



The Global Language of Business

# The benefits of using structured data

The role of GS1 standards in enabling seamless customer experiences



Consumers' appetite for multichannel retail is posing many challenges to merchants and brand owners, but one thing they are all certain of is the increased power of digital channels. Over time they have also become aware of the need to not only improve how their product data is structured online – but also to better link it to the physical products.

This is proving increasingly important in order that their goods appear prominently in search engine results and that the associated information is accurate and sourced from reliable sources that the retailers/brand owners trust.

This data is also vital to the operation of the growing number of smartphone apps that are frequently used by consumers while shopping in physical stores - which means the need to ensure consistency of product information across channels is critical.

To help enable this, GS1 is using its experience in the physical channel to deliver a digital solution. Over its 40 years of bringing together brand owners, manufacturers, distributors and retailers to work together on common standards for exchanging information within physical supply chains through barcodes, this GS1 community has developed extensive detailed data models and dictionaries for describing products, services and organisations.

The next step on our journey is a GS1 Digital Initiative that involves creating similar data models that are then made available as a web vocabulary for use online (utilising structured data).

Brand owners who apply barcodes to products lease a GS1 Company Prefix that enables them to create a series of globally-unique product identifiers called Global Trade Item Numbers (GTIN). The use of GTINs provides the consistent cross-reference across information that has been provided by the brand owner, retailers and resellers of the product. It is the common thread that can link the physical product with its online equivalent.

This unique identifier also enables search engines to quickly determine which data about a product is consistent – and which information is likely to be misleading. The GTIN makes it possible to confirm whether data about a product came from a trusted source, typically the brand owner or manufacturer.

Dr Mark Harrison is director of Auto-ID Labs within the University of Cambridge. He has been leading an initiative funded by GS1 to look at how global identification standards can be applied in the digital space. Harrison says: “The GTIN is the common identifier that links information about products. When a search engine looks across multiple websites it can look for consistent information about the GTIN and detect any different information. If one person describes a particular product incorrectly with a GTIN then the search engine will notice the inconsistency and can give it a lower rank in the search results.”

“GS1 standards have been invaluable in our efforts to improve the online shopping experience by harmonizing product data and eliminating inaccuracies.”

Eric Tholomé, Product Management Director, Google



'Using GS1 standards within structured data' - <http://youtu.be/1rGnv6DoA9s>

## Structured data and standards

To ensure this happens it is necessary for retailers and brand owners to put in place the digital infrastructure to be able to link their various data properties (such as technical specifications, ingredients, nutritional information etc) relating to their products.

The essential facts and figures contained within a webpage can be expressed using linked data standards developed by the [World Wide Web Consortium \(W3C\)](http://www.w3.org/).



This embedding can be done more easily via a single block of structured data within the header of a webpage using a mark-up format such as JSON-LD [JavaScript Object Notation for Linking Data]. This enables structured data to be encoded within a webpage in such a way that computer software – a search engine, for example – understands how to extract it automatically.

At the moment, retailers and brand owners are predominantly representing their product data in an unstructured format on their websites, which involves them creating their own individually-designed lists and tables of information that is very difficult for search engines to negotiate. In contrast, structured data involves holding information in a consistent format that is easily readable by software including search engines, irrespective of the visual layout of the webpage. “The facts and figures can be extracted from the page and there is no worry about how the data is laid out by the retailer on their webpages. The layout for the human-readable parts of the pages can be changed but the structured data [embedded] within it will remain intact,” explains Harrison.

The final element in creating the structured data framework for individual products (in webpages) is to assign values to various attributes or properties relevant to a particular type of product. This is done

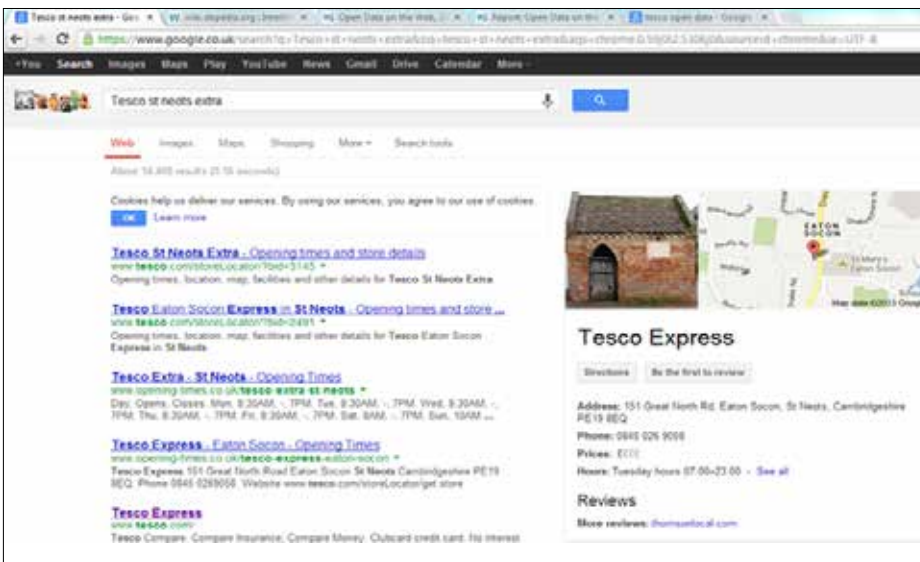
by using web vocabularies that consist of lists of characteristics such as packaging size, allergens, calories etc that can be populated with information by the retailer and brand owner. A web vocabulary also provides a definition for each characteristic or property.

The established de facto standard vocabulary is Schema.org (created by a consortium of the world's major search engines including Google, Yahoo, Bing and Yandex). This will be complemented by a new detailed web vocabulary that GS1 is developing for specifically describing products and services in rich detail.

Using a combination of the Schema.org and GS1 vocabularies, it is possible to associate the structured data about a product with its globally-unique GTIN and for rich structured data about the product to then be put online, made available to search engines, smartphone apps, etc.

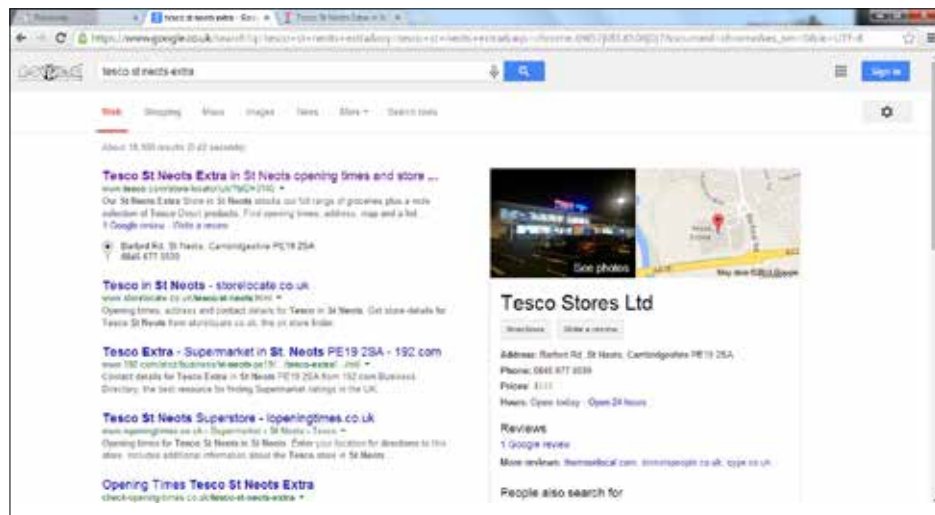
“Making data available in a standardised format and using unique identification for products is a no-brainer from a marketer’s perspective. We believe GS1’s semantic web initiative can be a really positive initiative for brands, manufactures, retailers, advertisers and of course consumers.”

James Grant, Search Manager, Performics



Google Rich Snippet search result for St Neots before using structured data

Google Rich Snippet search result for St Neots after using structured data



# In-practice – Rich Snippets

Established companies are only just beginning to consider embracing semantic web/linked data technologies and taking a structured data approach. For many of these organisations, the initial objective is to benefit from enhanced search results, such as having products appear within Google ‘Rich Snippets’ – these are the additional information extracts that appear both underneath and to the right of search results to give a better sense of what is relevant to the user on the pages that have been listed. This is only possible if Google understands the content on the pages it is interrogating.

At present Harrison says Google is generating Rich Snippets as a “free bonus in exchange for providing structured data online”, which is an incentive to retailers and brand owners to take this approach.

“Would you rather have a free enhanced ‘feature article’ for the search result about your product or retail store – or would you really prefer a plain text listing with a few lines and a boring blue hyperlink?” he asks.

Tesco has been experimenting with using linked data for its store location information - after finding that some information was being pulled in from various inconsistent sources. After just one month of utilising structured data, the results displayed in Google Rich Snippets were found to be far more accurate and reliably sourced, including correct locations on maps and accurate photos of the stores (see above).

US-based electrical retailer Best Buy has reported significant improvements in its web traffic from improved search results after it also experimented with

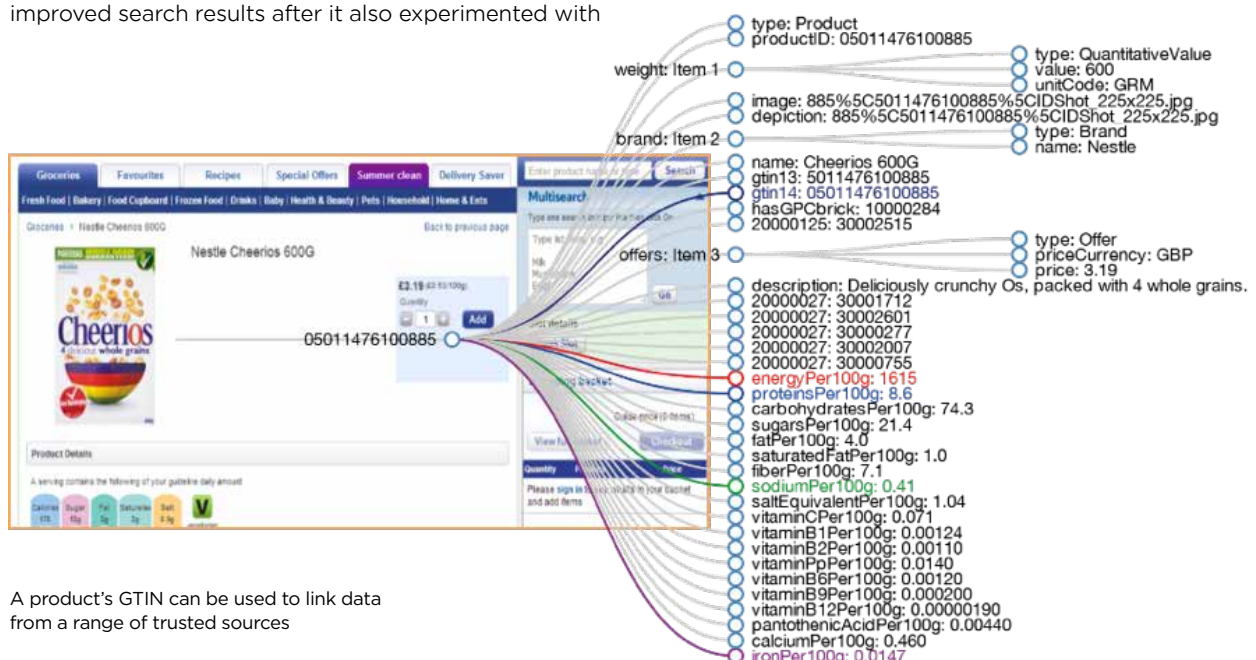
structured data. It has observed a 30% increase in web traffic to its store location pages.

But it is not just the major retailers and brand owners who can benefit from adopting a structured data approach to their products and services. “Even the smallest players can use it to improve their visibility from enhanced search results. Google Rich Snippets for instance offers the same size of screen real estate for all companies. It’s a level playing field for all businesses,” says Harrison.

Although it can seem a daunting prospect to undertake the task of implementing this approach to web data, many food retailers are already making significant changes to their information online as a result of the new EU food legislation (1169/2011) – effective from December 2014.

It requires purchasers of food items online to have access to the same information that is available on the physical labels of products – that includes ingredients, nutritional information, and other potentially important data about the items.

“This is driving retailers and brand owners to provide ever richer data on their webpages. The legislation does not force them into using structured data. However, if they make the relatively small extra effort required, then they can get the full benefits that a structured data approach brings. They’ll not just be doing it for compliance purposes but to also improve the search results of shoppers,” suggests Harrison.





# Future potential

With extra legislation in place, it clearly makes sense for grocers and food manufacturers to adopt structured data at this stage, but it will also likely be taken up in the near term by clothing and consumer electronics merchants as some of their items are sufficiently highly valued as to warrant lengthy online research by consumers.

It is therefore a sensible approach to provide structured data, with detailed product characteristics, as this enables much more detailed and accurate comparisons to be undertaken - through more accurate results from the search engines. This might not just be on price but on more specific technical parameters such as battery life and camera resolution.

Organisations thinking this way about utilising structured data are putting themselves in line to reap longer-term benefits, as shoppers using the next generation of smartphone apps and in-store scan-and-collect devices will ultimately enjoy an improved multichannel experience.

In Norway, Norgesgruppen has already developed a sophisticated digital shopping list app that enables consumers to make intelligent choices about products based on their individual dietary requirements, preferences on health, environmental and ethical issues.

Enabling customers to use apps to search on such things while in-store after scanning an optical barcode or QR code can clearly be very effective, according to Harrison: "A QR code could take the consumer to a structured data page where the app will have access to all the detailed product information. The customer could also be alerted to promotions and other alternative products could be highlighted."

After-sales service could also be enhanced in myriad ways. If barcodes or tags are included on electrical products, a simple scan using an app could take the shopper directly to associated content such as the instruction manual. And it need not all be one-way, as consumers could also then provide feedback on the product that can then be directly linked to that item through the linked data approach.

# Supporting adoption of structured data

Some companies are still unclear about how to proceed with this new unfamiliar technology – or are concerned about whether they will somehow relinquish control if they publish rich information on their products in a structured data environment.

Harrison suggests the reality is that initially much of the structured data on products will not be information that is commercially sensitive, but information that is already in the public domain because it already appears on the packaging or possibly in human-readable format within public webpages.

To encourage adoption of the structured data approach, GS1 is set to introduce standards in the summer of 2015. This will encompass its own web vocabulary, which will be aligned with the definitions developed by the GS1 community over many years.

We will also be providing tools such as JSON-LD examples and guideline notes that will make it much easier for companies to adopt this new approach. GS1 UK is encouraging and supporting a number of pilot activities in this area for experimentation to take place with strong guidance and assistance.

As the adoption of this new technology and approach to web data is a significant step for retailers and brand owners, we encourage you to experiment and provide us with feedback, as we take the view that a collaborative approach will be most effective in enabling all parties to gain maximum value from this major strategic imperative.

## Appendix - FAQs on the semantic web

### What is the semantic web?

According to W3C, the semantic web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF). It refers to a collection of technologies that can be used to transform the web of documents (e.g. web pages) into a global web of interlinked interoperable data that is machine-interpretable because the meaning of each data relationship is explicitly stated – and because the semantic web uses HTTP URIs (e.g. web addresses), it is possible to access related data as well as definitions of properties and attributes, multilingual names and descriptions simply via a regular HTTP web request.

### What is linked data?

According to LinkedData.org, linked data is about using the web to connect related data that was not previously linked, or using the web to lower the barriers to linking data currently linked using other methods. Linked data is sometimes considered as being either synonymous with the semantic web or being a subset of it. Linked data can be provided and retrieved via web requests, either as standalone data – or embedded within regular webpages, as additional semantic mark-up of the facts contained within the page, which are accessible without ambiguities to software including search engines, smartphone apps etc.

Linked data can be embedded within a webpage using either inline mark-up techniques (such as Microdata or RDFa) or using a single block of structured data (e.g. using JSON-LD (JavaScript Object Notation for Linking Data) – more info at <http://json-ld.org/>). RDFa and JSON-LD are W3C technical recommendations (standards) that enable structured factual data to be encoded within a webpage in a way that computer software understands how to extract it automatically. However, Microdata, RDFa and JSON-LD are orthogonal to vocabularies such as schema.org, GoodRelations and the new detailed web vocabulary that GS1 is developing for describing products and services in rich detail. So we need to choose various classes and properties from web vocabularies (also called ontologies) and use these with a specific mark-up format (such as JSON-LD or RDFa) for encoding those facts and figures into a webpage.

### What do we mean by a ‘trusted source of data’?

Trusted source of data refers to techniques that provide an assurance that the data was provided by an organisation that had the authority to provide that data, such as the brand owner or manufacturer of a product. This is in contrast with data from untrusted or non-authoritative sources, such as crowd-sourced data about products. Because most brand owners who apply barcodes to products lease a GS1 Company Prefix from a national GS1 member organisation (such as GS1 UK), GS1 is in a unique position to know which brand owner is associated with a given product barcode (GTIN – Global Trade Item Number) – and to be able to confirm whether data about a product came from a trusted source, typically the brand owner.

---

## What role does standardisation have to play in the semantic web?

Standardisation plays a critical role in promoting interoperability and reducing ambiguities and incompatibilities in the web of data. W3C has overseen the development of many of the fundamental technical standards that provide the framework for exchanging structured data using semantic technologies. Most of these standards are supported by commercial and free or open source implementations. For web vocabularies, there are some standardised ontologies such as Dublin Core [<http://dublincore.org/> – see also IETF RFC 5013, <http://tools.ietf.org/html/rfc5013>, ISO 15836:2009], as well as web vocabularies (such as schema.org) that were initially developed outside of a standards process, but which are now being further developed and extended within a collaborative community process, with involvement from W3C [<http://www.w3.org/wiki/WebSchemas>]. For over 40 years, GS1 has brought together a community of brand owners, manufacturers, distributors and retailers in a number of industry sectors to work together on common standards for exchanging information within supply chains. The GS1 community has already developed extensive detailed data models and data dictionaries for describing products, services and organisations and the GS1 Digital Initiative and GTIN+ on the Web work group is now making these available as a web vocabulary for use with linked data technologies.

## What are the challenges going to be in terms of getting retailers and brands on board?

At present (2014), many established companies are only just beginning to understand how the web is evolving and the benefits of embracing semantic/linked data technologies. For many, the initial incentive may be the opportunity to achieve enhanced search results, such as Google Rich Snippets, whereas a few are considering the longer term benefits and the new kinds of opportunities and product-related services that could be enabled in the next generation of smartphone apps if rich structured data about products is readily available on the web. As with any new technology, there can be a ‘first mover advantage’ but some companies are still unclear how to proceed with new unfamiliar technology – or are concerned about whether they somehow relinquish control if they publish such data using linked data technology. The fact is that initially, much of the structured machine-processable data will not be data that is commercially sensitive (such as traceability data) but information that is already effectively in the public domain because it appears on the packaging or products or in human-readable format in public webpages. For its part, GS1 is trying to educate its users about linked data

technology, the potential benefits – and provide not only a GS1 web vocabulary that is aligned with the precise definitions of terms developed by the GS1 community through a consensus process spanning several decades – but also to provide tools such as JSON-LD templates that should make it much easier for companies to adopt linked data technology. GS1 is also encouraging and supporting a number of pilot activities in this area.

## Why is it attractive to agencies?

For advertising agencies, linked data technology helps to fine-tune web search results and helps consumers to find exactly the products and services that are of interest to them, because the detailed product characteristics become more readily available to search engines and smartphone apps. Advertising agencies that become deeply familiar with semantic/linked data technologies will be in the strongest position to provide the greatest value to clients as the web evolves and in an omnichannel world.

## What are the benefits to retailers?

Retailers can benefit from linked data technology by ensuring that their product offerings are highly visible on the web (including details of promotions and special offers, availability, cross-selling opportunities for related products and accessories). Linked data can be used to provide an enhanced customer experience on the web and via retailer smartphone apps or in-store scan-and-collect devices. Some retailers (e.g. Norgesgruppen in Norway) have already developed highly sophisticated digital shopping list apps that enable consumers to make intelligent choices about products, based on their individual dietary requirements, or preferences on health, environmental or ethical issues.

## What are the benefits to brands?

Brand owners can benefit from linked data by ensuring that their products are highly visible on the web (including detailed product specifications, ingredients, nutritional information, health, environmental and ethical accreditations, as well as links to technical datasheets, instruction manuals, online help etc). Using linked data technology, even small manufacturers of specialised niche products can achieve just as much web visibility of their products as the larger manufacturers of mass-market products.

## What are the benefits to consumers?

By providing accurate structured data to consumers and retailers using linked data technology, the information needed by consumers becomes quickly available in ‘frustration-free’ packaging, enabling convenient side-by-side comparison of products not



---

only on price or value but a number of other factors, without consumers having to spend so many hours searching and researching online for their major purchasing decisions.

### **How will Google listings be impacted by structured data? (give example of 'red shoes' vs 'Lego Luke Skywalker'?)**

Using the schema.org, GoodRelations or GS1 vocabularies, it is possible to associate rich structured data about products with the globally unique identity of the product (its GTIN – typically the number on the barcode) by using properties such as <http://schema.org/gtin13>

The use of the GTIN as a consistent cross-reference across information provided by the brand owner and multiple retailers and resellers of the product enables search engines to quickly determine which data about a product is consistent – and which information is likely to be misleading.

Furthermore, GS1 is working on a mapping of its product classification systems (such as Global Product Classification) to a linked data representation, which means that consumers will be able to more reliably find products that match particular categories and criteria/attributes even when the consumer searches by keyword and does not have a specific brand in mind.

### **Will the semantic web become integral to SEO?**

Linked data technologies are already becoming integral to SEO. Many companies are already reporting benefits (e.g. BestBuy example). Even outside of retail and consumer products, organisations and individuals are improving their own SEO by using semantic technology. For example, many musicians and bands are already benefitting from Google Rich Snippets because they have put structured data about their discography into MusicBrainz, a biography into Wikipedia, their upcoming concert listings into BandsInTown or SongKick and they have cross-referenced between these sites and also linked to their websites and channels on various social media networks (e.g. Facebook, YouTube, Twitter), which means that search engines identify the connections across these constellations of linked data and begin to recognise them as 'Named Entities' with interesting facts and figures – and the bands or musicians benefit from enhanced search results such as Google Rich Snippets. At present (October 2014), Google are still generating Rich Snippets as a free bonus for providing structured data online. Would you rather have a free enhanced 'feature article' for the search result about your product or retail store – or would you really prefer a plain text listing with a few lines and a blue hyperlink?

## Contact us

### GS1 UK

Staple Court,  
11 Staple Inn Buildings  
London WC1V 7QH  
**T** +44 (0)20 7092 3500  
**F** +44 (0)20 7681 2290  
**E** [info@gs1uk.org](mailto:info@gs1uk.org)

Member Support Team 0808 178 8799

[www.gs1uk.org.uk](http://www.gs1uk.org.uk)

