient URLs indexed), maybe there's something new here…"

Key Takeaways: Google Search Console Index Noindex Detected

- Google Search Console (GSC) may report "noindex detected in X-Robots-Tag http header" even when that header is not present.
- CDNs, such as Cloudflare, may interfere with indexing. Steps were shared to check if Cloudflare's Transform Rules, Response Headers, or cache are affecting how Googlebot sees the page.
- Outdated indexing data on Google's side may also be a factor.
- Google's Rich Results Tester can verify what Googlebot sees because it uses Googlebot's user agent and IP, revealing discrepancies that might not be visible from spoofing a user agent.
- 401 Unauthorized responses can prevent indexing. A user shared that their issue involved login pages that needed to be blocked via robots.txt.
- John Mueller suggested CDNs and historically crawled URLs as possible causes.

https://www.searchenginejournal.com/google-on-search-console-noindex-detected-errors/540829/