

From Keywords to Meaning

*How web vocabulary and product identification make
the internet AI-proof*

Version 1.0, 08/05/2026

Document Summary

Document Item	Current Value
Document Name	From Keywords to meaning
Document Date	08/05/2026
Document Version	1.0
Document Issue	
Document Status	Pre-version
Document Description	Guideline

Contributors

Name	Organisation
Rien van den Bosch	GS1 Belgium & Luxembourg

Log of Changes

Release	Date Change	of	Changed By	Summary of Change
1				
2				
3				

Disclaimer

GS1, under its IP Policy, seeks to avoid uncertainty regarding intellectual property claims by requiring the participants in the Work Group that developed this manual to agree to grant to GS1 members a royalty-free license or a RAND license to Necessary Claims, as that term is defined in the GS1 IP Policy. Furthermore, attention is drawn to the possibility that an implementation of one or more features of this Specification may be the subject of a patent or other intellectual property right that does not involve a Necessary Claim. Any such patent or other intellectual property right is not subject to the licensing obligations of GS1. Moreover, the agreement to grant licenses provided under the GS1 IP Policy does not include IP rights and any claims of third parties who were not participants in the Work Group.

Accordingly, GS1 recommends that any organization developing an implementation designed to be in conformance with this Specification should determine whether there are any patents that may encompass a specific implementation that the organization is developing in compliance with the Specification and whether a license under a patent or other intellectual property right is needed. Such a determination of a need for licensing should be made in view of the details of the specific system designed by the organization in consultation with their own patent counsel.

THIS DOCUMENT IS PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF THIS SPECIFICATION. GS1 disclaims all liability for any damages arising from use or misuse of this Standard, whether special, indirect, consequential, or compensatory damages, and including liability for infringement of any intellectual property rights, relating to use of information in or reliance upon this document.

GS1 retains the right to make changes to this document at any time, without notice. GS1 makes no warranty for the use of this document and assumes no responsibility for any errors which may appear in the document, nor does it make a commitment to update the information contained herein.

This document is a time-bound guideline and reflects the state of knowledge at the time of publication. Due to the continuous evolution of technology, insights and recommendations may change. Each version of the document is therefore dated and versioned.

Introduction

Search behaviour is changing fundamentally. For years, companies optimised for search engines through SEO and content. Today, however, the reality is shifting towards AI-driven search: systems that no longer merely index products and information, but also understand, compare and summarise them. In that context, it is no longer enough to be “easy to find” through keywords. Machines need structure, meaning and unambiguous identification.

This paper explains why web vocabulary – standardised, machine-readable descriptions of product information – plays a crucial role in this evolution, and why product identification through official identifiers such as the GTIN (Global Trade Item Number) is key for companies to remain visible, trustworthy and future-proof within the digital ecosystem. The paper is primarily intended to help organisations move from insight to action. That is why, alongside the why and the what, it also answers the question that comes up after every presentation: how can you approach this concretely, without immediately turning it into a major IT project?

1 From search engine to AI assistant

For a long time, online search was a relatively simple and linear process. Users entered a query, search engines scanned their index and presented a list of web pages ranked according to relevance. Companies optimised their online presence around this mechanism: high-quality content, relevant keywords and effective technical SEO largely determined a page’s visibility in search results.

Today, this paradigm is evolving profoundly. Search interfaces are increasingly becoming AI-driven assistants that no longer simply provide links, but generate direct answers. Think of search engines that summarise results, compare products, make contextual recommendations or even prepare complete decisions for the user. Users are no longer searching for web pages containing information; they receive the information directly from an AI system.

In this new model, the role of the web page changes fundamentally. It is no longer the final destination of the search, but rather an information source that AI systems analyse, combine and reuse. The central question for these systems is no longer: “Which words does this page contain?”, but: “Which objects, products or entities can I identify and compare with certainty?”

Concrete example: the question of “the best washing machine”

Imagine someone asks: “Which energy-efficient washing machine is suitable for a family of four?” A traditional search engine can easily display links to online shops. An AI assistant aims to do something different: create a shortlist, compare models, suggest trade-offs and perhaps even recommend the “best choice”. This is only possible if the system can determine with certainty which products exist, which characteristics are linked to them (energy rating, capacity, dimensions) and whether the product is actually available for sale.

This is the core of the shift: visibility depends less and less on “having a good page” and increasingly on “being understood as a product”, with unambiguous characteristics and a reliable identity.

2 The problem: Fragmentation and Interpretation

Although AI systems are becoming increasingly sophisticated, they remain fundamentally dependent on the quality of the information they process. In practice, they face a structural problem: product information on the web is highly fragmented and inconsistent.

Today, the same product appears simultaneously on the manufacturer’s website, in multiple webshops, on marketplaces, on comparison websites and, increasingly, in datasets consulted by AI systems. In theory, these are identical products, but in practice they are often described in very different ways. Product names vary, specifications are incomplete, images are cropped differently, prices appear in another context, and fixed identification is missing.

For human users, this rarely causes major issues. People recognise patterns, make connections and understand that “X3000 professional device” is probably the same washing machine as “Model X Pro

3000". For machines, this interpretation is far less obvious. Without explicit signals, an AI system has to make assumptions. And when in doubt, the system will often choose caution: not comparing, not recommending or ignoring the information altogether.

In an AI context, uncertainty becomes a barrier. While humans can tolerate ambiguity, machines require clarity. As a result, products lose visibility, not because they are irrelevant, but because they cannot be identified unambiguously.

Example (webshop reality): one product, three names

A running shoe is listed as "Ultraboost 22" on the brand website, as "Ultraboost 22 – comfort running" on a reseller webshop, and as "Adidas UB22" on a marketplace. For a human, this is still understandable. For systems, however, uncertainty arises: are these variants, the same shoe, or different models? If the identifier is also missing, or replaced by internal SKUs, there is no reliable anchor to connect the three listings.

3 What is web vocabulary?

To understand why web vocabulary plays such a central role in AI-driven search, it is important to first consider how information on the web is consumed by machines today. While people read and interpret web pages based on context and experience, machines process information in a fundamentally different way. They rely on explicit markers that clearly define what something is in a structured and unambiguous manner, rather than interpreting it from free text alone.

Web vocabulary is designed precisely for that purpose. It is a collective term for standardised data models and vocabularies that enable information on the web to be described in a machine-readable way. For product information, this is primarily achieved through schema.org, a globally supported vocabulary used by search engines and many online platforms.

In essence, web vocabulary acts as a translation layer between human language and machine interpretation. While a product description written in free text can be ambiguous, web vocabulary explicitly defines that it is a Product, with properties such as brand, dimensions, price, availability and review rating. As a result, machines no longer have to infer this information but receive it in a structured and explicit way.

GS1 Web Vocabulary builds on this foundation. It aligns with schema.org wherever possible, while extending it where generic vocabularies fall short in supporting accurate product identification. A key element is the explicit use of official identification keys such as the GTIN. This ensures that product information is not only structured but also unambiguously linked to one and the same product, regardless of where that information appears.

Web vocabulary is therefore much more than an additional layer of metadata. It provides the semantic backbone that ensures product data retains its meaning in an increasingly complex digital landscape.

4 Web vocabulary is not classical SEO

Because web vocabulary is often mentioned in the same breath as SEO, it is frequently perceived as simply another optimisation tactic within the same framework. In reality, however, the two concepts differ fundamentally in purpose, approach and functionality.

Traditional SEO evolved from the need to make web pages more visible. Its focus lies on content, keywords, internal structure and external signals such as links. At their core, these techniques are aimed at human interaction: they are designed to ensure that a user finds a page, clicks on it and takes action.

Web vocabulary starts from an entirely different premise. It is not designed to make pages more attractive to people, but to make information understandable and reliable for machines. Where SEO relies on interpretation and probability, web vocabulary strives for unambiguous meaning. The objective is not to "rank higher", but to be recognised correctly.

In an AI context, this distinction is becoming increasingly important. AI systems do not rely solely on rankings or keywords; they select information they consider sufficiently reliable to reuse in summaries, comparisons and recommendations. A product may be described perfectly and perform extremely well in traditional search results, yet still remain invisible in AI-generated answers if the underlying structure is missing.

SEO and web vocabulary are therefore not opposites, but complementary layers. SEO remains essential for attracting users and providing context. Web vocabulary forms the structural layer beneath it, determining whether machines can correctly interpret and reuse the information.

Mini case: “perfect SEO, yet invisible”

A webshop invests in better product descriptions, fast loading times and strong meta tags. The page ranks well. Yet in an AI-generated overview comparing several products, the product does not appear. Not because the content is poor, but because the system cannot model the product as a reliable object (no identifier, no structured fields). This is typically the moment when companies realise that “SEO alone” is no longer enough.

5 The minimum to get started : From principle to practice

When the conversation turns to web vocabulary and structured data, many companies perceive it as a complex, technical and costly undertaking. In reality, the barrier to entry is lower than often assumed. The strength of web vocabulary does not lie in achieving completeness from day one, but in building the right foundations.

In essence, only a limited set of actions is needed to give machines a fundamentally better understanding of product information. The first and most important step is having a unique and stable product identifier. Without such identification, every product description remains contextual and temporary. With an official identifier, such as a GTIN, a product gains a fixed digital identity that is independent of marketing names, languages or channels.

The next step is a consistent product page where basic information is clearly available: a clear product name, an image, pricing information and availability. Most webshops already contain this information today, but it is often intended solely for human visitors. By also presenting this information in a structured format – for example through JSON-LD – it becomes explicitly readable for search engines and AI systems.

Adding structured data does not mean that existing content needs to be rewritten. It involves creating a parallel layer in which the same information is provided in a formal and unambiguous format. It is crucial that the GTIN is explicitly included as an identifying field. For machines, this makes the difference between “a product description” and “this specific product”.

5.1 Case: One product page, two worlds (human vs machine)

On screen, a visitor simply sees, for example:

- * product name
- * price
- * stock status
- * product image

For a machine, however, this is not automatically recognised as “data”. With structured data, the same page gains an additional layer that specifies: this is a Product, this is the GTIN, this is an Offer with a price and stock status.

Do: publish the information you already display, but in a structured format (name, brand, price, stock, GTIN).

Don’t: include information in structured data that is not visible on the page (for example, marking up reviews that are not displayed), or replace the GTIN with an SKU.

5.2 A practical starting approach (without project stress)

The most realistic way to get started is with a small pilot: one category or a selection of bestsellers. This makes it possible to quickly verify whether your GTINs are correct, whether your product fields are consistent, and whether your markup is coherent. Only once this pilot is stable does it make sense to scale up. This approach is often more efficient than activating everything at once and then spending weeks correcting errors.

6 Why product identification is critical

In a digital environment increasingly driven by AI, the focus is shifting from “findability” to recognisability. Products are no longer simply searched for; they are actively processed: filtered, compared, merged and linked to other data sources such as regulations, sustainability or safety information. In all of these processes, product identification plays a decisive role.

Without unambiguous identification, product information remains dependent on context. A product name can change, a description can be updated and a language can differ, but for a machine these are all new signals that need to be interpreted. This leads to uncertainty and fragmentation. With an official identification key such as the GTIN, however, a product gains a stable reference point that remains recognisable across time and systems.

That stability is crucial for AI systems. When the same GTIN appears on a webshop, a marketplace and within an external dataset, an AI system can determine with a high degree of certainty that it concerns exactly the same product. This not only enables reliable comparison, but also the aggregation of information: characteristics, reviews or sustainability data can all be linked to a single product instead of remaining fragmented.

6.1 Case Study: SKU versus GTIN (Why “Internal” Is Not Enough)

Many webshops use SKUs (internal stock keeping units) as identifiers. This works perfectly within their own systems, but it is not designed to function as a global identifier. On marketplaces and within AI comparison systems, this is often meaningless, because a SKU from webshop A has no meaning for webshop B.

Do: use SKUs for internal management, but publish the GTIN as the official identifier for the outside world.

Don't: enter an SKU in the GTIN field, or assign the same GTIN to multiple different products simply because it is convenient.

6.2 A bridge to regulation and the Digital Product Passport (Without the Doom Scenario)

Product identification is not only relevant for visibility today, but also for compliance and transparency tomorrow. The Digital Product Passport is explicitly based on the principle that product information must be consistent and reusable throughout the entire product lifecycle. Without a stable identifier, this becomes impossible. Identification is therefore not a detail, but a strategic choice that enables digital continuity.

7 Web Vocabulary as the foundation for the future

While web vocabulary is still often approached today from the perspective of findability and online visibility, its relevance in reality extends much further. It forms the structural foundation for a broader digital transformation in which product information is no longer published statically, but dynamically reused across different systems, applications and contexts.

An increasing number of digital ecosystems expect product information not only to be available, but also standardised and machine-readable. Marketplaces are building their processes around structured product data, AI systems require reliable identifiers to make comparisons, and consumers expect transparency regarding origin, composition and impact. In this context, web vocabulary acts as the connecting layer.

Regulation is also reinforcing this evolution. Initiatives such as the Digital Product Passport are based on the need for consistent, reusable product information throughout the entire lifecycle of a product. Web vocabulary makes it possible to avoid repeatedly translating or restructuring this information, by defining it correctly once and then reusing it across a wide range of applications. In that sense, web vocabulary supports not only compliance, but also operational efficiency.

Example: “Get it right once, use it everywhere”

Companies that structure product information for search engines today often discover that the same data quality later also benefits marketplaces, product feeds, QR landing pages or compliance workflows. The advantage is not limited to a single channel, but lies in the ability to maintain the same product identity and core data consistently everywhere.

8 The role of neutral standards

In a digital ecosystem characterised by a multitude of platforms, technologies and interests, neutrality is a crucial requirement. Without shared agreements on how information is structured and interpreted, data risks becoming trapped in silos or platform-specific solutions.

Neutral standards provide a counterbalance to this. They ensure that product information retains its meaning, regardless of where it is used or by whom it is consumed. By working with open and widely adopted standards, companies can innovate without becoming locked into one specific technology or commercial player.

Within this framework, GS1 plays a facilitating role. By managing globally recognised identification standards and translating them to the web through GS1 Web Vocabulary, GS1 contributes to a consistent and reliable data landscape. The focus is not on creating new silos, but on connecting existing ecosystems, from physical supply chains to digital platforms and AI applications.

9 Practical Implementation

This chapter answers the question that always follows: “Alright, but how do we actually do this in practice?” You do not need to be a developer to understand what is required. At the same time, it is concrete enough to serve as a briefing for an agency or IT partner.

9.1 The practical step-by-step approach (for every webshop)

Step 1 – Make sure your product identification is correct (GTIN discipline).

Start with the basics: does every product have a GTIN, and is that GTIN correctly included in your product data? This sounds simple, but this is where things often go wrong. If a GTIN is missing or used incorrectly, an AI system cannot reliably recognise your product.

Step 2 – Make your core fields consistent.

A machine relies on consistency. If the price displayed on the page differs from the one in your markup, or if your stock status is incorrect, you create uncertainty. And in a machine context, uncertainty often leads to the information being ignored.

Step 3 – Publish your core fields in a machine-readable format (structured data).

Structured data is an additional “language” on your page designed specifically for machines. You add it to the source code of your product page, usually as a script block. The standard format is JSON-LD.

Step 4 – Test with tools and make it maintainable.

You do not need to test this manually every day, but you do need a routine: whenever templates change, plugins are updated or pricing logic is modified, you should verify that your markup is still correct.

9.2 Glossary

Structured data: a way of “labelling” information on your page so that machines understand what it is. Not text they need to interpret, but fields with defined meaning.

Schema.org: the most widely used “vocabulary” defining which labels exist (Product, Offer, Brand, etc.).

JSON-LD: the format in which structured data is usually published. It is literally a block of JSON embedded in your HTML. Your visitors do not see it, but machines do.

Product / Offer: in schema terminology, a Product is the object itself (the shoe, the washing machine). An Offer is the commercial offer (price, currency, stock status, link to the product page).

Rich results: enhanced search results in which information such as price or stock status becomes visible. You do not “request” this from Google; you make your page technically eligible for it.

9.3 Case 1: Webshop with one offer per product page

Situation: one product page displays one price and one stock status.

Do: publish a Product with an Offer.

```
<script type="application/ld+json">
{
  "@context": "https://schema.org/",
  "@type": "Product",
  "name": "Adidas Ultraboost Hardloopschoenen",
  "image": "https://www.jouwwebshop.be/images/ultraboost.jpg",
  "brand": { "@type": "Brand", "name": "Adidas" },
  "gtin13": "4066751234567",
  "offers": {
    "@type": "Offer",
    "priceCurrency": "EUR",
    "price": "179.00",
    "availability": "https://schema.org/InStock"
  }
}
</script>
```

Don't: include pricing or stock information in the markup that does not match what the user sees on the page.

9.4 Case 2: Variants (size/colour) and why this often goes wrong

Situation: one product has multiple variants (for example shoe sizes 42, 43, 44).

In practice, there are two robust approaches:

1. Variant URLs: each variant has its own URL and its own GTIN (ideal).
2. One URL, multiple variants: in this case, you need to ensure very carefully that the correct variant information (and identifiers) remains consistent across your feeds and what is displayed on the page.

Do: define internally how variants are identified (GTIN per variant) and publish this consistently.

Don't: use one GTIN for "the model" when the variants actually have separate identifiers, or identify variants differently on marketplaces than on your webshop.

9.5 Case 3: Marketplace & Comparison – Why the GTIN is your "matching anchor"

Situation: your product appears both on a marketplace and on your own webshop. The marketplace wants to match products so that reviews, specifications and sellers can be grouped together.

Do: ensure that the same GTIN always represents the same product everywhere and that your structured data also publishes this GTIN.

Don't: use internal numbers or temporary codes instead of a GTIN. This leads to duplicate product pages, poor matching and reduced visibility in comparisons.

9.6 What you should ask an agency (A briefing that actually works)

Many clients ask agencies to "do something with SEO and AI". A better briefing is concrete and measurable:

- Publish a JSON-LD Product with an Offer on every product page.
- Explicitly include the GTIN (GTIN-13 or GTIN-14).
- Ensure that the price and stock status in the markup exactly match the visible page.
- Ensure there is a single source for schema output (no duplicate markup caused by multiple plugins).
- Define a set of reference URLs (for example 20) that must be technically correct.

9.7 Maintenance & Governance (The step that makes everything work)

Without governance, structured data often becomes nothing more than a one-off "SEO action". A sustainable approach requires defining:

- who owns the product data,

- who approves changes,
- how often quality checks are performed,
- and how releases are tested.

This does not need to become overly complex. Even a simple agreement (“product data owner”, monthly spot checks, testing after template changes) prevents quality from gradually deteriorating.

Conclusion

AI-driven search makes one thing clear: visibility is no longer a marketing trick, but a data challenge. Companies that merely describe their product information risk becoming invisible to systems that need to understand and reuse that information.

Web vocabulary and product identification provide a concrete and achievable answer. Not as a future vision, but as a practical way to make existing product information suitable for a machine-driven web. Companies that invest today in clear identification and structured product data are building trust, scalability and future resilience, regardless of which search interface or regulation becomes the standard tomorrow.

This paper is intended as a guiding document for companies, web agencies and policymakers who want to understand how visibility, data and identification come together in the era of AI-driven search.