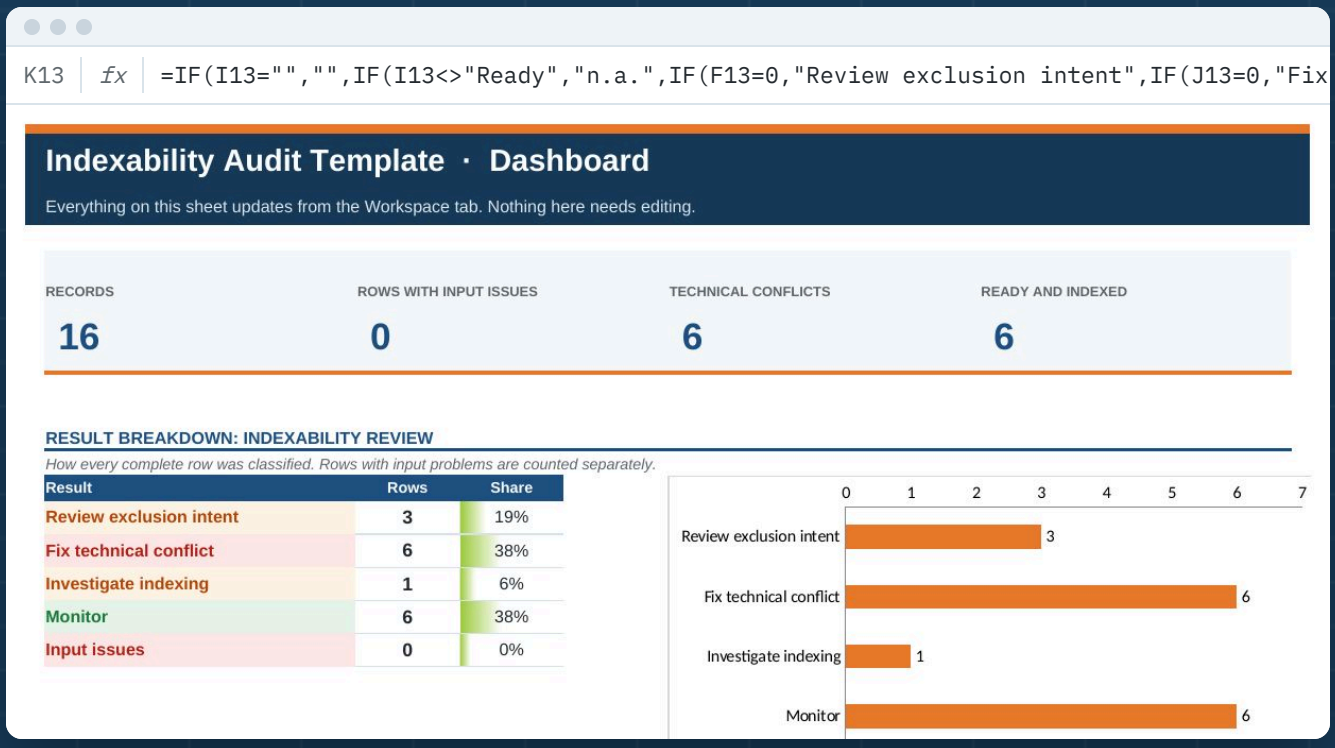
OptimizationTheory.com

SEO Migration Toolkit, resource 11

Indexability Audit Template

Check status, robots, noindex and canonical on every page you want indexed, then see exactly what to fix first.

Built for the DataForSEO API: On-Page API



Field guide, v3.0, November 2026

Jourdan Rombough, optimizationtheory.com

What this template decides

Which pages that should be indexed have a technical conflict, which are technically fine but still not indexed, and which were never meant to be indexed.

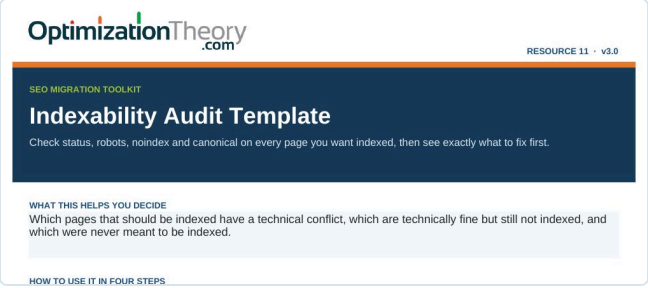
'Crawled, currently not indexed' and 'Duplicate, Google chose a different canonical' are some of the most common Search Console complaints, and some of the hardest to triage at scale. The cause is often mundane: a non-200 status, a robots block, a stray noindex or a canonical pointing somewhere else.

This indexability audit template checks each URL for those four conditions and compares the result with your intent and Search Console's indexing status. It sorts every row into one of four actions: Fix technical conflict, Investigate indexing, Review exclusion intent, or Monitor.

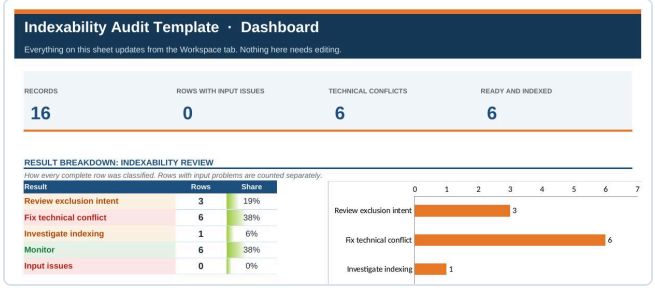
Who it's for

- SEOs triaging Search Console indexing reports
- Developers verifying templates after a release
- Migration teams checking the new site before and after launch
- Ecommerce teams with faceted or paginated URLs

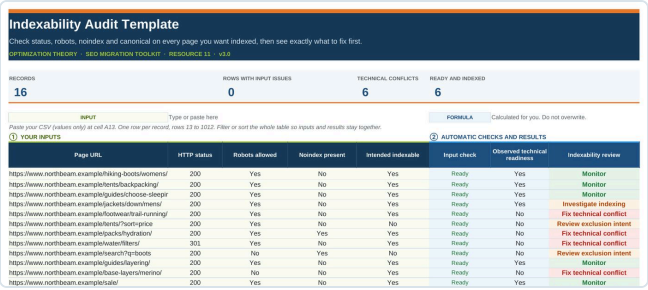
What's in the workbook



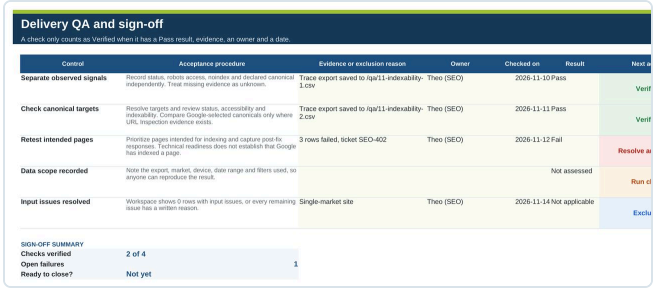
Start Here. Purpose, four-step setup and colour key.



Dashboard. KPIs, charts, Top 10 and an automatic action list.



Workspace. The working table, room for 1,000 rows.



Delivery QA. Evidence, owner and date for each validation step.

How the calculation works

Technical readiness = 1 when status is 200, robots allows crawling, noindex is 0 and the canonical equals the URL. The review returns Review exclusion intent when you don't intend to index the page, Fix technical conflict when it isn't technically ready, Investigate indexing when it's ready but not indexed, otherwise Monitor.

The key formula, as written in the workbook

```
=IF(I13="","",IF(I13<>"Ready","n.a.",IF(F13=0,"Review exclusion intent",IF(J13=0,"Fix technical conflict",IF(G13=0,"Investigate indexing","Monitor")))))
```

Result column	Plain-English formula	What it tells you
Observed technical readiness	IF(AND(HTTP status=200,Robots allowed=1,Noindex present=0,Declared canonical=Page URL),1,0)	Observed self-canonical 200 without known crawling/indexing block.
Indexability review	IF(Intended indexable=0,"Review exclusion intent",IF(Observed technical readiness=0,"Fix technical conflict",IF(GSC indexed=0,"Investigate indexing","Monitor")))	Keeps actual GSC indexing separate from crawl readiness.

The results you'll see

Review exclusion intent | Fix technical conflict | Investigate indexing | Monitor

Green means the row meets the checks, red means fix it first, amber needs a human review and blue marks an opportunity. Rows with missing or out-of-range inputs show in the Input check column and are left out of the results until fixed.

How to use it, step by step

1

List URLs and your intent

Include pages you want indexed and the ones you don't. Mark intended indexable as 1 or 0.

2

Crawl for technical signals

Record status code, whether robots.txt allows crawling, whether the page has noindex, and its canonical URL.

3

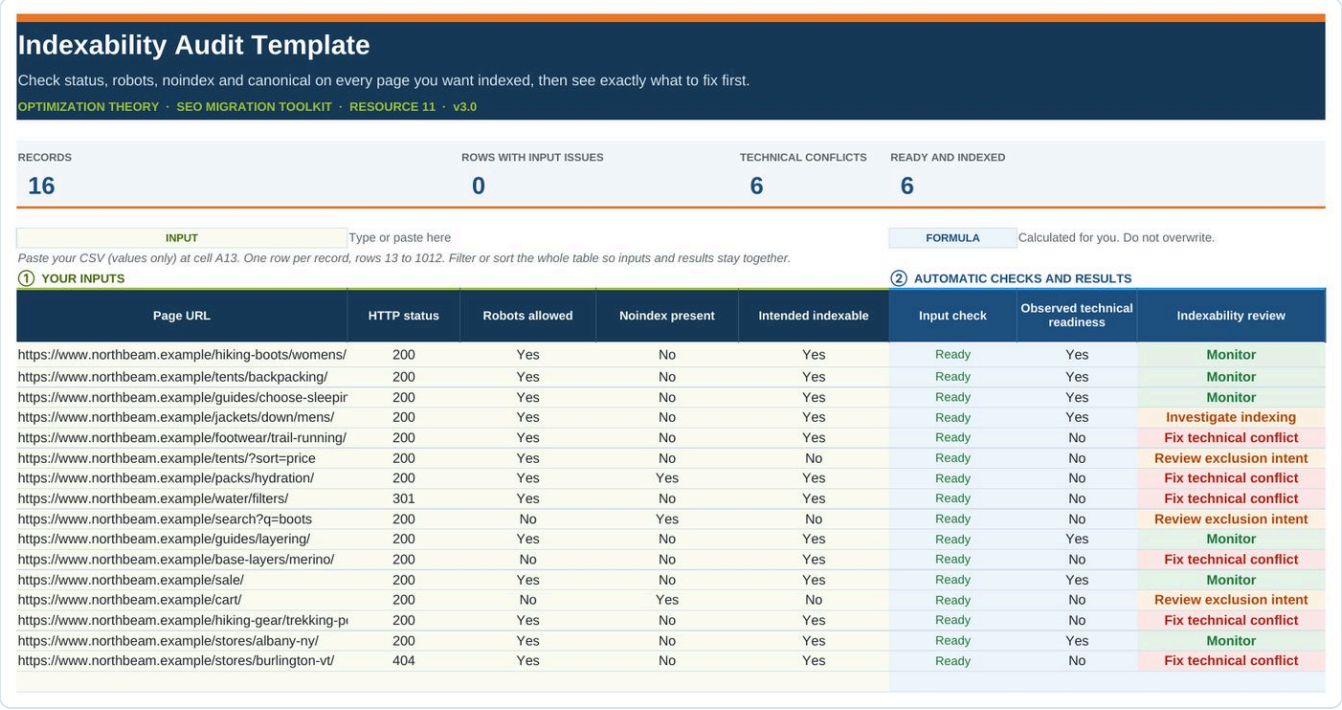
Add indexing status

Use URL Inspection or the page indexing report to record whether Google has indexed the URL (1/0).

4

Work the queue in order

Fix technical conflicts first, then investigate indexing on technically ready pages, then confirm exclusions.



Indexability Audit Template

Check status, robots, noindex and canonical on every page you want indexed, then see exactly what to fix first.

OPTIMIZATION THEORY · SEO MIGRATION TOOLKIT · RESOURCE 11 · v3.0

RECORDS	ROWS WITH INPUT ISSUES	TECHNICAL CONFLICTS	READY AND INDEXED
16	0	6	6

INPUT Type or paste here — FORMULA Calculated for you. Do not overwrite.

Paste your CSV (values only) at cell A13. One row per record, rows 13 to 1012. Filter or sort the whole table so inputs and results stay together.

① YOUR INPUTS — ② AUTOMATIC CHECKS AND RESULTS

Page URL	HTTP status	Robots allowed	Noindex present	Intended indexable	Input check	Observed technical readiness	Indexability review
https://www.northbeam.example/hiking-boots/womens/	200	Yes	No	Yes	Ready	Yes	Monitor
https://www.northbeam.example/tents/backpacking/	200	Yes	No	Yes	Ready	Yes	Monitor
https://www.northbeam.example/guides/choose-sleepir	200	Yes	No	Yes	Ready	Yes	Monitor
https://www.northbeam.example/jackets/down/mens/	200	Yes	No	Yes	Ready	Yes	Investigate indexing
https://www.northbeam.example/footwear/trail-running/	200	Yes	No	Yes	Ready	No	Fix technical conflict
https://www.northbeam.example/tents/?sort=price	200	Yes	No	No	Ready	No	Review exclusion intent
https://www.northbeam.example/packs/hydration/	200	Yes	Yes	Yes	Ready	No	Fix technical conflict
https://www.northbeam.example/water/filters/	301	Yes	No	Yes	Ready	No	Fix technical conflict
https://www.northbeam.example/search?q=boots	200	No	Yes	No	Ready	No	Review exclusion intent
https://www.northbeam.example/guides/layering/	200	Yes	No	Yes	Ready	Yes	Monitor
https://www.northbeam.example/base-layers/merino/	200	No	No	Yes	Ready	No	Fix technical conflict
https://www.northbeam.example/sale/	200	Yes	No	Yes	Ready	Yes	Monitor
https://www.northbeam.example/cart/	200	No	Yes	No	Ready	No	Review exclusion intent
https://www.northbeam.example/hiking-gear/trekking-p	200	Yes	No	Yes	Ready	No	Fix technical conflict
https://www.northbeam.example/stores/albany-ny/	200	Yes	No	Yes	Ready	Yes	Monitor
https://www.northbeam.example/stores/burlington-vt/	404	Yes	No	Yes	Ready	No	Fix technical conflict

The Workspace tab with the worked example. Lime cells are inputs; blue cells are formulas.

Fill it with DataForSEO

This template is built to run on **DataForSEO On-Page API**: a crawl of your site: URLs, status codes, canonicals, inlinks and sitemap membership.

Most of the data columns fill automatically. Add the business inputs listed below, then paste.

1. Request the data

```
POST https://api.dataforseo.com/v3/on_page/task_post
[
  {
    "target": "yourdomain.com",
    "max_crawl_pages": 1000
  }
]
```

Post the crawl task, wait for it to finish, then request on_page/pages with {"id": "<task id>", "limit": 1000} and save that response.

2. Save the full JSON response

Save the complete response, not just the task ID, next to the template folder.

3. Convert it to the input CSV

The script in the Automation folder runs offline: no API call, no credentials.

```
python3 Automation/dataforseo_import.py import --template 11 --input response.json --output my-inputs.csv
```

Fills automatically	You add
Page URL, HTTP status, Declared canonical	Robots allowed, Noindex present, Intended indexable, GSC indexed

Source reference fills automatically with the task ID, endpoint, location, language, device and snapshot time, so every row is traceable. Paste the finished CSV at Workspace cell A13. Estimate the cost first with the DataForSEO Cost Calculator (template 38).

Input dictionary

Keep the CSV columns in this order and paste values at Workspace cell A13. A missing observation stays blank; use 0 only for a measured zero or a verified No.

CSV field	Column	Meaning and rule	Typical source
`url`	Page URL	Exact absolute URL. Keep case and query parameters unless a documented normalization rule applies.	DataForSEO: url
`status_code`	HTTP status	Observed crawl status.	DataForSEO: status_code
`robots_allowed`	Robots allowed	1 only after robots.txt permits crawling for the intended bot. 1 = Yes, 0 = No.	Your records or manual review
`noindex`	Noindex present	1 if a noindex directive is present in rendered meta or HTTP header. 1 = Yes, 0 = No.	Your records or manual review
`canonical_url`	Declared canonical	Exact resolved declared canonical URL. Enter MISSING if absent.	DataForSEO: canonical_url
`intended_indexable`	Intended indexable	1 when the approved strategy intends this URL to be indexed. 1 = Yes, 0 = No.	Your records or manual review
`gsc_indexed`	GSC indexed	1 only when inspected in GSC and reported indexed. Enter 0 for checked and not indexed. 1 = Yes, 0 = No.	Your records or manual review
`source_ref`	Source reference	Task ID, export filename and observation scope. Retain the untouched source response separately. Optional.	DataForSEO: source_ref

The worked example

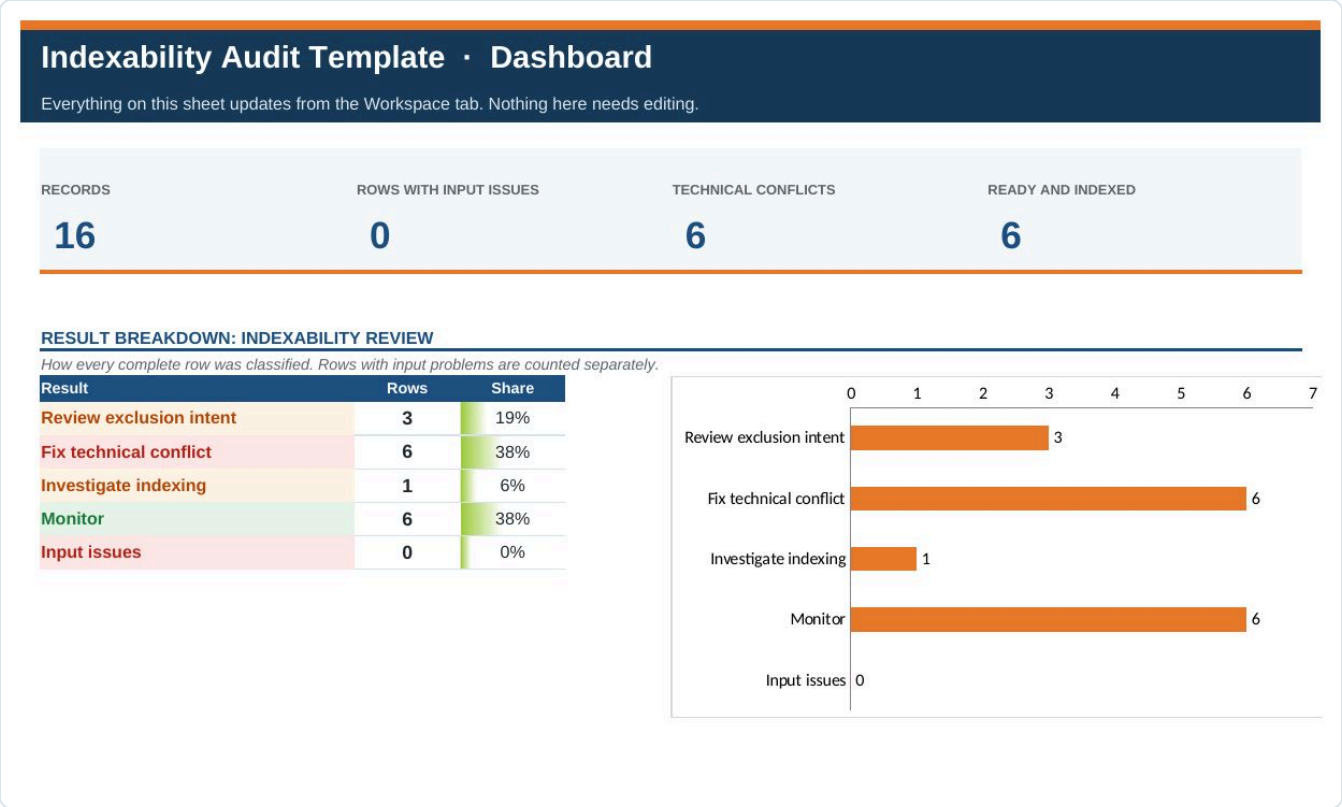
Every workbook ships filled with data from Northbeam Outdoor, a fictional outdoor-gear retailer. This one has 16 example rows.

16	0	6	6
Records	Rows with input issues	Technical conflicts	Ready and indexed

Six example URLs have technical conflicts: a canonical pointing at ?page=1, a stray noindex, a 301, a robots-blocked category, a canonical to a parent folder and a closed store that 404s. One more page is technically clean but not indexed, which is a quality or demand question, not a tag problem.

Result breakdown

Result	Rows	Share
Review exclusion intent	3	19%
Fix technical conflict	6	38%
Investigate indexing	1	6%
Monitor	6	38%
Input issues	0	0%



The Dashboard updates automatically from the Workspace tab.

Reading the result honestly

A ready crawl response does not prove Google selected this canonical or indexed the page.

Limits. Robots blocking can prevent a crawler from seeing noindex. Blank inspection data stays missing, not an assumed zero.

Mistakes to avoid

Fixing 'not indexed' with a request indexing click. If the page has a technical conflict, requesting indexing won't help. Fix the signal first.

Canonicalizing to URLs with parameters. Canonicals should point at the clean, indexable URL. Parameter canonicals usually signal a template bug.

Blocking pages you want deindexed. If robots.txt blocks a page, Google can't see its noindex. Allow crawling until it drops out, then block if needed.

Tips from the field

Run this on a template sample after every release.

Combine with the sitemap audit: URLs in the sitemap should all be technically ready.

Investigate indexing rows often need better content, internal links or consolidation, not more tags.

Questions people ask

What is indexability in SEO?

Whether a page can be crawled and indexed: it returns 200, isn't blocked by robots.txt, has no noindex directive and declares itself (or no other page) as canonical.

Why is my page crawled but not indexed?

Usually because Google doesn't find it valuable or distinct enough, or because it's a near-duplicate of another URL. Technical conflicts should be ruled out first, which this template does.

Does a canonical tag guarantee which URL is indexed?

No. It's a strong hint. Google may choose another canonical if signals conflict.

Should robots.txt be used to remove pages from Google?

No. Use noindex and allow crawling so Google can see it. Robots.txt controls crawling, not indexing.

Templates that work well with this one

XML Sitemap Audit Template

Reconcile sitemap URLs against status, canonical, robots and noindex to see what to add, remove or fix.

Hreflang Audit Template

Validate every hreflang relationship: alternate status, self-reference, return tags, canonical match and locale codes.

Redirect Mapping Template

Map every old URL to its new home and verify each redirect: one permanent hop, a 200 page, no noindex, matching canonical.

Crawl Budget Analysis Template

Group verified Googlebot requests by URL pattern, measure crawl waste and pick the right fix for each pattern.

Schema Markup Audit Template

Check presence, syntax, required fields, visible-content match and policy for every structured data type you ship.

Technical SEO Prioritization Template

Turn a technical SEO audit into a ranked fix queue by recovered clicks, profit and profit per hour of dev time.

Read the full tutorial online: **optimizationtheory.com/download-view/indexability-audit-template/**
Get all six templates in the **SEO Migration Toolkit**, or all 40 in the **Complete SEO & GEO Toolkit**.