MarketingSignals

Free Build Standard

Version 2.2 · Updated 9 November 2026

SEO And GEO Website Build Standard 2027

A platform-agnostic standard for building websites that search engines and AI assistants can crawl, understand and cite, while staying fast, accessible, accurate and built to convert.



Peer reviewed by **Jono Alderson** and **Mark Williams-Cook**

About this standard

Build websites that are easy for search engines and AI-assisted discovery systems to access, understand and reference, while remaining fast, accessible, accurate and effective at generating enquiries or sales. This is a platform-agnostic implementation standard, not a guarantee of rankings or citations.

Version	2.2, updated 9 November 2026
Audience	Developers, technical SEOs, content teams, designers and project owners
Scope	New websites, redesigns, CMS migrations and major template changes
Online	marketingsignals.com/seo-geo-website-build-standard
Checklist	Free checklist Google Sheet (File > Make a copy to use it)

Peer review

This standard has been peer reviewed by Jono Alderson and Mark Williams-Cook. Their review comments are incorporated in version 2.2, covering content negotiation for Markdown versions of pages, image loading priorities, crawlable pagination, share image dimensions and terminology.

Contents

MANDATORY BUILD REQUIREMENTS

1 Environments, hosts and deployment safeguards
2 URL design, migrations, redirects and errors
3 Crawlability, indexability and XML sitemaps
4 Rendering, HTML and technical architecture
5 Page-level metadata and social previews
6 Structured data and entity consistency
7 Information architecture and internal linking
8 Content, GEO and evidence standards
9 Images, video, performance and accessibility
10 CMS and editorial governance

CONDITIONAL REQUIREMENTS

11 Activate the right modules for the site

EXPERIMENTAL GEO ENHANCEMENTS

12 Machine-readable publishing experiments
13 Search and AI crawler policy
14 Measure GEO outcomes rather than compliance with formatting myths

DELIVERY, LAUNCH AND ONGOING VERIFICATION

15 Required acceptance gates
16 Definition-of-done scorecard

Reference documents and maintenance

How to use this standard

Every project must be assessed against three levels:

Level	Meaning	Implementation rule
P0: Mandatory	Essential technical SEO, quality, accessibility and launch safeguards	Must pass before launch; exceptions require written approval
P1: Conditional	Features required when the business model, target market or website architecture calls for them	Decide applicability at discovery; deliver applicable items before launch or document an approved phase
P2: Experimental	Emerging GEO formats and integrations with unproven or platform-specific search benefit	Implement only with a stated use case, owner, cost and measurement plan

Ownership: The developer owns technical implementation and automated tests; the SEO lead owns search intent, indexing rules and redirects; the content owner owns factual accuracy and editorial quality; the project owner signs off the release. Assign named individuals at project kick-off.

Small builds: where one person, or one person working with an AI coding assistant, does the build, these roles can collapse to a single named owner who signs off each gate in Section 15. Keep the gates and the evidence; drop the hand-offs.

Definition of done: A requirement is not complete because it exists in the CMS or works on one example page. It must work on every applicable template, survive a deployment, and have a documented test and result.

Tier 1 · P0 Mandatory

Mandatory build requirements

1 Environments, hosts and deployment safeguards

- **Production host:** Nominate one canonical HTTPS hostname (`www` or non-`www`). Redirect alternative production host/protocol variants permanently to it. Enforce HTTPS and keep certificates valid.
- **Staging and preview:** Protect staging, preview and platform-generated URLs with authentication or equivalent access controls. Do not rely on robots.txt to keep confidential or preview environments out of indexes. Make sure previews are not publicly linked or included in sitemaps.
- **No accidental alternative sites:** Platform hosts and obsolete environments must not serve an unrestricted duplicate production site. Where practical and safe, redirect public alternate production hosts; do not redirect private staging previews to unrelated live pages.
- **Release controls:** Production indexability must be explicitly checked at launch. A production `noindex`, blocked crawler, incorrect canonical host or test-domain reference is a release blocker.
- **Development and production parity:** Templates, SEO fields, redirects and canonical rules must be tested in a production-like environment. Feature flags must not silently override per-page SEO requirements.
- **Post-deploy smoke test:** On every deployment, fetch representative live templates and check status, robots directives, canonical, title, visible primary content and navigation.

> **Acceptance test:** No public indexable staging URLs; the production homepage and representative content pages return the expected `200` HTML response; all required metadata points to the production host.

2 URL design, migrations, redirects and errors

- **Stable URL conventions:** Define routes for each content type before development, including lower/upper case, hyphens, trailing slash, language prefix and category path. Avoid unnecessary URL changes.
- **Migration inventory:** For rebuilds, combine old XML sitemaps, an independent old-site crawl, Search Console landing pages, analytics landing pages, backlink targets and existing redirect rules into one source URL inventory.
- **URL mapping:** Map each existing valuable URL to the same URL wherever possible, or to the closest genuinely relevant replacement. Obtain SEO approval for content consolidation and changed slugs.
- **Redirect system:** Maintain editor-manageable redirects as structured records, with source, target, status and notes. Exact rules take precedence over configured patterns. Validate wildcard rules to avoid accidental broad redirects.
- **One-hop redirects:** Aim for a single direct `301` or `308` to the final destination; detect loops and chains. Test HTTP, HTTPS, www and slash variations. Use temporary redirects only when the change is genuinely temporary.
- **Removed pages:** Return a real `404` or `410` when no relevant replacement exists. Do not redirect every missing URL to the homepage. Design a helpful 404 page that still returns a 404 status.

- **Route correctness:** Pages that do not exist must not return `200` with an empty template or an error message (soft 404). Forms and route handlers must not expose arbitrary indexable URLs.
- **Internal links:** Update all internal references to canonical destination URLs rather than relying on redirects.
- **Sitemaps and analytics:** Update sitemap paths, navigation, breadcrumbs, canonicals, hreflang links and campaign destinations together with any URL migration.

> **Acceptance test:** Every legacy priority URL has a documented outcome (`200` , one-hop relevant redirect, or intentional `404/410`); no redirect loops or critical soft 404s; no broken high-value internal links.

---

3 Crawlability, indexability and XML sitemaps

- **robots.txt:** Provide a valid `robots.txt` on the canonical host. Block only paths that should not be crawled and for which blocking is appropriate. Do not block critical CSS, JavaScript or rendered resources.
- **Crawler policy:** Establish separate access decisions for traditional search crawlers, AI search crawlers, interactive assistant fetchers and AI training crawlers. See Section 13. Do not assume all AI agents have the same purpose.
- **Indexing directives:** Make `index` / `noindex` an explicit per-page field with secure defaults by content type. Use a robots meta tag or `X-Robots-Tag` for non-HTML resources where needed. A crawler must be allowed to fetch a URL to observe its `noindex` directive.
- **Canonical URLs:** Every intended indexable HTML page has one absolute, valid preferred canonical, normally self-referencing. Canonical targets should return `200` , be indexable and represent duplicate or substantially equivalent content. Cross-domain canonical overrides require review.
- **Duplicate URL rules:** Document treatment of query parameters, campaign tracking, filtering, faceting, sorting, internal search, pagination, print variants and syndication. Choose between canonicalisation, crawling restrictions and `noindex` based on what the URL actually does; do not use them interchangeably.
- **XML sitemaps:** Generate sitemap indexes and sitemaps automatically where needed. Include only canonical, indexable URLs that return `200` . Exclude blocked, redirected, deleted and noindexed pages. `lastmod` must change when meaningful page content changes, not on every build.
- **Discovery:** Include the sitemap URL in robots.txt, submit the relevant properties/sitemaps in Google Search Console and Bing Webmaster Tools, and keep ownership access documented.
- **Internal discovery:** Important pages must be linked from other accessible, relevant HTML pages; an XML sitemap alone is not a replacement for navigation.
- **IndexNow:** Where applicable, automate notifications for newly published, meaningfully updated and deleted URLs to participating search engines. Store the verification key correctly; retry failed submissions sensibly. Submission is notification, not an indexing guarantee.

> **Acceptance test:** Crawl the deployed site using a search-engine user agent and a normal browser. Indexable pages are reachable, self-canonicalised, included in sitemaps and free of accidental blocks. No sitemap URLs redirect, error or carry `noindex` .

4 Rendering, HTML and technical architecture

- **Crawlable initial HTML:** Serve essential body copy, headings, navigation, internal links and metadata in the initial response using SSR, SSG or another reliable server-rendered approach. Client-side enhancement is fine; critical information must not require clicks, scrolling or application state to exist.
- **Real links:** Use `<a href="/path/">` for navigation. Do not depend on JavaScript `onClick`, `javascript:` URLs or fragment-only pseudo-links for discoverability.
- **Semantic structure:** Use meaningful page landmarks (`header`, `nav`, `main`, `footer`), lists, tables, buttons and form labels. Keep core content in a sensible reading order.
- **Page headings:** Use one clear primary H1 as a house style, with descriptive H2/H3 headings in logical order. Titles and H1s should be aligned in meaning, not mechanically identical.
- **Unique page identity:** Each indexable URL must expose its own primary content, page title, canonical and relevant structured data. Watch for client-side routers inheriting metadata from the previous page.
- **Pagination:** Provide real paginated URLs and navigable links when lists span multiple pages. Avoid infinite-scroll-only discovery. Do not canonicalise every paginated page to page one unless they are genuine duplicates.
- **HTTP semantics:** Content pages return `200`; redirects return a true redirect response; deleted or missing content returns `404/410`; genuine server failures return `5xx`.
- **Caching:** Cache static assets safely, invalidate changed HTML and sitemaps promptly, and use versioned asset filenames to prevent stale frontend bundles.
- **Testing:** Inspect both the raw response HTML and browser-rendered DOM for every template. Confirm content equivalence for users and crawlers.
- **Alternative formats:** Any alternative representation of a page (Markdown, plain text, print or AMP variants) must not compete with the HTML page. Prefer content negotiation on the same URL (see Section 12). Where a separate URL is used, serve it with an HTTP `Link: <html-url>; rel="canonical"` header or `X-Robots-Tag: noindex`, and keep it out of XML sitemaps.

> **Acceptance test:** Disable JavaScript and fetch source HTML on all important templates: their principal text, anchor links, title, canonical and basic context remain accessible.

5 Page-level metadata and social previews

Provide editable or predictably generated values for each applicable page:

Element	Standard
Title	Unique, descriptive, naturally worded, aligned with search intent and appropriate branding. Roughly 50 to 60 characters can be an editorial preview target, not a hard search-engine limit.
Meta description	Unique, useful summary with a clear reason to visit. Roughly 140 to 160 characters can be a drafting guide, not a mandatory threshold.
Canonical	Absolute production URL, using the rules in Section 3; optional override with validation and approval.

Element	Standard
Robots directive	Correct for the page type; never inherited accidentally from a staging environment.
H1	Visible, specific primary heading; does not need to reproduce the title word for word.
Social metadata	Open Graph title, description, image and URL; Twitter/X summary-large-image where appropriate. Use a sensible default and page-level overrides.
Language	Correct HTML `lang` for the content and accurate locale fields where relevant.
Date and authorship	Accurate published/updated dates and named authors on content where those facts genuinely matter.

- Prevent accidental duplicate titles, missing descriptions, placeholder metadata and production pages referencing staging asset URLs.
- A 1200 × 675 (16:9) share image is a sensible universal default: it suits Google Discover and the main social platforms. It is a design asset, not a search ranking requirement. Social image alternative text can be supplied through supported platform tags where useful.
- Allow large image previews with `max-image-preview:large` in the robots meta tag, so pages are eligible for large images in Google Discover.
- Do not generate hundreds of near-identical titles just to insert keyword variants.

> **Acceptance test:** Template sampling finds the correct metadata in raw HTML. Automated reports flag missing fields and unintended duplication; exceptions are reviewed rather than blindly padded to a character count.

---

6 Structured data and entity consistency

- **JSON-LD implementation:** Use supported, valid schema.org types that accurately describe the visible page. A connected `@graph` and stable `@id` values are preferred when they simplify reuse, but one graph is not compulsory.
- **Single source of approved facts:** Store brand name, legal identity, site URL, logos, contact details, service areas, leadership and verified public profiles in one governed data source used by templates and schema.
- **No invented attributes:** Do not add ratings, awards, pricing, credentials, author details or entity relationships that are not substantiated. Never add markup solely because a validator accepts it.
- **Visible-content alignment:** Structured data claims must correspond to genuine page content and the current offer. Remove or update markup when the underlying information changes.
- **Validate correctly:** Run Google Rich Results Test for Google-supported rich-result types and Schema Markup Validator for general schema correctness. Passing validation does not guarantee an enhanced search result.

Core and common conditional schema modules:

Content	Relevant types and properties
Site identity	`Organization` or a suitable subtype; `WebSite`; stable organisation `@id`, official logo, verified profiles
Site hierarchy	`BreadcrumbList` on appropriate non-homepage templates, matching the visible breadcrumb trail
Editorial	`Article` / `BlogPosting`, accurate `author`, `publisher`, `datePublished`, `dateModified`, image
Services	`Service`, `provider`, `areaServed`, relevant offers where truthful and useful
People	`Person` on genuine author/team pages with verified biographies and profiles
Video	`VideoObject` with actual upload date, thumbnail, location and duration where known
Employment	`JobPosting` on active jobs, following current search-engine requirements and removing expired postings
Ecommerce	`Product`, `Offer`, `AggregateRating` where permitted and accurate; variants and availability as appropriate
Location or events	`LocalBusiness`, `Event` and relevant business-specific types when applicable
Glossaries / ordered lists	`DefinedTerm`, `DefinedTermSet`, `ItemList` where they truthfully describe page content

FAQ note: Useful FAQs are an editorial choice; FAQ schema must not be treated as an assured Google rich result or AI citation technique. Google's FAQ rich results were discontinued in 2026.

> **Acceptance test:** Representative URLs from every schema-bearing template validate without errors; all substantive claims are visibly supported; references resolve to the correct brand/person/page entities.

---

7 Information architecture and internal linking

- **Design for user intent:** Plan a hierarchy that distinguishes navigational, informational, comparison and transactional needs. Do not create a page solely because a keyword variant exists.
- **Navigation, breadcrumbs and pagination:** Main navigation, contextual links, breadcrumbs, pagination and HTML lists expose the most important pages without JavaScript-only interaction. Pagination is a common failure: every page in a series needs a real `<a href>` link, not only a JavaScript "load more" button or infinite scroll.
- **Topic relationships:** Service pages link to applicable sub-services, industries, supporting guides and case studies. Editorial content links to related content and relevant commercial destinations where genuinely helpful.
- **Internal links as content data:** Prefer CMS relationships for links to internal pages so slug changes update destinations automatically. Provide link checks for any manual URLs.
- **Anchor text:** Descriptive and natural; avoid repeated exact-match anchors where they read unnaturally. Accessible labels must make sense outside their immediate sentence.

- **External citations:** Link normally to useful evidence and authoritative sources. Use `rel="sponsored"` for paid placements, `rel="ugc"` for user-generated links, and `rel="nofollow"` where a link should not be endorsed. Do not automatically nofollow all sources or comparison-list references.
- **New tabs:** Open links in the same tab by default unless there is a clear UX reason not to. When using `target="_blank"`, use safe `rel` values and communicate behaviour when helpful.
- **Orphans and depth:** Automatically report indexable pages without relevant incoming internal links, including those only discoverable in sitemaps.

> **Acceptance test:** No important orphan pages, unintended broken links or misleading anchors. Key commercial and editorial journeys can be navigated using ordinary HTML links.

---

8 Content, GEO and evidence standards

Content is built first to answer real customer questions and demonstrate why the business is credible. The same qualities support search engines and AI-assisted discovery; no particular formatting guarantees citation.

- **Clear answer or proposition:** Explain what the page is about near the beginning. Commercial pages should establish the offer, audience, benefit and proof without making visitors wade through generic content.
- **Original value:** Include first-hand expertise, examples, credible comparisons, unique data, transparent methodology or direct experience that cannot be reproduced from generic summaries alone.
- **Entity accuracy:** Name products, people, brands, standards, platforms and places precisely where relevant. Use specific examples rather than vague claims; do **not** impose arbitrary entity-density quotas.
- **Claims and evidence:** Maintain a governed claim register containing the exact approved claim, source/evidence, methodology, scope, approval owner, date checked and permitted reuse. Use consistent language when repeating important results.
- **Structure for reading:** Use clear headings, concise paragraphs, bullets, comparison tables, FAQs and short takeaways when they genuinely make the page more useful. No fixed rule for FAQ count, paragraph length, first-answer position or percentage of question headings.
- **FAQs:** Answer meaningful questions in natural language, with self-contained explanations and links where needed. Avoid filler FAQs added only for keywords.
- **Freshness:** Show a visible reviewed/updated date where useful. Only change `dateModified` for meaningful updates; schedule reviews for time-sensitive pricing, product claims, legislation and platform guidance.
- **Authorship and trust:** Make relevant author credentials, company identity, contact routes, editorial policies and relevant regulatory disclosures available and accurate.
- **Source quality:** Prefer primary documentation and original data for factual assertions. A sources section is desirable for research-led and reference content; outbound citations should remain normal followed links unless a specific qualification applies.
- **Editorial quality:** Use UK English where appropriate to the brand; preserve proper names and product terminology. Check spelling, grammar, accessibility, accuracy, duplicated sections and unsupported figures before publication.
- **AI-assisted production:** Human review, factual verification and original contribution are mandatory for material produced with AI. Avoid large volumes of near-duplicate or unoriginal pages created mainly to

target search traffic.

Recommended default patterns

These are starting points that have worked well on service and editorial sites, not quotas. Use them unless the page has a better reason to do something else.

- **Summary block near the top:** On commercial pages, a short panel directly below the hero with three or four complete, standalone claims (each one sentence that makes sense quoted on its own), led by the strongest approved proof and followed by a clear call to action. Build it as a removable block with a site-wide off switch, so it can be tested and withdrawn without editing every page.
- **Sell in the top half, answer in the bottom half:** Keep benefit-led headings, proof and the offer above the fold and in the first sections. Put informational content (how it works, accordions, FAQs) lower down, where question-style headings suit readers looking for detail.
- **Answer echo:** Where a heading is a question, open the section with a sentence that restates its subject and answers it directly ("Law firm PPC campaigns are structured by case type, because...").
- **Specific names over vague references:** "Ahrefs and Semrush" rather than "SEO tools"; "the FCA" rather than "regulators". Use this as an editing habit and an audit signal, not a percentage target.
- **Self-contained FAQ answers:** Each answer stands alone and links to the deeper page where it helps.
- **Consistent proof:** When a case-study result is repeated (summary blocks, FAQs, schema), reuse the exact approved wording from the claim register.

> **Acceptance test:** Each priority template has a defined user goal, clear main answer/offer, relevant real-world proof and verified material claims. Content owners approve evidence before launch.

---

9 Images, video, performance and accessibility

Images and video

- Use descriptive filenames where practical, responsive sources (`srcset` / `sizes`), correct image dimensions and modern formats where compatible.
- Write meaningful alternative text for informative images. Use empty `alt=""` for purely decorative images. Do not repeat adjacent captions verbatim or force keyword stuffing into alt text.
- Use original imagery and charts where they add evidence. Reusing a logo or illustration is acceptable when it is the right design choice; a blanket rule against repeat images is unnecessary.
- Never lazy-load the primary above-the-fold (LCP) image. Give it `fetchpriority="high"` and preload it (`<link rel="preload" as="image">` , with `imagesrcset` / `imagesizes` for responsive images).
- Use `fetchpriority="low"` and `loading="lazy"` on below-the-fold images, and `loading="lazy"` on below-the-fold `<iframe>` s (embedded video, maps, forms).
- For below-the-fold `<video>` , use `loading="lazy"` where browsers support it, with `preload="none"` and a poster image as the fallback.
- Give important videos an accessible player, thumbnail, useful summary or transcript and captions as needed. Add `VideoObject` and video sitemaps when justified by the video programme.

Performance

Meet the following **good Core Web Vitals** thresholds at the 75th percentile of real user visits, on mobile and desktop where data are available:

Metric	Target
Largest Contentful Paint (LCP)	**≤ 2.5 seconds**
Interaction to Next Paint (INP)	**≤ 200 milliseconds**
Cumulative Layout Shift (CLS)	**≤ 0.1**

- Before launch, set realistic template-level lab budgets for JavaScript, CSS, image transfer, third-party tags and server response times; test representative mobile devices and connections.
- Prefer efficient image delivery, caching, limited third-party scripts, self-hosted fonts where practical, and reserved dimensions for media and dynamic blocks.
- Treat Lighthouse as a diagnostic tool. Its lab score is not a substitute for field Core Web Vitals data.

Accessibility and agent usability

- **Target WCAG 2.2 AA** for applicable interfaces. Provide semantic controls, visible focus states, adequate contrast, keyboard access, correct labels and error handling.
- Menus, accordions, tabs, search, forms and checkout must work without a mouse. Use standard HTML controls with appropriate accessible names, roles and states.
- Respect `prefers-reduced-motion` and make essential actions understandable without animation, hover or colour alone.
- Make important service, price, eligibility and contact information available in normal page text, not only as image text or inaccessible widgets.

> **Acceptance test:** No critical accessibility failures in automated checks or keyboard/manual testing; all essential journeys work; template performance is measured and a field-monitoring plan is active.

---

10 CMS and editorial governance

The CMS must make correct implementation easy and incorrect implementation difficult.

Required capabilities

- Reusable templates, structured content types and editable titles, descriptions, slugs, headings, social images and appropriate SEO overrides.
- Controlled index/noindex defaults per content type; changing a page to `noindex` must automatically remove it from XML sitemaps.
- Automatic canonical generation with tightly controlled overrides; link/redirect validation on slug edits.
- Structured related-content and internal-page selectors rather than fragile pasted destination URLs wherever possible.

- A central source for organisation facts, contributors, approved results, case-study proof, sources and reviewed dates.
- Editorial states (draft, review, approved, scheduled, published, archived), permissions, revisions and rollback.
- Scheduled content reviews and warnings for expired jobs, offers, price claims and older time-sensitive evidence.
- Meaningful publish webhooks for cache invalidation, sitemap updating, SEO QA and IndexNow where applicable.
- Fields for language/region relationships on multilingual sites, and structured product/location fields where relevant.

> **Acceptance test:** Editors can publish and revise standard content without developer intervention, while critical SEO defaults remain predictable. A slug change cannot silently create an orphaned broken URL.

Tier 2 · P1 Conditional

Conditional requirements

11 Activate the right modules for the site

11.1 International and multi-region SEO

Use when content targets more than one language or country.

- Choose a consistent URL strategy (subdirectories, subdomains or country domains) and map each locale to a real landing page.
- Implement valid, reciprocal `hreflang` references, including each page's own language/region variant and `x-default` where appropriate.
- Keep canonical and alternate signals consistent; localised pages should not all canonicalise to a generic market page merely because the topics overlap.
- Localise currency, availability, spelling, contact details and substantive content where relevant; do not create thin regional duplicates solely to capture keywords.
- Test redirect and geo-routing behaviour so crawlers and visitors can access alternate markets without forced IP-based navigation.

11.2 Ecommerce and catalogue SEO

- Define indexing rules for categories, variants, filters, sorting, pagination, out-of-stock products and discontinued items.
- Generate accurate Product/Offer data including currency, price, stock status, variant relationships and permitted review information; keep feed and page facts in sync.
- Provide crawlable category/product links, unique helpful descriptions, product specifications, images, returns/shipping information and visible customer support routes.
- Consider Merchant Center, feeds, images and commerce-agent integrations when commercially appropriate.

11.3 Local SEO

- Define canonical location pages for real staffed branches or genuinely distinct service areas; avoid duplicative doorway pages.
- Maintain consistent business identity, address/telephone/service information and relevant local proof across the site and business profiles.
- Use LocalBusiness schema only for genuine relevant entities and keep opening hours and location details current.

11.4 Editorial publishers, video-led sites and large libraries

- Design topic hubs, author profiles, article archives, navigation and pagination based on actual content volume.

- Set rules for thin taxonomy pages, duplicate tags, expired content, content consolidation and archive indexing.
- Create media metadata, transcripts, appropriate sitemaps and content decay review workflows where useful.

11.5 Service businesses and sector-led site architecture

A suitable service site may have a service hub, parent services, specialist child services, industry pages, helpful guides and case studies. Build only where the intent and evidence support it.

Suggested content relationships:

- **Parent service page:** service proposition, scope, evidence and links to genuinely distinct specialist services.
- **Specialist service page:** differentiated need or platform, process, proof, related guidance and a clear enquiry path.
- **Industry page:** actual experience and sector-relevant examples, not a find-and-replace version of a generic service page.
- **Playbook or guide:** substantial useful advice, firsthand insight and links to appropriate next steps.
- **Case study:** client/context, challenge, work undertaken, timeframe, methodology, exact approved outcomes and limitations.
- **Glossary:** one clear definition per concept when it has genuine search/user value, linked to deeper articles and relevant services.
- **Author/team profile:** substantive biography, authentic credentials and authored material; do not publish empty profiles merely for schema.
- **Trust pages:** contact, company details, policies, terms, privacy, team, methodology and relevant security information.

Independent ranked lists: Comparisons or “best providers” pages are optional, not a required page for every service and sector. Publish them only with defensible selection methodology, genuine usefulness, sufficiently independent evaluation and clear disclosure of the publisher's interests. Do not hide them from normal discovery while attempting to rank them, or create large numbers of near-duplicate rankings.

11.6 Lead generation and analytics

- Define primary conversions before development: qualified enquiry, booked meeting, purchase, lead submission, phone call or relevant other outcome.
- Implement consent-aware GA4/event tracking and configure appropriate conversion events in advertising/analytics systems.
- Verify form submissions, thank-you states, phone/email clicks, consent mode where appropriate, spam protection and cross-domain journeys.
- Preserve UTM parameters through genuine user journeys without allowing tracking variants to become duplicate indexable landing pages.
- Record an organic search baseline before migration and compare performance after launch.

Tier 3 · P2 Experimental

Experimental GEO enhancements

12 Machine-readable publishing experiments

The following are **not universal search requirements**. In particular, Google Search says that `llms.txt`, special AI text files and Markdown alternatives do not improve eligibility or rankings in Google Search. Only build these where a product, integration, partner or clearly defined experiment justifies them.

Experiment	Possible implementation	Guardrails and decision rule
`/llms.txt`	A concise, curated guide to major site pages, grouped by services, guides, evidence and contact information	Generate from approved content; avoid claims of universal crawler adoption; measure actual consumers/use
`/llms-full.txt`	A larger plain-text corpus for systems that request it	Protect licensing, performance and sensitive information; avoid republishing material without permission
Markdown page variants	**Preferred:** content negotiation on the same URL. When a request sends `Accept: text/markdown`, return the Markdown version of that page; otherwise return HTML. **Fallback:** a separate `.md` URL for selected public pages, declared via `rel="alternate" type="text/markdown"`	Ensure the variants are current, equivalent and safe. With content negotiation, send `Vary: Accept` so caches and CDNs never serve Markdown to browsers or HTML to agents that asked for Markdown, and keep the HTML as the default for any other `Accept` value. With separate `.md` URLs, serve each with an HTTP `Link: <html-url>; rel="canonical"` header (or `X-Robots-Tag: noindex`), keep them out of XML sitemaps, and check they do not appear in search results as separate pages
EntityMap	A structured entity map tied to approved public facts and evidence	Treat it as an experimental publishing convention, not an established search-engine directive
Read-only MCP server	Explicit, permissioned access to approved service, product, pricing or documentation data	Require authentication where appropriate, rate limits, monitoring and security review; never expose internal CMS APIs or private data
Agent-focused interfaces	Better form semantics, clear labels, consistent prices, accessible workflows, or purpose-built agent APIs	Prioritise standard web accessibility first; test with actual target agents and permitted use cases

Default for generated files: `llms.txt`, `llms-full.txt` and Markdown variants may be switched on by default when they are generated automatically from the same approved content as the HTML pages, cost little to maintain and cannot drift from the live site. Record them in the P2 register with an owner and a review date. Anything maintained by hand (an EntityMap with quoted evidence, an MCP server, agent APIs) still needs a full experiment brief.

Experiment brief required: State the user/system that consumes the output, what problem it solves, hypothesis, cost, owner, baseline, success measure and review date. Remove or suspend experiments that introduce maintenance or security risk without demonstrated benefit.

13 Search and AI crawler policy

Document separate policies for:

1. **Traditional search indexing:** for example Googlebot and Bingbot, where inclusion is wanted.
2. **AI search discovery:** for example OAI-SearchBot for ChatGPT search; other vendors' published search crawlers where supported.
3. **User-triggered fetchers:** for example ChatGPT-User, which has different behaviour from automatic search crawling.
4. **Model training:** for example GPTBot and other training-related agents; set independently from search discovery according to the site's legal/content policy.
5. **Security and abuse controls:** WAF, CDN, IP rules and rate limits must permit authorised crawling without weakening access protections for private content.
   - Verify crawler names and current publisher controls against each provider's official documentation; do not maintain an unreviewed allowlist copied from another site.
   - Test real bot requests and logs, including legitimate IP verification where the provider publishes ranges.
   - A robots.txt rule is not access control and cannot secure confidential data.
   - Record the policy owner, approval date and a scheduled recheck; vendor behaviour changes.
   - Never serve inaccurate or materially different claims to bots compared with human users.

> **Acceptance test:** Key public templates are accessible to authorised search crawlers through both robots.txt and the CDN/WAF. Training controls reflect the approved policy, not an assumed SEO default.

14 Measure GEO outcomes rather than compliance with formatting myths

- Build a **buyer-intent prompt set** by market, sector and journey stage. Keep the prompt list, test conditions and observation dates consistent enough to compare trends.
- Track **brand inclusion and mention rate**, **site citations**, **competitor visibility**, **answer accuracy** and **sources mentioned**. Treat repeated results as samples, not precise deterministic rankings.
- Monitor **AI referral traffic** and identifiable campaigns in analytics, recognising that not every AI influence or no-click interaction is attributable.
- Use **Google Search Console's Generative AI performance reporting** where available, alongside ordinary organic search reports, rather than assuming AI Overview data must be inferred from general queries.
- Connect visibility to business outcomes: qualified leads, enquiries, sales and assisted journeys where the data permit. Report uncertainty honestly.
- Review inaccuracies and unsupported brand claims surfaced by AI systems; correct underlying public evidence and entity inconsistencies before chasing artificial citation tactics.

Editorial GEO review: Assess whether a page offers a clear answer, original insight, reliable evidence, correctly identified entities, current claims and useful internal context. Do not score it on arbitrary percentages of entity names, answer location or FAQ word count.

Launch

Delivery, launch and ongoing verification

15 Required acceptance gates

Gate A: Discovery and architecture sign-off

- [ ] Confirm target markets, site types and applicable P1 modules.
- [ ] Approve information architecture, page templates, primary conversion events and URL conventions.
- [ ] Baseline rankings, Search Console, analytics, indexing, top landing pages and priority backlinks for migrations.
- [ ] Produce and approve the legacy URL map and redirect plan.
- [ ] Assign owners for development, SEO, content, measurement and release.

Gate B: Pre-launch technical QA

- [ ] No public staging/preview sites are indexable; production hosts resolve correctly.
- [ ] HTML content and crawlable links are present on every priority template.
- [ ] Status codes, canonicals, robots directives and parameter handling are correct.
- [ ] XML sitemaps contain only indexable canonical `200` URLs.
- [ ] Important pages are discoverable through internal links, not just the sitemap.
- [ ] Redirect-map testing shows relevant one-hop destinations and no loops.
- [ ] Titles, descriptions, headings, social previews and schema are checked across templates.
- [ ] Accessibility testing, mobile checks, performance budgets and forms pass.
- [ ] Structured data matches the visible page and validates where applicable.
- [ ] International, ecommerce and local modules are tested where relevant.
- [ ] Source claims, case-study figures, authors and approved brand details are checked.
- [ ] Analytics, consent and primary conversions have successful test events.
- [ ] CDN/WAF rules permit approved bots and do not expose private resources.

Gate C: Launch day

- [ ] Remove staging restrictions from the **correct** public host without unintentionally removing protection from staging.
- [ ] Confirm no accidental global `noindex`, `Disallow: /`, staging canonicals or password challenge on production.
- [ ] Deploy and verify redirects, metadata, sitemaps, robots.txt and live content.
- [ ] Crawl the live site and compare priority URLs with the approved inventory.

- [ ] Verify Search Console/Bing Webmaster Tools access, submit sitemaps and confirm IndexNow where applicable.
- [ ] Verify production analytics and real lead/transaction events.
- [ ] Log launch time, deployed version, exceptions and named approvers.

Gate D: After launch

Within 7 days

- Review server/CDN errors, bot access, indexing warnings, redirect hits, 404s, sitemaps and first conversion events.
- Resolve critical crawl/indexing problems immediately and compare the highest-value migrated URLs with expected outcomes.

At 30 days

- Compare organic landing-page performance, indexing, CTR, priority rankings, conversions and Core Web Vitals against baseline.
- Review AI visibility samples where measurement has been scoped, and prioritise substantial technical/content issues.

At 90 days

- Assess business outcomes and organic/AI discovery trends; review pages needing consolidation, strengthening or new evidence.
- Document recurring technical issues as automated tests and update this build standard.

Continuous checks

Frequency	Automated or manual checks
Every deployment	Sampled statuses, production hostname, canonical, robots, title, content, navigation and schema
Every publish	Valid published page, content/claim validation, sitemap/cache update, broken links, IndexNow where applicable
Weekly after launch	Crawl errors, redirect failures, indexing changes, form/conversion failures and bot blocks
Monthly	Core Web Vitals, internal linking/orphans, content freshness, rankings, search demand, leads and relevant GEO outcomes
Quarterly	Crawler-policy review, template audit, outdated evidence, evolving platform documentation and experimental GEO return on effort

16 Definition-of-done scorecard

A project cannot receive final SEO/GEO sign-off while any P0 launch blocker remains unresolved.

Workstream	Owner	Evidence required
Environments and deployment	Developer / DevOps	Host, access-control and production smoke-test results
URLs and migrations	SEO lead / Developer	Approved URL map; redirect crawl; error report
Crawl and indexing	Technical SEO / Developer	robots, sitemap, canonical and indexability crawl
HTML and templates	Developer	Raw HTML and rendered-DOM template tests
Schema and entity facts	SEO lead / Content owner	Template validation and approved fact register
Content and internal links	Content owner / SEO lead	Content QA, evidence approval, orphan/broken-link report
Performance and accessibility	Developer / QA	CWV/lab results, accessibility findings and remediation
Analytics and conversions	Analytics owner	Test events and production conversion verification
P1/P2 scope	Project owner	Signed applicability decision and experiment briefs
Release approval	Project owner	Completed gates, exceptions, approvers and release date

Approved exceptions: Record the requirement, reason, business risk, compensating control, owner and target resolution date. Never silently waive P0 requirements.

Reference documents and maintenance

This playbook is a practical implementation standard, with project-specific decisions where required. Platform documentation and features change. Review the source guidance at least quarterly and after major Google/Bing/OpenAI updates.

- Google Search Central: AI features and your website
- Google Search Central: Optimising for generative AI
- Google Search Central: JavaScript SEO
- Google Search Central: Canonical URLs
- Google Search Central: robots.txt and noindex
- Google Search Central: International hreflang

- Google Search Central: Structured data policies
- Google Search Central: Qualify outbound links
- Google Search Central: Core Web Vitals
- OpenAI: Crawlers
- OpenAI: Publisher and developer FAQs
- IndexNow: Protocol documentation

Principle: Reliable crawling, genuinely useful content, accurate evidence and measurable outcomes take precedence over GEO files or formatting conventions that have not demonstrated benefit.

MarketingSignals

Want help applying this standard?

Marketing Signals is a search marketing agency, running since 2007, covering SEO, PPC, Digital PR, AI SEO and GEO. Our own website, marketingsignals.com, is built to this standard.

Every requirement in this standard is also in a free checklist Google Sheet, ready to copy, assign and track.

Web marketingsignals.com
Email contact@marketingsignals.com
Phone 0330 043 4676

Free to download, use and share. Platform documentation changes: check the online version for the latest release.