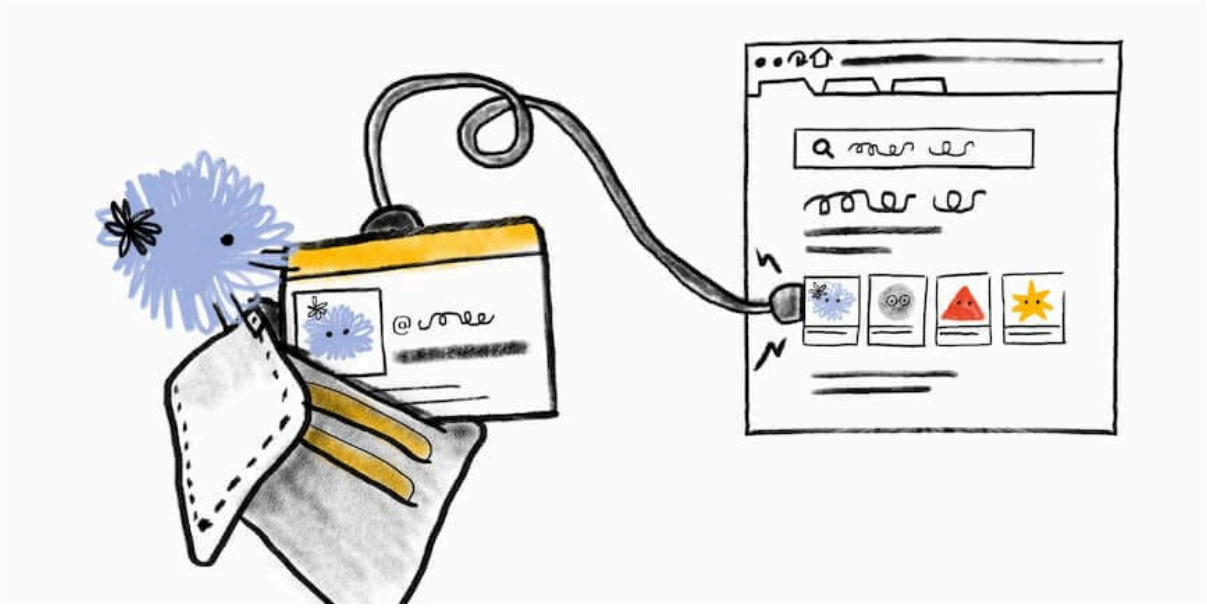
Structured Data 2025



Content team

*Hello content team! This is your personal doc to collaborate on and plan the contents of your chapter. Click **Request edit access** above to get started.*

Please add your name and email address below so we can @tag each other in the comments. You can also subscribe to all comments by opening the comment history, clicking the notification bell, and selecting All.

***Authors:** …*
***Reviewers:** …*
***Analysts:** …*

The objective of your chapter is to write a data-driven answer to this big question:

"What is the state of Structured Data in 2025?"

Learn more about the chapter lifecycle and refer to your chapter's tracking issue on GitHub for more info. Thank you all for your contributions! Official call

Outline

The purpose of this section is to define the scope of the chapter by creating an ordered list of all of the topics to be explored. You can think of this outline as the chapter's table of contents. This list will become your narrative, so consider how the content should be sequenced and how much additional depth is needed for major topics. You may choose to start with last year's outline and add or remove content as needed. Every chapter must have an introduction and conclusion, but everything in between is up to you.

Every chapter must also be data-driven, so for each topic in the outline below, clearly enumerate which metrics you'll need to substantiate your narrative. Work with your analysts to clarify what data is needed and how the results should be formatted. For example, if you're measuring the usage of a particular HTTP header value, you can measure it as the percentage of pages having that header, as the percentage of headers having that value, as a distribution of values, what the largest value is, etc. Clarify those expectations upfront so that the analysts know how to write the corresponding queries and whether the metrics are even feasible in the dataset.

First meeting to outline the chapter contents by June 1
Custom metrics completed by July 1
HTTP Archive crawl by July 1
Querying all metrics and saving the results by September 1
First draft of chapter by October 1
Reviewing & Editing of chapter by October 20
Publication of chapter (Markdown & PR) by November 15

Proposed 2025 Structured Data Chapter Outline

1. Introduction: From Metadata to a Machine-Executable Web

- **A Look Back:** Briefly recap the established role of structured data in powering search engine rich results, referencing key adoption metrics from the 2024 chapter.
- **The 2025 Thesis: The Great Translation.** This year's central theme is the maturation of structured data from a passive descriptor for search engines into the **foundational grounding protocol for AI**. We'll argue that for a website to be relevant in the new AI-driven landscape, it must be not just human-readable but **machine-executable**. We are witnessing the "API-ification" of the web page, where structured markup is the contract that makes content understandable and actionable for AI agents.

---

2. The Mechanics: Why Structure Unlocks AI Reliability

- **From Conversation to Contract:** Explain the core problem that structured data solves: the inherent non-determinism of LLMs. Contrast a vague natural language prompt with a precise JSON-based instruction, framing structured data as an enforceable "contract" that guarantees predictable, machine-readable outputs.
- **Speaking the Native Language:** Introduce the "Training Data Hypothesis" — that LLMs are highly effective with formats like JSON because their training data is saturated with code, APIs, and configuration files, making it a "native tongue."
- **Constrained Generation: Guaranteed Reliability.** Briefly explain how modern LLM APIs (e.g., Google Gemini, OpenAI) now offer schema-enforced output modes. This moves structured data from a suggestion to a **programmatically guaranteed format**, eliminating parsing errors and making AI a reliable component in automated systems.

---

3. State of the Web: Adoption & Vocabulary

- **The Big Picture: Adoption Rates.** Analyze HTTP Archive data for year-over-year growth in structured data usage on both desktop and mobile pages.
- **Formats & Syntax:**
    - **JSON-LD's Absolute Dominance:** Confirm JSON−LD's position as the universal standard, showing its market share versus the legacy formats of Microdata and RDFa.
    - **Vocabulary Landscape:** Highlight the most-used `schema.org` types (`Product`, `LocalBusiness`, `Recipe`, etc.) and note any significant new additions or trends in vocabulary since the last analysis, reflecting emerging areas of interest (e.g., `FinancialService`, `PublicAct`).

---

4. The Rise of the AI Consumer: From Search Snippets to Agentic Execution

- **The First Wave: AI-Powered Search.** Discuss how Google's SGE and Microsoft's Copilot leverage schema not just for rich snippets but for generating synthesized

answers. Analyze Google's deprecation of certain schema types as a strategic move to focus on high-value, verifiable data.

- **The New Wave: Native AI Agents.** This section moves beyond traditional search engines to the new class of AI consumers.
    - **Case Study: Live Model Analysis (e.g., GPT-OSS-120B).** Present findings from recent experiments showing how new models natively fetch web content and **directly parse embedded JSON−LD** to build synthesized knowledge panels. This provides direct evidence that a page's schema is being ingested as its primary source of truth.
    - **From Keywords to Entities:** Emphasize the critical strategic shift. AI agents retrieve **"things, not strings."** A page succeeds by clearly defining its entities and their attributes through structured data, not just by ranking for keywords.
- **SEO 3.0: Optimizing for Agents.** Introduce the concept of **SEO 3.0** as the necessary evolution of the discipline. Success in the agentic era depends on making a site's content and purpose unambiguously machine-readable through robust structured data.

---

5. Building the Agentic Web: Protocols and Challenges

- **The Vision: A Web That Does Things.** Briefly define the "Agentic Web" as an executable layer where autonomous agents accomplish goals. Contrast its pragmatic, bottom-up adoption (using existing standards like JSON and Schema.org) with the top-down vision of the original Semantic Web.
- **Protocols and Vocabularies for the Future:** Introduce emerging standards designed to formalize agent communication. This includes protocols like the **Model Context Protocol (MCP)** for tool use and **Agent-to-Agent (A2A)** for inter-agent collaboration, as well as vocabulary initiatives like **SEOntology**—a potential extension of schema designed to provide the nuanced, goal-oriented data that agents require for complex web interaction tasks.
- **Implementation Hurdles:**
    - **The Quality & Validation Gap:** Discuss the persistent challenge of incorrect or incomplete schema and the heightened importance of validation tools. In the agentic era, bad data isn't just a missed rich snippet; it's a broken API call for an agent.
    - **Adversarial Tactics & Security:** Address the risks of structured data spam and how it could be used to mislead or manipulate AI agents, creating a new and complex security landscape.
    - **The Reasoning Limit:** Acknowledge that while LLMs can process structure, they lack true symbolic reasoning. Hallucinations can still occur, even within a perfectly valid schema, creating a need for human oversight and verification.

---

6. Conclusion: The Human-Agent-Data Triad

- **The Virtuous Cycle:** Summarize the key findings through the lens of a self-reinforcing cycle: the need for reliability led to structured prompting, which enabled robust AI applications, which in turn drives the adoption of public structured data as a core **SEO 3.0** practice, creating the foundation for the Agentic Web.
- **The Road Ahead: From Operator to Architect.** Conclude that the focus has irrevocably shifted from simply marking up content to enabling machine action. In this new paradigm, the human role evolves from a direct operator to a **system architect**, responsible for defining the goals, tools, and structured data schemas that guide autonomous agents. The future of the web will be defined by the quality of this structured collaboration between humans and AI.