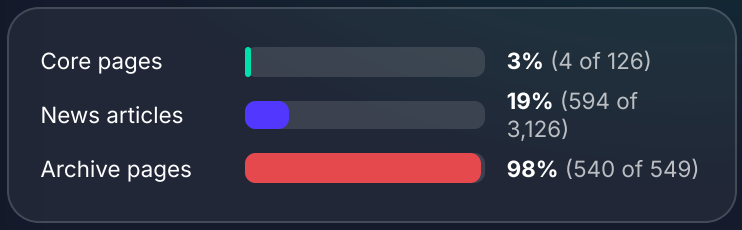
What Google Knew About 3,801 Pages

A crawl of Google's "About this result" panel for every URL on a large, news-heavy information site in a regulated niche. Which pages Google recognised, which it had never seen, and what it knew about the site itself.

Archive study · Site anonymized

SHARE OF URLS GOOGLE SAID IT HADN'T SEEN, BY SECTION





3,801	30%	0
URLs checked	"never seen" by Google	site descriptions found

Ameneh Saeednia
Senior SEO and Generative Search Strategist

Summary

Key findings

In late 2023, a crawler requested Google's **"About this result"** panel for every URL on a large, news-heavy information site and saved what Google said. The panel answered two questions: **what does Google know about this site**, and **has it seen this page**.

1,138
URLs (**30%**) Google said it **hadn't seen before**

97%
of core pages lacked this panel message

98%
of archive pages returned this panel message

0
URLs where Google found a **reference description** of the site

1

No source description displayed

On all 3,801 URLs, the panel reported no matching reference description. This records the panel display, not the presence or absence of an entity in Google's systems.

2

Core pages differed from archives

122 of 126 core pages lacked the "hasn't seen" message. This is not a verified index-status result.

3

Archives had more panel gaps

540 of 549 paginated archives returned the "hasn't seen" message. Crawl cost, traffic and ranking effects were not measured.

4

Older URLs had more gaps

The message appeared for up to 31% of older articles, versus under 3% in newer ID groups. IDs are approximate age indicators.

Why this matters

The archived panel exposed differences in the messages displayed for page groups. It did not establish which URLs were indexed, why messages differed, whether content had been forgotten, or how Google represented the site as an entity.

Method

How the data was collected

Item	Detail
Site	One large, news-heavy information site in a regulated niche (anonymized)
Size	**3,801** unique URLs, all on the site's main domain
Date	Late 2023
Request	For each URL, Google's "About this result" view: a search for *About [URL]* with the "ilp" result type
Captured	HTTP status (all **200 OK**) and the panel text
Analysis	November 2026, from the archived export

Only two messages appeared

"Google looked for a description of [Site] on a reference site, but couldn't find a good match." On every URL. This is the site-level *About the source* line.

"Google hasn't seen this page before, but found some info about the site." On 1,138 URLs. This is the page-level line.

What "hasn't seen" means

This is a message shown by one panel at one time. It is not a definitive index or crawl status, and does not expose every record Google holds. Check indexing separately with Search Console URL Inspection.

Segment	How it was defined	URLs
Core pages	Everything outside /news/ and paginated archives: guides, reviews, law pages, product and payment pages, about pages	126
News articles	/news/[slug]/[post ID]/	3,126
Author archive pages	/author/[name]/page/N/	257
News archive pages	/news/page/N/ and category archives /news/.../page/N/	292

Finding 1
No matching source description was displayed

Every one of the 3,801 panels carried the same site-level line: Google **looked for a description of the site on a reference source and couldn't find a good match**.

What the panel is looking for

The panel may display a source description from Wikipedia or other available context. A missing match does not prove that Google has no entity record, no trust signals, or no knowledge of the organization.

Review public identity information for accuracy and clarity, without inferring a ranking penalty from this message.

Useful identity checks

A clear **About page and editorial policy**, stating ownership and purpose.
Organization markup with consistent `sameAs` profiles.
Independent coverage: interviews, citations, press, industry directories.
Where notability allows, a **Wikidata item**. (See Lessons 9 and 11.)

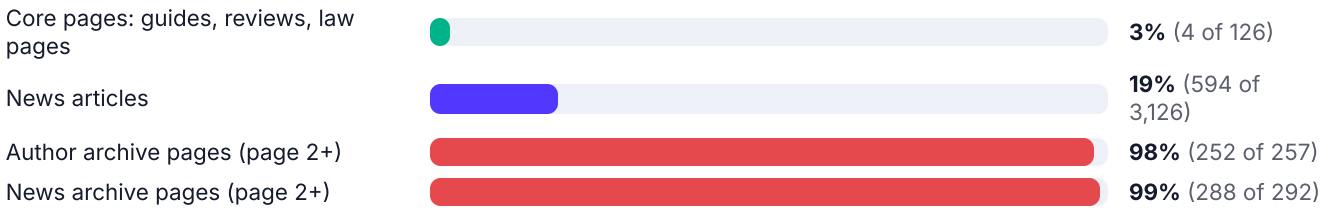
The takeaway

A missing source description is a prompt to inspect public identity information. It is not an entity-status test or evidence of a ranking disadvantage. A Wikidata item or markup does not guarantee a panel or citation.

Finding 2: Panel messages differed by section

The observed message varied by page group; verified index status was not measured.

URLs Google said it hadn't seen, by section



Core-page panel observations

122 of 126 core pages lacked the message. It appeared for three author-profile pages and one payments guide. Actual indexing of these pages was not verified with Search Console.

Archive-page panel observations

540 of 549 paginated archives returned the message. The crawl does not establish whether Google indexed them, which discovery paths it used, or whether archives were retained or excluded.

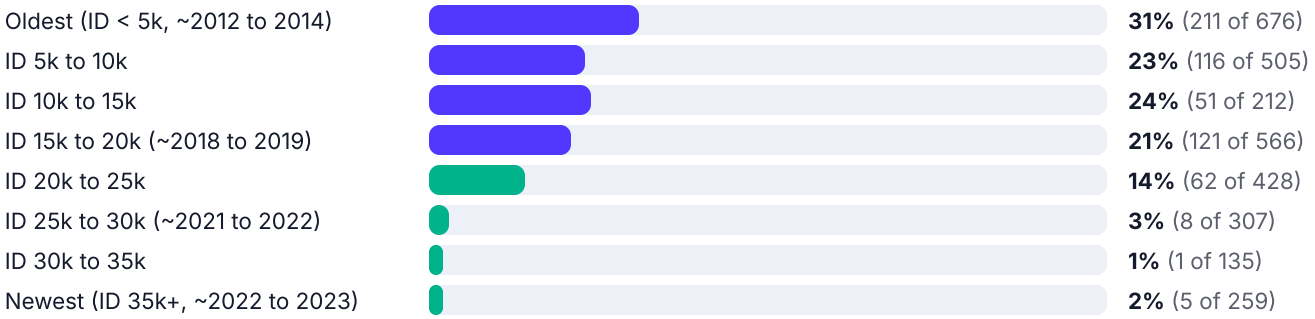
Two quick wins

Inspect author-page links, visible biographies and accurate attribution. Use Search Console to check index status where available. Keep archive links useful and crawlable when they provide access to articles; do not remove or noindex a page on the basis of this panel message alone.

Finding 3: Older URLs returned the message more often

Grouping news URLs by WordPress post ID suggests an age-related pattern in the displayed panel message. IDs are approximate, and the crawl does not establish indexing, ranking changes or the cause of the pattern.

News articles Google said it hadn't seen, by post ID (older at top)



Reading the chart

Up to 31% of the oldest ID group returned the message, versus under 3% in newer groups. Years are approximate, from dates that appear in some slugs. IDs aren't perfectly chronological. 38 news URLs had no ID and were left out of this chart.

Possible explanations, not findings

Age may correlate with changed links, relevance or maintenance, but these causes were not tested. Post IDs are only approximate age indicators. The observed relationship is with the panel message, not verified deindexing or loss of rankings.

Connection

Use the age pattern to choose pages for a closer review. Check accuracy, reader value, links and first-party search evidence before deciding to keep, refresh, merge or remove content. The panel alone cannot justify a purge.

What it means

Lessons for large content sites

1

Check what Google knows

The panel supplies contextual messages. URL Inspection and the indexing report address index status. Compare the tools without treating them as equivalent.

2

Review identity facts

Keep ownership, purpose, About content and identity links accurate. A missing panel description does not establish a ranking disadvantage.

3

Protect author pages

Link bylines to useful author profiles and show relevant evidence. A panel message does not prove that authorship evidence is invisible.

4

Audit old content by age

Group pages by age or ID to select a review sample. Assess usefulness and accuracy separately from the panel-message pattern.

5

Review archives by purpose

Keep useful discovery paths. Review indexing needs separately, using page purpose and first-party evidence.

6

Re-run it

A single crawl is a snapshot. Repeat it after a cleanup to **measure the change**.

Limits of this study	Why it matters
One site, one moment (late 2023)	Results describe this site, not all sites
"Hasn't seen" isn't an official index status	Confirm with Search Console before acting
Google has since changed its search features	The panel's wording and behaviour may differ today
Post IDs approximate age	Some IDs don't follow publishing order

Analysis by Ameneh Saeednia, from an archived public-panel crawl, with AI assistance. No private analytics or business data are used. Interpretation corrected for public release: panel context is not an entity or index-status test.

Public-release interpretation references
Google Search Help: Learn more about a web page
https://support.google.com/websearch/answer/11127743
Google Search Console: URL Inspection API
https://developers.google.com/webmaster-tools/v1/urlInspection.index/inspect