Linked Data Standards and the Semantic Web

How are these relevant to users of GS1 standards?

Mark Harrison and Alexander Ilic

Auto-ID Labs White Paper WP-SWNET-030

Mark Harrison
Director, Auto-ID Lab (Cambridge)
University of Cambridge

Alexander Ilic
Director, Auto-ID Lab (Switzerland)
ETH Zürich / University of St. Gallen

Contact:

Auto-ID Lab (Cambridge),
Institute for Manufacturing,
17 Charles Babbage Road, Cambridge, UK,
CB3 0FS

Phone: +44 1223 338178
Email: mark.harrison@cantab.net
http://www.autoidlabs.org.uk

Contact:

Auto-ID Lab (Switzerland)
ITEM HSG
Dufourstrasse 40a
CH-9000 St. Gallen

Phone: +41 71 224 7240
Email: alexander.ilic@autoidlabs.ch
http://www.autoidlabs.ch

WNET-030

Executive Summary

The proliferation of the World Wide Web has driven a step-change in the global availability of data, and recent work in the field of information encoding standards for the Web is beginning to drive new levels of consistency into the way that data is represented on the Internet.

Concurrently, the dissolving boundaries between online and offline retailing are posing new challenges for supply chains. Consumers use their smart phones as a medium to connect both worlds and are ready to engage with brands and retailers in completely new ways. New opportunities are being created by the global rise in social media engagement for individuals, businesses, campaign groups and governments to source (and to make accessible) information about products and services. In addition, massive sets of data are being published in open-source Linked Data formats and these data sets are increasingly being leveraged to refine and augment decision-making processes across supply chains in conjunction with real-time analytics, reality mining and metrics of consumer behaviour.

Mobile and online consumers are increasingly relying on rich online information and digital product profiles to guide their buying decisions and to identify products and services that best fit with their lifestyle and with their personal needs. For them, data quality and data accessibility are critical measures of trust in any online source of data. For businesses, assurance that all representations of their products and services online are accurate and representative is paramount. Discoverability and search visibility are increasingly key contributors to market share for most manufacturers and sellers in today's online world.

This step-change in global data availability is disruptive, and has already driven fundamental changes in the way that consumers approach their buying decisions. The companies behind three major search engines have joined forces to promote the schema.org 'vocabulary' that provides a data model for use with embedded structured data so that factual information can be more easily found and extracted. Leveraging this, it is conceivable that web search engines and mobile applications (apps) will provide considerably richer contextual search interfaces in the near future. Search technology will be more personalized and more adapted to personal needs, preferences, dietary concerns, allergies, priorities, interests, sensitivities and even intentions and travel plans.

The current state of development and adoption of standardized web vocabularies has major relevance to the GS1 System. GS1, working across multiple industry sectors for 40 years, has defined a detailed Global Product Classification (GPC) system, which can be used to identify products that match particular search criteria. Leveraging the Global Trade Item Number (GTIN) as a globally unique key, systems exist for accessing even more detailed and quantitative 'Master Data' about a manufacturer's products across a global supply chain (GDSN). Combining existing GS1 Standards with online tools that allow manufacturers and retailers to conveniently export a customer-facing subset of their 'Master Data' in formats that are aligned with developing Linked Data standards would allow companies of any size to take maximum advantage of the availability of accurate online data about their products and

services. Such solutions could have significant implications for ensuring availability of trusted data in a global marketplace (especially considering the existence of recent legislation such as the European regulation EU1169/2011 regarding food labelling requirements that require detailed and accurate product information to be available online).

In this paper, we explore some of the challenges and opportunities posed by the widespread availability of Linked Data technologies in a Business to Consumer (B2C) context as well as in a Supply Chain Management context. In particular, we explore the value of standardized product IDs on the web and Linked Data technologies that connect physical goods with digital product profiles and mobile/online consumers.

A fairly substantial technical appendix provides background information and introduction to the underlying technical standards, in addition to the methods that have been developed for data structuring and querying. Analysis of knowledge-based and rule-based automated reasoning is also covered in the context of generating accurate inferential data.

Disclaimer: This paper focuses on providing an introduction to the applications and technical standards rather than a comprehensive research overview.

1. Introduction

Governments, government agencies and companies are increasingly making large amounts of useful and interesting data available on the web. For governments, this is done for a number of reasons including open government and greater transparency and accountability, as well as providing data that has been collected or generated at taxpayer's expense (e.g. aggregated census data, mapping data, public transport data) back to the public, to help to boost technical and economic innovation of new digital services and applications (often on the web or mobile devices) that make use of that data and create mash-ups and visualizations in new interesting ways that are helpful to the public. Please see http://www.data.gov, http://data.gov.uk , http://publicdata.eu and http://datahub.io for links to open datasets as well as some examples of data mash-up applications and visualizations. For companies, the motivation may be to ensure that detailed information about products, services (and the companies that offer these) can be efficiently indexed by search engines and found by customers, especially as search technologies are becoming increasingly sophisticated, personalized and focusing on the semantics of information on the web.

GS1 has a unique strength of bringing together a global community of manufacturers, distributors and retailers in multiple industry sectors, to work together on the development of open standards that enable supply chains to operate more efficiently through interoperable data models and interfaces for data exchange. Open standards in turn foster a competitive marketplace of technology solutions and avoid a lock-in to a single solution provider, operating system or programming language. To date, much of the focus has been on business-to-business information exchange but there are already some GS1 initiatives concerned with mobile commerce and business-to-consumer applications.

Business-to-business, mobile commerce and business-to-consumer applications each have their own characteristics. For B2B, a major concern is interoperability, including robust validation and checking of data formats and data types, in order to minimize errors and ambiguities. For this reason, technologies such as XML and Web Services provide this kind of robust validation, while remaining neutral, with respect to operating system or programming language.

For mobile devices, the available bandwidth and data transfer rates can be a limiting factor, so many websites designed for mobile devices use more lightweight mechanisms for formatting and exchanging structured data. These include JavaScript Object Notation [JSON] as a lightweight alternative to XML, [REST] interfaces and Asynchronous Javascript And XML [AJAX] requests as an alternative to traditional XML SOAP-based Web Services or re-loading of entire web pages, when it is only necessary to exchange a small package of structured data, e.g. for the automatic suggestions in a search field as the user types.

For business-to-consumer applications, a key characteristic is the need to support collation and comparison of products, services and their associated information from multiple sources, even though these might be provided by several different organizations and presented in various layouts. This therefore requires the ability for computer software and web search engines to efficiently extract the meaning of information from different websites and data sources, in order to support meaningful comparisons that enable consumer choice.

The purpose of this paper is to provide a concise overview of the key principles and technologies of Linked Data and the Semantic Web and to discuss how these can be relevant to GS1 and supply chain users of GS1 standards. We consider that this is particularly timely, since web search engines are encouraging the use of semantic markup within websites through initiatives such as schema.org [http://schema.org] and the GoodRelations ontology [http://purl.org/goodrelations/], while GS1 is developing standards such as Trusted Source of Data [http://www.gs1.org/source/standards] and solutions such as GS1 Source [http://www.gs1.org/source] and GS1 UK TrueSource, to ensure that consumers are provided with correct trustworthy master data about products, particularly when they perform web searches about products or use applications on mobile devices and smartphones to scan products and retrieve additional information.

Through Global Data Synchronization Network technology [GDSN], such master data is already exchanged between businesses but in this paper we explore the opportunities for making a subset of that data available for business-to-consumer use, aligning with linked data techniques. This also opens up some new possibilities for exploiting Trusted Source of Data initiatives in the *opposite* direction, 'Consumer-to-Business' or the product/provider *discovery* phase, where instead of scanning a physical product to retrieve information, we start with a search keyword and a number of consumer preferences (e.g. category of product, organic, fair trade, nut-free, gluten-free) and constraints (e.g. budget, size, urgency, proximity etc.) in order to discover products and services that meet their needs, as well as local or online providers and their location, availability and pricing information. Figure 1 shows both the initial 'Search and Discover' phase ('C2B') as well as the 'Identify, Inform & Interact' phase more familiar in current B2C applications [GS1 B2C]. Further applications can be envisaged beyond the point of purchase, using hyperlinked digital receipts to link to a number of product-centric services, such as retrieval of instruction manuals, automated warranty registration, discovery of providers of spare parts, refills and accessories, among others.

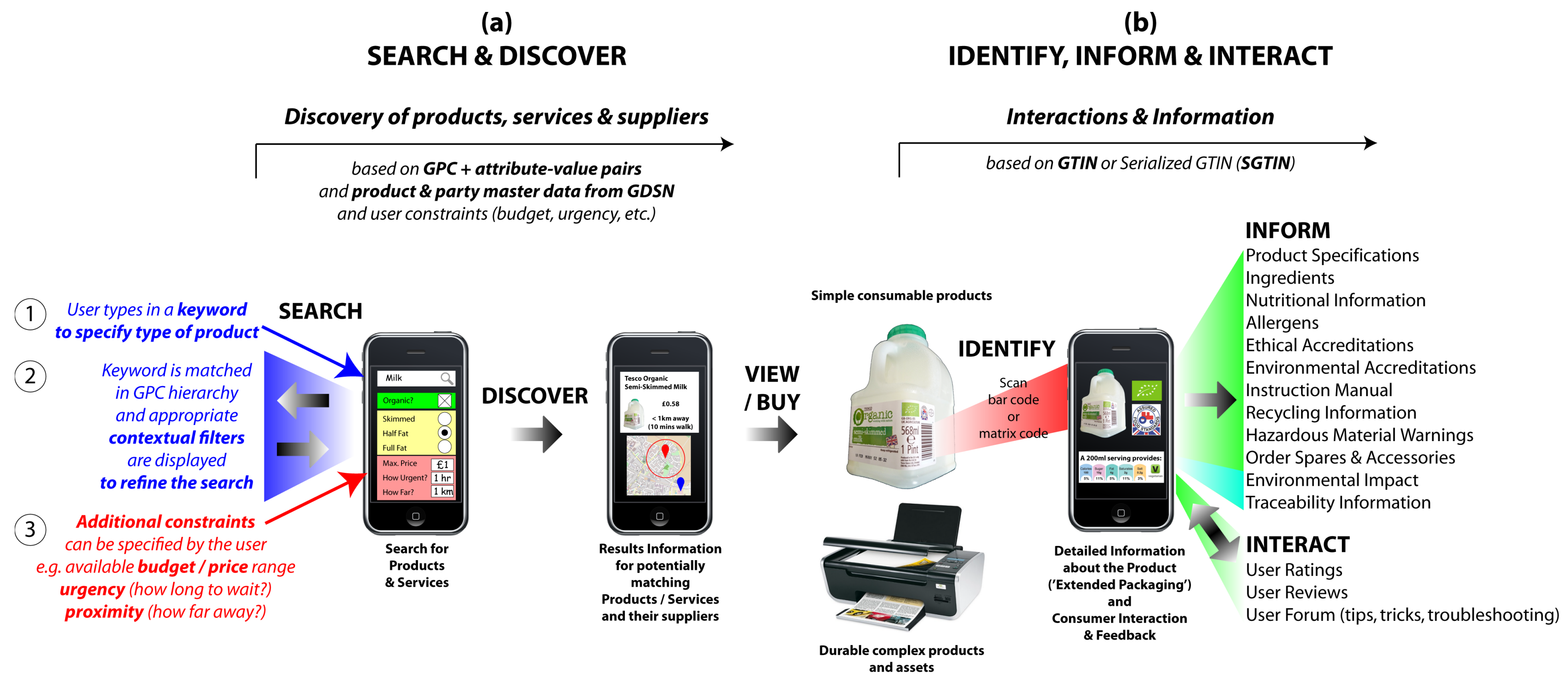


Fig. 1. Consumer-to-Business (C2B) and Business-to-Consumer (B2C) interactions

2. Business perspective

The boundaries between online and offline retailing are continuously dissolving. With this paradigm shift, new challenges and opportunities for retailers and brands emerge from the consumer side. A standardized representation for product IDs and related linked data on the web will help to engage with the new generation of mobile and connected consumers. Also, it will enable holistic and real-time analytics that can be transformed into actionable insights to drive revenue, customer loyalty, and profitability.

2.1. Key Challenge Areas

This paper looks at how data related to physical goods and digital product profiles on the Web can be linked and associated with consumers. In this context, smartphones have become the key media to connect the physical with the digital world in a personal and social context.

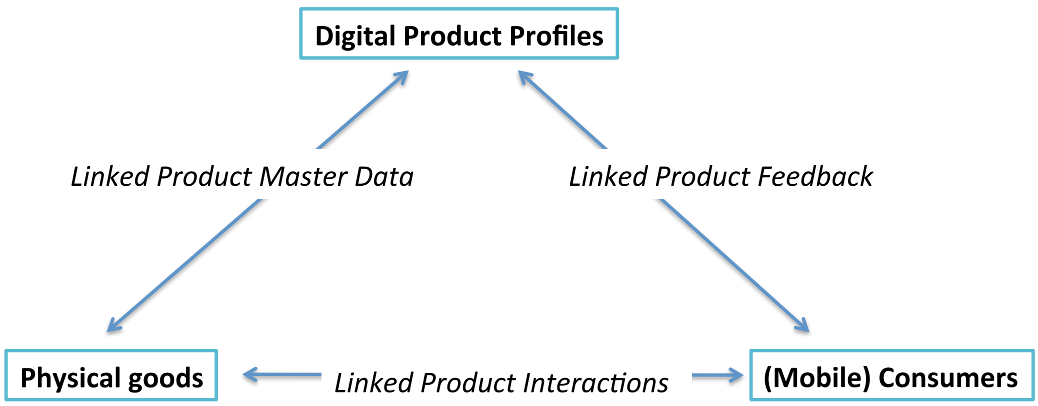


Figure 2 – Three key areas for standardization of Linked Data about products

Retailers and brands enhance their data strategies beyond a B2B case towards a B2B2C scenario. As shown in Figure 2 above, there are three key areas for standardizing product IDs and related linked data that need to be addressed:

- Linked Product Master Data: Providing a digital profile and linking it to physical goods where the linked data comprises aspects such as the relation of different identifiers for describing the same or similar products, product categories and alternatives, attributes, text and other rich descriptive information.
- Linked Product Feedback: Linking VoC (Voice-of-customer), endorsements, ratings, reviews, signalling of brand affiliation ("likes") to digital product profiles
- Linked Product Interactions: Linking interactions of physical products with people (e.g. comparing, purchasing, selling, repairing). Interactions are typically enabled by

scanning bar codes, RFID tags, object recognition or digital receipts (online and offline) that leverage the Global Trade Item Number (GTIN) or Serialized GTIN (SGTIN) or other identifiers to map the product interaction into the digital world.

The following technical analysis discusses the three challenge areas in more detail.

2.2. Implications of a GS1-enabled ID on the Web

GS1 has a unique strength of bringing together a global community of manufacturers, distributors and retailers in multiple industry sectors, to work together on the development of open standards that enable supply chains to operate more efficiently through interoperable data models and interfaces for data exchange. With a standardized representation for ID on the Web, the GS1-enabled services and language of business can be extended to the B2C world.

This will lead to significant boosts for companies supporting a GS1-enabled ID on the Web. For example, retailers and brands might benefit from the following aspects:

- Future proofing against dissolving boundaries between online and offline shopping
- Faster communication of value proposition for complex products (e.g. by using observational learning, social cues, popularity information for cross-selling, reviews, etc.)
- 360 degree actionable analytics and visibility of customer needs beyond POS
- Increased consumer engagement and loyalty
- Enabling of product-based loyalty programs / reward schemes

Last, but not least consumers will benefit from this standardisation by being able to enter into a constructive dialogue around products with retailers and brands yielding several benefits including:

- More efficient ways to discover/explore products and faster decision-making
- Ability to track own consumption footprint and behaviour (kcal, CO_2, $, ...) across channels and brands
- Ability to give efficient and direct product feedback to retailers and brands
- Ability to signal status to peers and product communities for self-help and networking
- Faster/better access to product-related services

3. Technical Perspective

A standardized ID on the web based on a semantic markup framework is required to make the web more efficient to understand for machines and serves as the basis for the Linked Data opportunities outlined before.

The World Wide Web (WWW) is a global collection of documents (web pages) with hyperlinks to link one document to a related document. The WWW is massively distributed in nature and anyone with access to a computer and an Internet connection can easily access or create web pages of information or views about any subject. In recent years we have seen the emergence of websites such as Wikipedia, a crowd-sourced encyclopedia containing information on all kinds of subject, contributed by individuals. Web search engines (such as Google, Yahoo, Bing and others) periodically crawl the web, following links to discover new web pages and also to rank these according to the keywords or text that they contain and the relevance, judged by various factors including the presence and contents of metadata, number of hyperlinks from the page to related pages - and from other pages to that page, as well as other factors such as information about web traffic flows and user interactions such as the search keywords and frequency with which particular search results are clicked or followed in response to particular search criteria.

Web pages were initially designed to be read and understood by human beings. The visual layout of a page is very flexible and creators of web pages may choose different page layouts as well as different ways of presenting information, e.g. in graphical charts, in tables, as well as within sentences of text. Human beings who read a web page can usually extract the information they require and understand the context and semantic meaning of the content, regardless of how it is presented. A human being can easily extract meaning from a chart or a table of nutritional information that is embedded as a bitmap image - but to most computer systems, the meaning and underlying data that were used to generate the chart or table remain completely opaque and inaccessible. Most computer systems and search engines do not yet have this flexible ability to extract meaning from words and images as a standard feature and therefore need assistance (in the form of embedded semantic markup) in order to 'understand' the meaning of the information that is presented in a web page or document and in order to make meaningful comparisons and correlations across websites that present similar information in different formats and layouts.

Search engines are now placing much greater emphasis on understanding the semantics of information in web pages and are even rewarding those who add semantic markup with the prospect of more prominent, visually appealing search results, such as Google Rich Snippets.

For example, a web page about a food product might contain information about its ingredients, as well as nutritional information (e.g. calorific value and amounts of protein, fat, sugar, fibre per 100g or 100ml), as well as advice about potential allergens within the product (e.g. 'may contain traces of nuts') and various ethical and environmental accreditations and certifications (e.g. Fair Trade, organic / bio, sustainable farming/fishing, approval by the Marine Stewardship Council, Forest Stewardship Council, Soil Association etc.). A human being can easily recognize such information (so long as they can understand the human language in which it is presented) and interpret it, regardless of where or how it appears on a web page. A computer generally needs semantic markup or metadata within the page in order to help it to find and extract the same kind of information from each page, regardless of the layout of the page or whether the information is provided as text or within a bitmap image such as a logo, chart or photo. In Figure 3, we show an example of a web page that has been enhanced with semantic markup for the product characteristics. The visualisation shows the various facts about the product (e.g. nutritional information) that can be automatically extracted by search engines and other software when the appropriate semantic markup is added to the web page templates.

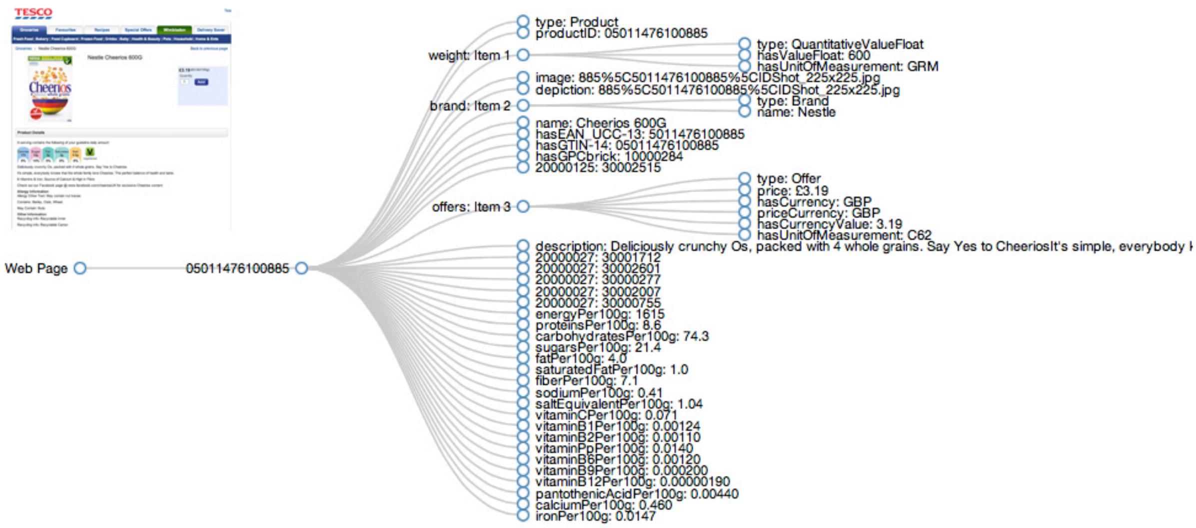


Figure 3 – example of a graph of structured data that can be extracted from a web page about a product when appropriate semantic annotations are added to the web page template.

Semantic markup therefore enables computer software and search engines to compile much richer information about web pages - not only the keywords and phrases that appear within the page, but also much richer information about the kind of product or service that is described by the page, thus enabling this richer information to be extracted and indexed within a search engine database that is populated from crawling several pages. In turn, this is fundamental to enabling meaningful comparisons of products and services across different websites, as well as much more sophisticated search techniques including contextual search and passive search agents.

In order to enable these scenarios and extend the GS1 world to the consumer, computers and search engines need greater ability to extract the meaning of information from web pages and to use this to intelligently match against our various multiple requirements, preferences and priorities. Linked data and semantic web technologies enable this vision and are now reaching sufficient maturity in terms of standardization at W3C and recognition by major web search engines that it makes sense to consider exploiting this technology in order to improve the visibility of products and services to potential customers and end-consumers.

We consider that GS1 is in a particularly strong position to help its user community (many of which are small to medium enterprises (SMEs) to take maximum advantage of this, by providing them with tools that allow them to make master data about their products and services available in Linked Data format so that this much richer information can be more readily indexed by search engines and can be more easily discoverable by consumers.

A further benefit to all members of the GS1 community of greater alignment and familiarity with linked data technologies is the prospect of easier exploitation of various public open data sources by businesses to improve their own business intelligence. This can include open data sources such as weather forecasts, alerts about traffic congestion, planned roadworks or other incidents that can disrupt or delay normal logistics processes within supply chains. Linked data technologies such as the SPARQL query language [http://www.w3.org/TR/rdf-sparql-query/] make it relatively easy to combine internal company data with public open data in order to enhance the value of the internal data or enable additional or better decisions to be made using it.

As a very simple example, a company will typically have a list of suppliers and regular customers, together with contact details. These might include a postal address and postcode. In the United Kingdom, the postcode is very granular, typically mapping to an individual street, a part of a very long street or sometimes an individual building; the combination of postcode and house number is usually sufficient to uniquely identify any residential address in the UK. Recently, the mapping between postcodes and other attributes (such as latitude and longitude co-ordinates) has been made available as public open data by the Ordnance Survey as part of the UK government open data initiative [data.gov.uk]. This means that any company can now easily use linked data tools, such as the SPARQL query language (discussed in section f of the appendix) to enrich their internal contacts database to include geographic co-ordinates (e.g. terrestrial latitude & longitude co-ordinates) as well as postcodes. Using the geographic co-ordinates, they can calculate distances to particular suppliers or customers, sort a list of suppliers to find the nearest suppliers - or consider the geographic distribution of their major customers when planning the location of a new retail store or warehouse. Furthermore, socio-economic information such as aggregate census data, demographics, house purchase prices or crime statistics are increasingly being published by government agencies as open data and linked to postcodes, electoral wards or constituencies, so these factors can also be taken into account when planning new locations for business or the types of products to stock in a new store. Please see http://www.data.gov, http://data.gov.uk , http://publicdata.eu and http://datahub.io for links to open datasets as well as some examples of contributed data mash-up applications and visualizations.

3.1. Linked Data for Product Master Data

Linking digital profiles with product master data of physical products is important for helping computers to extract meaning and enable more efficient searches by consumers.

Many of us are familiar with contextual search within online shopping sites. For example, the website of a clothing retailer may allow us to progressively refine our search for a product by applying a number of successive filters, e.g. selecting menswear vs ladieswear, trousers / shirts / pullovers / coats and jackets, size, colour, fabric, price range until we find a subset of product classes that match all of our selected criteria.

Now imagine that this kind of contextual search becomes available as a standard feature in all major web search engines, so that when someone types a keyword such as 'jacket', the search page understands the context and knows that it should display appropriate additional filters about size, colour, fabric, price range etc. so that it is possible to find not only the products or services from the website of each individual retailer or manufacturer but to do this kind of *contextual, criteria-driven search* using a general-purpose web search engine to find products or services from any manufacturer or retailer potentially anywhere within the world - *or* within the immediate local area.

Auto-ID Labs researchers have already developed an initial prototype of contextual search using the Global Product Classification (GPC) system and its attributes and values in order to refine a search. In the prototype demonstration system, shown in Figure 4, a user types a keyword and is then provided with a filtered list of potentially relevant product categories. After selecting a category from the list, the relevant attributes and values for that product categories are displayed. Tri-state checkboxes are provided, to allow the user to indicate whether they wish to include or exclude a particular attribute and value combination or whether they have no preference regarding that attribute. Behind the scenes, the contextual search interface is collecting the following information: the product category (e.g. the GPC brick value), an inclusion list of attribute-value pairs and an exclusion list of attribute-value pairs.

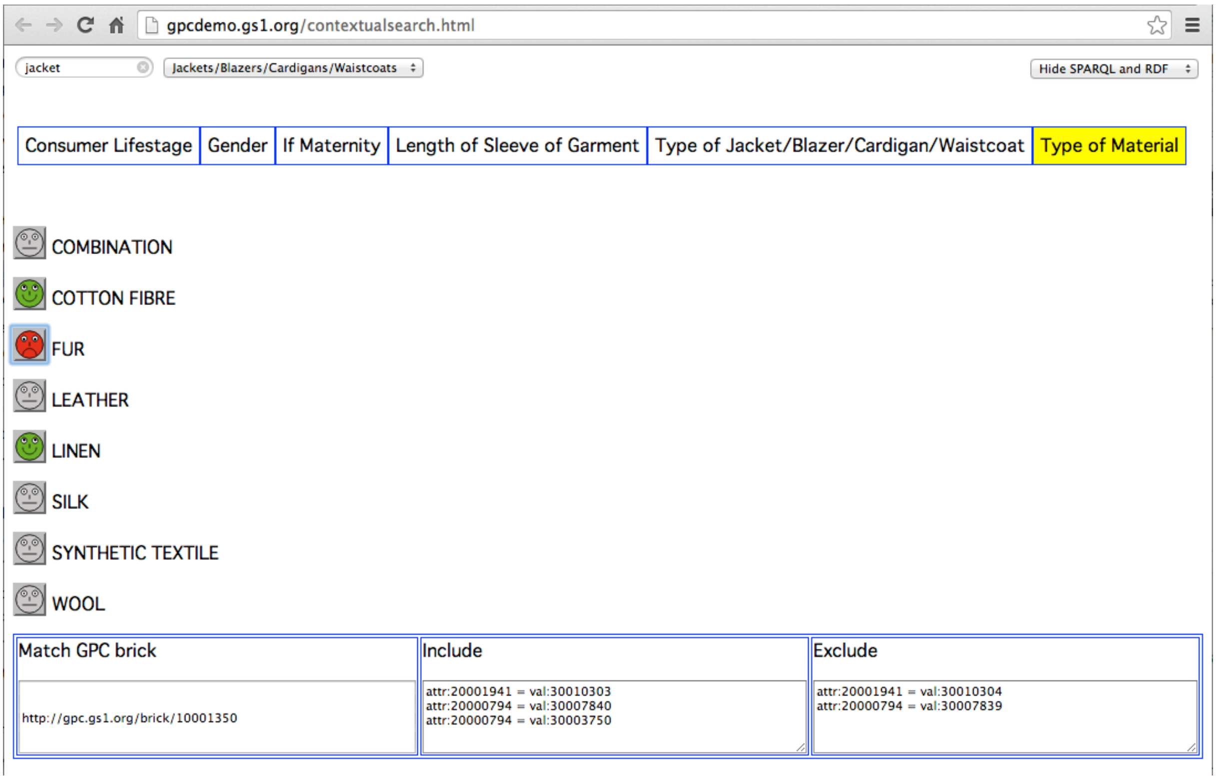


Figure 4 - A screenshot for an initial prototype for contextual search based on GPC.

The next logical step for extending the demo is to connect it to actual product master data in order to find GTIN identifiers for products that fall into that particular product category (have that GPC brick) and which have attribute-values pairs that match the inclusion list criteria and which do not have attribute-value pairs that match the exclusion list criteria.

Currently, when we are planning a major purchase (e.g. a new car, house, digital camera, computer or home entertainment system), we typically spend some considerable time performing online research across a variety of websites in order to collect all the information we need to make a balanced decision. This might include retrieval of technical datasheets and brochures from a manufacturer's website, reading independent impartial reviews, checking price comparison sites to find the best deals, both locally at traditional 'high street' retailers or online, then checking current availability and if buying locally, obtaining a map and directions to the retail store. If planning a house purchase, we might want to check the local crime statistics, road accident statistics, public transport connections, risk of flooding, reputation of nearby schools etc. Since this kind of information is being released as open data, often using linked data formats, it becomes much easier to get an overview about locations. Web developers are also developing map-based data mash-ups that overlay such socio-economic profile data over a streetmap, using contour maps with colour gradients to provide an easily digestible view of the data for each location. One example of such a mash-up can be found at http://apps.seme4.com/see-uk .

One of the visions for the GS1 Digital GTIN+ on the Web activity is to provide a foundational platform for product-centric information and services in a somewhat analogous way to the use of mapping tools provided by search engines (e.g. the Google Maps API) provides a platform for mash-ups based on location-centric information and services.

Expected research results by the Auto-ID Labs:

- Digital out-of-stock / lost sales: We want to demonstrate how to reduce digital out-of-stocks and lost sales by increasing accuracy of online/mobile search with a standardized ID and structured data for products on the web

- Trusted data sources: We aim to show how to manage consistency, provenance and trust for digital product profiles based on open, distributed data from various sources

- Mapping of product identifiers, attributes and categories: We want to enable a seamless interoperability between different systems by showing how products, their identifiers and attributes can be structured in a consistent way

- Consistency and timeliness: We aim to facilitate analysing linked data of digital product profiles by providing insights about how to deal with outdated or conflicting information

- Enable attribute-based search/ discovery of product alternatives: We aim to demonstrate how linked data can enable consumers to easily discover product alternatives based on various criteria, personal preferences and context

3.2. Linked Product Feedback

Linking digital profiles with feedback information from mobile/online consumers is important to simplify search, enable faster decision-making and enable network effects. Mobile and online consumers typically operate with their very specific context in mind.

Thanks to standardized linked data, we expect that in the future, we will spend individually less time actively searching the web for such information we need but will instead enter our intentions/desires and constraints (e.g. budget, rating threshold, urgency etc.) into the interface of a search agent and use it to intelligently and periodically search for the options of products, services and other offerings that meet our search criteria, together with ratings, reviews and recommendations, so that instead of users having to spend time searching across multiple sites for information, the search agents are able to retrieve this, collate it and present us with an intelligently sorted list of options and associated 'information packages' for us to consider. Of course this is not only restricted to purchases of material goods. We might configure such search agents with details of our musical preferences and other interests, so that it can alert us to concerts, performances, events and exhibitions that might be of interest to us, either in our local area or at locations that we plan to visit in upcoming business or leisure trips. Furthermore, this kind of contextual searching and matching can be done anonymously by a search agent acting on behalf of a person, without needing to reveal our identity to the providers of such goods and services; more precise information about such offerings can be discoverable by interested potential consumers based on their search/preference profiles rather than their identity or contact details and without subjecting them to unsolicited marketing / 'spam' - it is a form of passive targeted advertising that is entirely driven by the intentions, preferences and constraints that a citizen has voluntarily specified in a passive search service to which they have opted in. A user might interact with such a search agent through a web interface or through an 'app' (software application) on a mobile device or smartphone, although some of the background computational processing might be done on the network / 'in the cloud' and periodically synchronized with the mobile device. 'Google Now' [http://www.google.com/landing/now/] and 'Siri' from Apple [http://www.apple.com/ios/siri/] are some familiar early examples of personalized search agents that use semantic technology and natural language processing to understand context and aim to provide helpful personalized information and recommendations.

Expected research results by the Auto-ID Labs:

- Mobile/online search in personal context: We aim to prove how linked data can enable targeted and criteria-based search while still retaining privacy

- Alerts and notifications: We aim to demonstrate how consumers can be engaged in passive search services and receive notifications about product profile updates

- Contribution of feedback, rating and reviews: We aim to demonstrate how all reviews and ratings related to a product can be easily consolidated regardless of the channel by using a standardized linked data feedback representation for digital product profiles

- Signalling driven by consumers: We aim to show how consumers can be enabled to signal their relation and attitude towards digital product profiles in search and towards brand owners

- Product communities: We aim to show how owners of the same products can share their experiences in a privacy-preserving manner

3.3. Linked Product Interactions

Linking the interactions of mobile/online consumers with physical products on the web is a key element to bridge the gap between the physical and digital world. Product interactions can range from touching/looking at a product to buying it, selling it, etc. The following paragraphs illustrate several technologies for supporting easy creation of linked data in this context.

So far, most of the use cases focused on pre-purchase interaction scenarios of price comparison of physical products. This is usually done by an app on the consumer's smartphone that extracts the GTIN / SGTIN from a bar code or RFID tag.

In addition to the pre-purchase case, consumers are increasingly interested in post-purchase data for quantifying their own shopping behaviour and getting better services for re-selling their products, assessing the value of their products, sharing them with friends, and for easier warranty management. Therefore, it is not surprising that the popularity of receipt management applications that scan paper receipts for documenting proof-of-purchase associations between consumers and physical products is rising significantly. Examples of such functionality can be found in applications such as Delicious Library (for keeping track of the CDs, DVDs and books that you own or might lend to friends and family) or Immobilize.com, the UK National Property Register, which allows anyone to register the valuables they own (including registration of serial numbers and uploading of scanned receipts) so that police can reunite stolen goods with their rightful owners and so that second-hand dealers can check whether someone actually owns the goods they present.

Brands and retailers (e.g. Apple, Best Buy, Macy's) have started to provide digital receipts instead of paper-based ones to their consumers. While these initiatives are a good starting point and good for the environment, it is still not a fully seamless experience for consumers. Much more could be done with the data if digital receipts would be enriched with standardized hyperlinked representations of product IDs and data encoded in a more suitable machine-readable way, so that data can be directly used in mobile/online apps.

Therefore, a more efficient and promising way might be to use digital receipts that carry itemized and standardized GTIN / SGTIN representations directly from the POS to the user's smartphone. This scenario is applicable to online as well as to physically purchased goods and therefore could be an attractive way for trusted proof-of-purchase associations of their products.

Last but not least, consumers could manually enter their product interactions with manual, reverse lookup functions for products. A typical example could be when a user wishes to sell a product on platforms such as eBay and tries to find the product identifier by using descriptive names and specifying exact make/model or other information to ensure a better mapping and find the exact Global Trade Item Number (GTIN) for the product.

Expected research results by the Auto-ID Labs:

- Proof-of-purchase: We aim to evaluate options for providing reliable proof-of-purchase data as a transition from a pre-purchase scenario to a post-purchase scenario for consumers
- Reverse look-up: We aim to demonstrate how simple descriptive terms or image recognition can be used to do reverse look-ups from the physical product to its corresponding digital identifier
- Consumer analytics: We aim to demonstrate how to enable consumer analytics that encompass online and offline retailing
- Timeliness: We aim to investigate how to ensure timeliness of mapping product interactions to the web and consistency checks
- Privacy of interactions: We aim to give consumers tools and frameworks that enable them to control the degree of what they want to share on their product interactions

4. Opportunities for GS1 and its community

GS1 works with end-users and technology solution providers to develop standards to improve the efficiency of operations across the supply chain. These include standards for classification of product categories, identification of product classes and product instances, shipments, assets, documents etc., standards and solutions for efficiently exchanging and synchronizing master data about product classes and parties (organizations), as well as standards for capturing and sharing data about physical events such as observations and changes in containment.

Until recent years, much of the focus has been on efficient business-to-business (B2B) communications, using standard identifiers, standardized data formats and standardized interfaces to eliminate ambiguities, reduce the need for manual re-entry of data and to enable efficient sharing of data across the supply chain, to better balance supply and demand.

More recently, increased attention has focused on business-to-consumer (B2C) operations and particularly electronic commerce and mobile commerce and the ability to provide consumers with accurate information about products when they scan the bar codes of products (or even QR codes) to obtain additional information about products, as well as trying to ensure that the descriptions of products appearing on the websites of online retailers are trustworthy data. Such initiatives include standards such as GS1 Trusted Source of Data [GS1 TSD] and solutions such as [GS1 Source] and [GS1 UK TrueSource].

However, many B2C approaches to date focus mainly on the information flow in one direction, starting with a product identifier, to find additional information about the product, as illustrated in the right-hand side of Figure 1. It is very important to not overlook the other direction, where a consumer might be looking for a particular product or service with particular attributes but does not yet know if such a product exists or who provides it and whether it is available locally or only online. This is in fact much closer to the way most people use web search engines - users input a number of keywords or phrases, rather than starting with the GTIN code of a particular product. So the question is how can GS1 and its subscriber community help consumers to *find or discover* the products and services they seek, based on *product categories* and searching of *attributes*, rather than assuming that the consumer already knows the GTIN because they just scanned it? By providing this capability, GS1 and its whole community can benefit by closing the loop on searches for products and services, helping consumers to find what they are really looking for, resulting in increased sales for those who can provide such products and services and match the consumer's requirements.

Searching by attribute relies heavily on the availability of accurate master data about product classes and organizations such as local businesses or online businesses that offer these products for sale or rental. The GDSN business message standards [GS1 GDSN] already provide a rich vocabulary for expressing the properties of many products, including details about environmental or ethical accreditation marking or information about allergens that might appear on the product marking. This is very helpful for electronic commerce, because a consumer ordering groceries online will typically not be able to see the full packaging label that appears on a product but may need to know whether it is gluten-free or whether it may contain traces of nuts, so this information needs to be readily available when buying goods online (including mobile commerce). It should be noted that recent food labelling legislation (such as EU 1169 / 2011) is requiring the same level of detailed product information to be provided before purchases for online sales as is visible on the product labelling in-store. It should be possible and relatively easy to develop *tools to export* such master data that is currently used for B2B purposes and make a *relevant subset* of it available to GS1 subscribers (e.g. manufacturers and retailers) in a linked data format that they can readily include within their websites and mobile commerce sites, as well as ensuring that the data attributes are available in recognized linked data formats for harvesting by the major search engines, to enable *contextual attribute-driven search*.

4.1. Schema.org and GoodRelations

In June 2011, Google, Yahoo and Microsoft Bing launched an initiative called schema.org to provide a common set of schema for marking up structured data in web pages. Schema.org defines a number of types (or classes) including Organization, Place, LocalBusiness, Product, Offer, Review, AggregateRating, etc, as well as many data fields within each of these. The full list of types is shown at http://schema.org/docs/full.html

Another initiative, the GoodRelations ontology has been developed for over 10 years by the research team of Prof. Martin Hepp at the Universität der Bundeswehr in Munich. GoodRelations is intended as a web vocabulary for electronic commerce and defines a number of classes including BusinessEntity, ProductOrService, Offering, PriceSpecification and many data fields within these. In November 2012, it was announced that the

GoodRelations ontology is now recognized by schema.org and can be used from within the schema.org vocabulary. This means that the much more expressive fields in GoodRelations can be used together with the more general fields that were already defined by the schema.org initiative.

The full class diagram for the GoodRelations ontology is currently available at http://www.heppnetz.de/ontologies/goodrelations/v1#uml

What should be immediately apparent to any manufacturer or retailer looking at these two initiatives is that the major search engines are actively encouraging markup of structured data within web pages, so that important factual data such as location, address, contact details, opening hours, price of products or services can be harvested more easily and more correctly by the search engines and used to support search strategies that are more contextual, filtered or location-centric, as well as providing key summary information about a business or the products and services it offers, illustrated in enhanced search listings (such as Google Rich Snippets [RichSnippets]) with the key information directly visible in the search results page. Schema.org and GoodRelations are already being used by a number of companies, including BestBuy, Sears, K-Mart, Volkswagen UK.

4.2. Repurposing B2B master data for B2C use

There is considerable overlap between the fields in schema.org / GoodRelations and the attributes in GDSN master data about parties and trade items, so there is great potential for GS1 to develop tools that make it easy for manufacturers or retailers to export a subset of the B2B master data that they already contribute to the Global Data Synchronization Network and to make this available for export as linked data using recognized names of attributes defined in schema.org and/or GoodRelations or for GS1 to develop its own GS1 Linked Data Vocabulary that leverages the existing precise definitions in GDSN data models and the GS1 Global Data Dictionary (GDD) and where appropriate cross-references to related concepts and attributes in schema.org and GoodRelations.

There are a number of options for embedding structured data within web pages. These include W3C standards such as [RDFa], [RDFa Lite] and [JSON-LD], as well as other approaches such as Microdata. There are also a number of tools available that can translate between these different ways of encoding structured data facts, such as http://any23.org. User-friendly tools are available to help anyone to get started with adding structured data to their web pages. Examples of tools for adding structured data to web pages are the GoodRelations Rich Snippet generator tool [http://www.ebusiness-unibw.org/tools/grsnippetgen/] and the Google Structured Data Markup Helper [https://www.google.com/webmasters/markup-helper/].

Development and use of such master data export tools would be particularly helpful for small to medium enterprises (SMEs) that may lack substantial IT capabilities. If the tools can reliably provide them with HTML web page markup with the correct embedded Linked Data annotations, the organization can simply copy and paste that into their website and improve the likelihood that consumers will be able to find the products and services they seek, especially when the consumer is searching by attribute, e.g. for a product with particular characteristics or specifications and is interested in knowing current availability locally or online, as well as pricing information, associated ratings and reviews, etc.. This helps consumers to discover better matches for what they are really looking for and can also help increase sales, as well as giving high street retailers and services a much needed boost in these challenging economic times. Figure 4 provides an illustration of how GS1-related linked data can help with the search and discovery of products or services relevant to a user's search criteria.

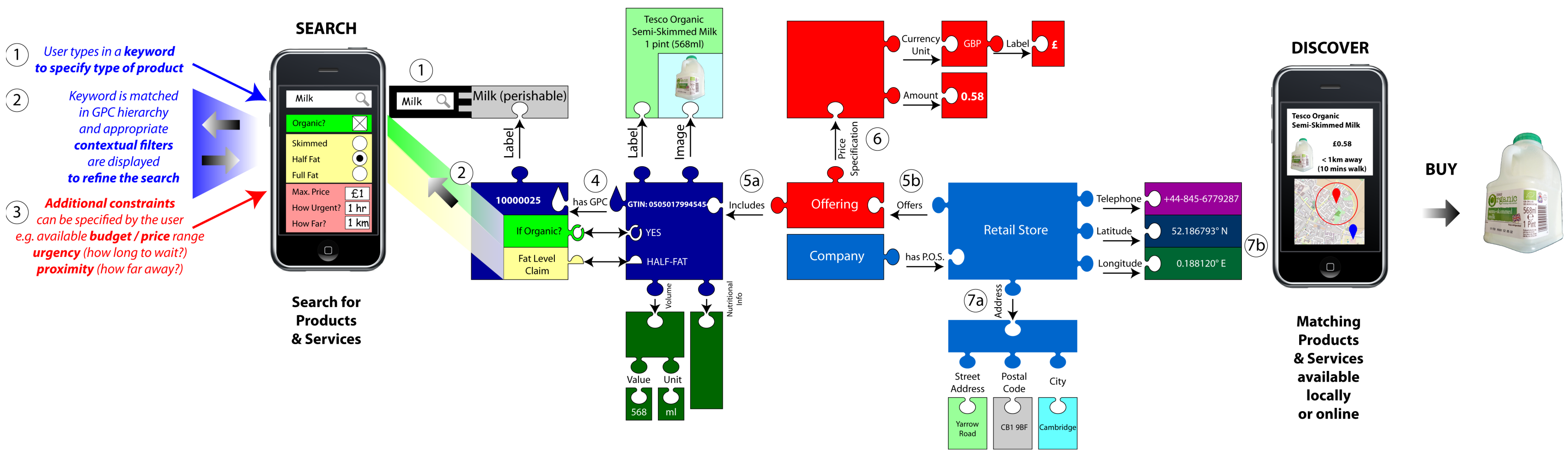


Fig. 4 An end-to-end example of the use of Linked Data relationships to perform contextual search and discovery products and services.

In step 1, a user provides a keyword to indicate the product category

In step 2, the keyword is matched against the labels of GPC identifiers (ideally GPC bricks) and relevant attributes and values are shown

In step 3, the user refines their search by setting values of such attributes and specifying additional criteria, e.g. budget, urgency, proximity

In step 4, GTINs that match the GPC and attribute-values are considered.
The GTIN identifier links to additional information such as product specifications, weight, ingredients, nutritional information, image or description.

In step 5, suppliers offering that GTIN can be identified

In step 6, the offering is associated with price specifications, current availability etc.

In step 7, the supplier's point of sale links to address information as well as geographic co-ordinates, which can be used to display the location on a map.

In principle, this kind of end-to-end searching can be achieved within a small number of SPARQL queries, although in practice, these SPARQL queries will usually be hidden from users when they interact with B2C applications and search engines.

4.3. Rethinking how GS1 provides services for its users

GS1 provides a range of services to users through a variety of mechanisms. In addition to the Global Data Synchronization Network (GDSN), some master data can also be provided via GEPIR, the Global Electronic Party Information Registry. Code lists and definitions are provided via the GS1 Global Data Dictionary (GDD) and GLN Registries are being considered for use in certain sectors, to provide authoritative party information for any GLN, with an endorsement that GS1 has verified this information to be correct.

Furthermore, GS1 operates as a federation of national GS1 member organizations (MOs) that provide a number of services to their local subscriber companies, in addition to any services that are provided globally by GS1 Global Office. Sometimes these services are provided electronically via web interfaces and there are occasions when multiple national member organizations identify a similar need to provide a particular kind of service. This can result in the development of separate disconnected local services that address a common need, although there are also examples where one GS1 MO has made one or more of its services to another GS1 MO, to be used as-is – or further extended to address local needs.

These existing practices have been discussed within the 'Future Network Services' sub-group of the GS1 Architecture Group and Linked Data technology has been identified as a potential solution for semantically describing services, their interfaces, methods and data models, so that when there is sufficient motivation and justification to consider merging such service offerings into a multi-national or global service, it is much easier to connect them together in a meaningful way so that they interoperate seamlessly together, even if the local services use different names for data fields or functional methods that have the same meaning or serve the same purpose.

At present, it is difficult to ensure that master data that is available to such services is complete and also consistently synchronized across all of these channels. Definitions in the web interface for the Global Data Dictionary are currently provided in English only because it was not originally architected to support definitions in multiple human languages. However, this situtation is changing and many GS1 MOs have assisted with the translation of the Global Product Classification system into local languages, so we can anticipate similar efforts to also convert the Global Data Dictionary and other GDSN data models into truly multi-lingual offerings.

GS1 currently uses Uniform Resource Names (URNs) in many of its standards, rather than *HTTP URIs* that can serve as unique names but can also directly *self-resolve* to additional information about those identifiers, *without* the need for specialized lookup services such as the Object Name Service (ONS). Indeed the whole GS1 Global Data Dictionary can be made much more accessible to Linked Data tools if GS1 defines HTTP URI aliases for most of its existing URNs via a small number of simple mapping rules that convert the existing dot and colon delimiters to slashes and map an initial URN prefix to an initial HTTP URI prefix under the gs1.org domain or one or more subdomains dedicated for GS1 linked data resources and ontologies. Note that for URNs derived from GS1 keys (such as the Pure Identity URIs defined in the GS1 Tag Data Standard and used in other GS1 standards such as GS1 EPC Information Services (EPCIS)), our consensus thinking is that the Pure Identity URN format will remain canonical and formed under a GS1-controlled URN prefix, while the brand owner and retailer of the product are free to create HTTP URIs for the product using their respective domain names. This will be discussed further in the GS1 GTIN+ on the Web mission-specific work group that launches in early 2014.

Furthermore, using labels (such as `rdfs:label` or `skos:prefLabel`) and the language tags, e.g. `@en`, `@fr`, `@de`, etc. would provide a W3C standard mechanism for definitions of terms and codes to be provided in multiple languages, rather than solely in English. The Global Data Dictionary would no longer be constrained within an isolated 'walled garden' accessed primarily through a website interface - it would become available also as a fully fledged Linked Data ontology that could use `owl:equivalentClass` , `owl:equivalentProperty` , `owl:sameAs` or `owl:seeAlso` to cross-reference to equivalent terms / data fields, classes and properties defined in other ontologies that are already in use, including schema.org, GoodRelations, FOAF, vCard or the European Commission's Core Business Vocabulary [EU Core Business Vocabulary] and emerging W3C Registered Organization Vocabulary [W3C Reg Org Vocab]. The Simple Knowledge Organization System [SKOS] standard from W3C provides additional ways of expressing more general relationships between concepts defined in different vocabularies. This is discussed in section s of the appendix.

Just as the founders of schema.org have shown a willingness to incorporate the more expressive terms developed in the GoodRelations ontology, it is conceivable that there could be a willingness to incorporate even more specialized granular terms that have already been defined by the GS1 community in the form of the Global Product Classification (GPC) and GDSN Business Message Standards or to recognize a GS1 Linked Data vocabulary as being a perfectly acceptable (perhaps even recommended) way to express rich structured data about products.

For manufacturers, retailers and providers of consumer services, this means that the effort they put into providing accurate and complete master data is leveraged not only for B2B use but that a relevant subset can also be made available for them to use for B2C use and as part of their Search Engine Optimization strategy - this may provide them with a much greater incentive for providing and actively maintaining their master data. For consumers that are searching for products and services that meet their exact needs, it means less time spent actively searching and the possibility of intelligent search agents finding suitable well-researched options that match all of their search requirements, preferences and priorities. Just as GS1 US has developed tools such as Data Driver [http://www.gs1us.org/resources/tools/data-driver] to help companies provide master data into GDSN, we could conceive of GS1 developing tools that allow manufacturers and retailers to export their master data in Linked Data format (e.g. as RDFa or Microdata), ready to embed in their own web pages. The GS1 US B2C Sandbox is one example of such a tool.

However, there are some further complications when embedding Linked Data markup in web pages, because the major search engines currently appear to prefer that any semantic markup is included inline, directly alongside the visible human-readable markup, to reduce the risk that the semantic markup is misleading and unrelated to the visible content of the web page. This means that the process of including Linked Data inline within a web page is difficult to fully automate and needs to take into account the existing web page structure and element hierarchy of blocks of information within the web page. More positively, many websites are data driven, using a small number of web page templates and a database to dynamically generate the pages. This means that it is often sufficient to modify a small number of page templates to include the semantic markup, rather than having to edit a large number of individual web pages and the partially manual analysis of understanding the existing web page structure only needs to be performed for each template rather than for each product GTIN.

Having said that, Linked Data technologies in their current state do not yet represent a complete solution to address all kinds of data used by the GS1 community. At present, the SPARQL 1.0 and SPARQL 1.1 recommendations (W3C standards) lack specialized functions for temporal reasoning or spatial reasoning, although the GeoSPARQL initiative [GeoSPARQL] from the Open Geospatial Consortium proposes an extension to SPARQL to support a number of spatial reasoning functions, e.g. to calculate distances between points, find nearby places or check whether a point is enclosed within a particular bounded region. Since EPCIS event data should be considered as a stream of events representing the physical world history of individual objects, it is currently more appropriate to use Complex Event Processing techniques, rather than SPARQL in order to detect or match patterns of events over a period of time - and to use analysis of such sequences of events to determine or estimate the current state or location of each object. However, it is entirely conceivable that in the future, SPARQL will be further enhanced to support similar capabilities and at that time, a Linked Data approach also to EPCIS event data may also be fruitful. Even then, companies may be understandably hesitant to openly publish serial-level traceability data such as EPCIS event data, because of its commercial sensitivity and the possibility that it can be analysed to determine production rates, inventory volumes, trading relationships etc. For now, there are clearly some easy benefits to be gained by using Linked Data technology to express and access fairly static master data, in order to extend GS1's B2C capabilities before the point of scan.

By aligning with Linked Data techniques and technologies to make its vocabularies, identifiers and data models more accessible, GS1 can more successfully play its part in bringing the Global Language of Business to the web.

5. Further reading and resources

This document is intended to provide a rapid introduction to the opportunities of Linked Data and the Semantic Web, together with some initial ideas about how it can be applied within a GS1 / supply-chain context, particularly to help with B2C strategies.

The Technical Appendix of this paper provides an introductory tutorial on the fundamental W3C technical standards that support Linked Data.

There are many useful books and online resources that provide further information on this topic. These include the W3C Semantic Web Activity home page [W3C SW], the online resources at [LinkedModel], the books [Linked Data], [Linking Enterprise Data], [StructuredData], [Learning SPARQL] and [Working Ontologist].

Acknowledgements

We would like to thank the following members of the GS1 Digital team for constructive feedback on earlier drafts of this paper and for fruitful discussions related to these ideas: Hein Gorter de Vries (GS1 Netherlands), Ralph Tröger (GS1 Germany), Robert Beideman (GS1 Global Office), Ken Traub (Ken Traub Consulting), Steve Bratt (GS1 Global Office), Nicholas Fergusson (GS1 Global Office), Scott Gray (GS1 Global Office), Henri Barthel (GS1 Global Office)

Technical Appendix

This appendix is intended to provide an introductory tutorial about the underlying technologies behind Linked Data. We will focus specifically on the following technologies:

- Uniform Resource Identifiers (URIs) to identify things and relationships
- Resource Description Framework (RDF) to write 'facts' in a machine-readable way
- SPARQL Protocol and RDF Query Language (SPARQL) to match patterns in the data, retrieve facts and construct new facts using rules or matching conditions.
- RDF Schema (RDFS) and Web Ontology Language (OWL) for representing general / background knowledge within a particular domain - and using this to infer additional facts and relationships.
- Simple Knowledge Organization System (SKOS) for expressing relationships between concepts, particularly those defined in separate vocabularies and ontologies.

a. Fundamental standards for Linked Data

Most of us are familiar with the World Wide Web (WWW). It is built upon a small number of fundamental technologies that have been developed by the World Wide Web Consortium (W3C) and the Internet Engineering Task Force (IETF). These standards include:

- **Uniform Resource Identifiers [URI]**
    - URIs are identifiers in string format, which can be globally unique [URI]. URIs include both Uniform Resource Locators (URLs) [URL] and Uniform Resource Names (URNs). More recently, Internationalized Resource Identifiers [IRI] are a generalization of URIs that support the use of a much wider range of international [Unicode] characters within the identifier string. Throughout this document, all references to URIs apply equally to IRIs.
      Within the WWW, URLs are typically used to identify web pages or the location of images or video files/streams to be displayed within a page, as well as the address of scripts that should receive and process data from forms on web pages.
- **HyperText Markup Language [HTML]**
    - HTML provides a way to define the structure and contents of a web page. It is recommended that the presentational layout of a web page is defined using Cascading Style Sheets (CSS), which are referenced within the HTML

markup. Web browsers understand how to process HTML and CSS to render it as a human-readable page.

- **eXtensible Markup Language [XML]**
  - XML provides a way to represent structured data including hierarchical data tree-like structures as well as tabular data. XML is widely used in business-to-business message exchanges because it is independent of any programming language and tools exist in most modern programming languages for transforming it into other representations [XSLT], mapping (data binding) to/from relational databases, as well as performing queries to extract information from its structure [XPath][XQuery]. Furthermore, it is possible to automatically validate XML documents to ensure that they conform to the expected structure and data types for any message format for which an XML Schema Definition file [XSD] is defined.
- **HyperText Transfer Procol [HTTP]**
  - HTTP provides the high-level communications protocol (above TCP/IP) for retrieval of web pages and other multimedia files (e.g. images, videos), as well as sending data to a web resource, deleting web resources or creating new resources.
    A secure version of HTTP (HTTPS) using Transport Layer Security can be used to send HTTP traffic over an encrypted communication channel.

Table 1 summarizes some of the existing technologies that are used for exchanging structured data. We also include a column about Linked Data / Semantic Web technologies that are concerned not only with representation of exchange of structured data, but also the representation and exchange of the *meaning* of the data in a consistent manner.

	Internal (within a company)	**B2B**	**Web**	**Mobile Web**	**Linked Data / Semantic Web**
Identify	fields (column names) within specific relational databases or URIs	URIs, elements in XML namespaces	URIs (primarily URLs) Identify documents, images, scripts	URIs + local element IDs / names within web page local variable names in JSON	URIs (mainly HTTP URIs) Identify anything *including relationships*
Store and Represent Structured Data	Relational databases (Oracle, Postgres, MySQL etc.) - tables of rows, col.s - defined datatypes - joins across tables	XML Lists, Tables, Trees Neutral: Any operating system or programming language Robust validation Supports multiple namespaces	HTML Structure & content of web pages that are designed primarily to be read and understood by humans (difficult for machines to extract meanings)	JSON JavaScript Object Notation Less verbose than XML Natively consumed by JavaScript JSON-LD allows Linked Data to be expressed using JSON syntax and HTTP URIs	RDF Resource Description Framework Stores data as a 'graph' of arcs & nodes Can represent any data structure Can be embedded within web pages using RDFa, RDFa Lite, JSON-LD or Microdata
Query Structured Data	SQL SELECT columns FROM table WHERE condition + joins across tables	XPath XQuery Often transform back to relational data then use SQL internally	Document Object Model (DOM) and XPath, JQuery & CSS selectors - but page structure varies between sites, so extraction of specific data values from web pages is laborious, lots of re-work for each site		SPARQL Query language for matching 'graph' data & Retrieving data from multiple sources
Exchange Structured Data **&** **Perform Remote Functions**	Exchange spreadsheets via e-mail SQL + JDBC/ ODBC connector URIs	AS2 & Web Services	HTTP / HTTPS and REST (representational state transfer) Retrieve / send web pages, images, form data, retrieving Linked Data Use HTTP verbs PUT / GET / POST / DELETE to create / read / update / delete web resources HTTP Content Negotiation Preferred format for returned data can be specified (e.g. HTML, XML, JSON-LD, RDF/XML, Turtle) AJAX (XMLHttpRequest) Allows data to be retrieved without reloading a web page		

Table 1. Technologies used for exchange of structured information and its meaning

Linked Data builds upon existing web technologies to enable individual 'facts' or statements to be asserted (and embedded within documents), searched, retrieved and combined, in order to generate additional 'facts'. The most fundamental Linked Data technologies are:

- **Uniform Resource Identifiers [URIs]**
    - When URIs are used in Linked Data and the Semantic Web, they can be used not only to refer to web pages but also at a much more granular level, to provide globally unique identifiers for the factual information that appears within a web page. A URI might refer to a person, a place, an organization or any kind of relationship. Note that any 'thing' might be referred to by one or more URIs, but a single URI may refer to a single individual 'thing' or to a class of individuals - or to any kind of relationship or property. Two or more URIs can refer to the same thing; this is similar to URL aliases that are used in the world wide web or the use of nicknames for people.
    - Linked Data typically uses HTTP URIs that visibly resemble URLs and can be used either for identifying / naming things in a globally unambiguous way or for retrieving related factual information about those things. Using HTTP URIs for Linked Data means that anything can be easily looked up, whether facts about a product or place – or the definition or a particular attribute or relationship. There is a widely used technique called HTTP Content Negotiation that allow software to select whether they would like to retrieve a web page representation of something or just the underlying structured data in a preferred format specified by the user or software. This is used to request structured data either in a human-readable or machine-readable way.
- **Resource Description Framework [RDF]**
    - Somewhat analogous to the role of HTML in defining the structure and content of information in a web page, RDF is used to represent the factual information contained in a web resource. RDF allows simple logical triples (Subject–Predicate–Object) to be expressed. RDF can be represented in multiple serialization formats, including an XML representation [RDF/XML] and more terse formats (such as N3 and [Turtle] notation), as well as embedding within HTML (using RDF in annotations [RDFa][RDFa Lite] or Microdata or JavaScript Object Notation for Linking Data [JSON-LD]).
- **SPARQL Protocol And Query Language [SPARQL]**
    - SPARQL provides a protocol for retrieving Linked Data from any appropriate web resource, as well as a query language that operates on RDF data, to extract 'facts' that match particular criteria or even to construct new 'facts' based on specified matching rules or patterns that allow such new 'facts' to be inferred from existing 'facts'.
- **RDF Schema [RDFS] and Web Ontology Language [OWL]**
    - RDF Schema [RDFS] and Web Ontology Language [OWL] are both used to express at an abstract level the relationships between classes of entities (e.g. kinds of things) and predicates (property relationships that connect them), as

well as some logical constraints on these, which allow us to infer additional facts and relationships. They can also be used to make statements about individual entity instances, e.g. to say whether two or more things are identical or different.

- **Simple Knowledge Organization System [SKOS]**
  - Simple Knowledge Organization System [SKOS] provides a mechanism to express knowledge systems, thesauri and controlled vocabularies as linked data. SKOS introduces some additional terms to those provided by RDFS and OWL, with subtly different semantics and generally with less strict inference features.

We will discuss these in further detail in sections c-s of this appendix but will first provide some real examples of linked data, to show that it can be very simple to write and query.

b. Examples of Linked Data

At this point, it is probably useful to give some real examples of Linked Data to show what it looks like and how it can be accessed. The Linked Open Data cloud diagram [LOD Cloud] illustrates a number of major sources of open data (many of which provide their data in RDF format), together with an indication of the cross-references among these sources.

One major source of open data is DBpedia [http://dbpedia.org], which provides factual information extracted from the 'info boxes' of facts and figures appearing on Wikipedia pages in RDF format. Another major resource is GeoNames [http://www.geonames.org], which includes geographic information about locations.

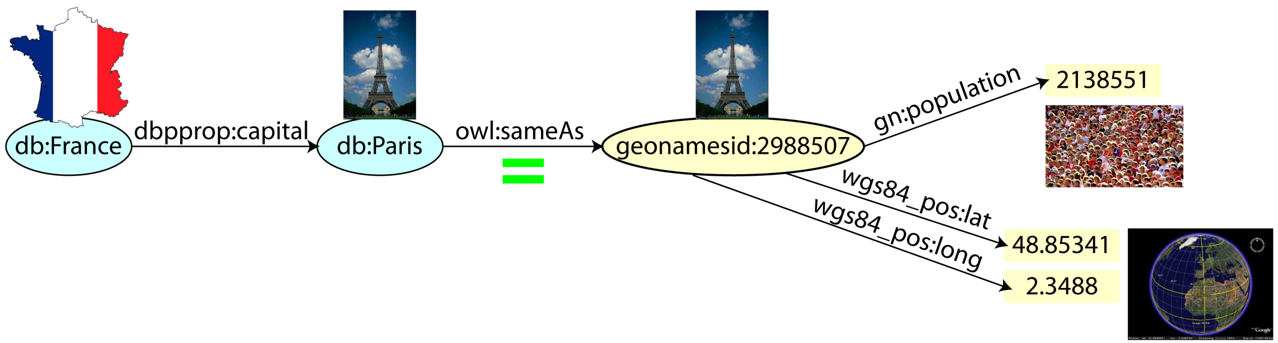
DBpedia contains machine-readable facts such as "The capital of France is Paris", while GeoNames contains additional facts such as "Paris has latitude 48.85341°N and longitude 2.3488°E and a population of 2,138,551". Although DBpedia uses the URI `http://dbpedia.org/resource/Paris` to refer to Paris, whereas GeoNames uses the URI `http://sws.geonames.org/2988507` , DBpedia contains an RDF triple representing a link to say that its URI is semantically identical to the URI that GeoNames uses to refer to Paris. DBpedia uses the OWL '**`sameAs`**' relationship (`http://www.w3.org/2002/07/owl#sameAs`) to express this equivalence between individual resources.

This ability to cross-reference linked data at a very granular level allows any individual or organization the autonomy to independently define and use their own data models (or re-use the data models of others) and to assert facts and link these to facts asserted by other organizations, *without* needing to duplicate information. Note also that Linked Data allows each organization to create its own globally unique identifiers for use in its facts, both for identifying entities (the subject or object of a 'fact') and relationships (which relate the subject to the object); this is achieved though the use of URIs. Organizations may create and use URIs that they define within their own namespace (e.g. derived from a domain name under their own control) or where appropriate, they may also make use of existing URIs that are already defined by other organizations. Figure 5a illustrates graphically how the facts in the example above are linked together, as a *graph* of nodes and arcs, whereas Figure 5b shows how these same facts are expressed in text using RDF Turtle notation.

We can see that URIs are used to identify named entities, as well as the relationships that connect them. RDF notation represents facts as Subject – Predicate – Object triples. These can be read as simple logical sentences having a subject, a predicate or property (acting as the verb) and an object, representing the other entity that is related to the subject.

We can make a simple SPARQL query to retrieve Linked Data that matches our search criteria. An example of a simple SPARQL query and its result is shown in Figure 6.

(a) Graph view



(b) RDF triples

db:France	dbpprop:capital	db:Paris	.
db:Paris	owl:sameAs	geonamesid:2988507	.
geonamesid:2988507	gn:population	2138551	.
geonamesid:2988507	wgs84_pos:lat	48.85341	.
geonamesid:2988507	wgs84_pos:long	2.3488	.

Prefixes:

db: = http://dbpedia.org/resource/
dbpprop: = http://dbpedia.org/property/
dbpedia-owl: = http://dbpedia.org/ontology/
geonamesid: = http://sws.geonames.org/
gn: = http://www.geonames.org/ontology#
owl: = http://www.w3.org/2002/07/owl#
wgs84_pos: = http://www.w3.org/2003/01/geo/wgs84_pos#

e.g. db:France = http://dbpedia.org/resource/France

Fig 5. Representation of some facts from DBpedia and GeoNames
(a) as a graph diagram and (b) as RDF data in Turtle notation

Web page about UK postcode CB3 0FS	http://data.ordnancesurvey.co.uk/doc/postcodeunit/CB30FS.html
Page of RDF data about UK postcode CB3 0FS	http://data.ordnancesurvey.co.uk/doc/postcodeunit/CB30FS.ttl

a ***SPARQL*** *query:*

```
PREFIX ns1: <http://data.ordnancesurvey.co.uk/id/postcodeunit/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX spatialrelations: <http://data.ordnancesurvey.co.uk/ontology/spatialrelations/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX postcode: <http://data.ordnancesurvey.co.uk/ontology/postcode/>
```

Declaration of namespaces and URI prefixes

```
SELECT ?postcode ?latitude ?longitude ?easting ?northing ?district ?ward
```

the ***SELECT*** *clause lists the variables for which values are to be found. Variables names begin with a question mark (?)*

```
WHERE {
?s geo:lat ?latitude .
?s geo:long ?longitude .
?s spatialrelations:northing ?northing .
?s spatialrelations:easting ?easting .
?s rdfs:label ?postcode .
?s postcode:district ?d .
?d rdfs:label ?district .
?s postcode:ward ?w .
?w rdfs:label ?ward .
}
```

the ***WHERE*** *clause defines the graph matching patterns*

The Results:

postcode	latitude	longitude	easting	northing	district	ward
"CB3 0FS"	52.209400	0.087313	542717	258849	"The City of Cambridge"	"Newnham"

Fig. 6. Example of a simple SPARQL query and its results when it is executed at the SPARQL endpoint of DBpedia, http://dbpedia.org/sparql

c. Linked Data technologies in more detail

In the remainder of this technical appendix we describe each of the core technologies in further detail. In Sections d - f, we describe the use of Uniform Resource Identifiers (URIs) to identify all kinds of things, concepts and relationship predicates, Resource Description Framework (RDF) to provide machine-readable 'facts', SPARQL query language to query for facts or create new facts. These are probably the most important Linked Data techniques to be aware of.

Sections g - s introduce RDF Schema (RDFS), the Web Ontology Language (OWL) and Simple Knowledge Organization System (SKOS), which are used to express general knowledge or domain-specific knowledge at an abstract level. Although we have tried to provide familiar examples, these sections can be skipped by readers who are less interested in understanding the techniques used for data modelling.

d. Uniform Resource Identifiers (URIs)

URIs are defined in IETF RFC 3986 [URI] and provide the standard way of identifying resources on the Internet. Resources might be the addresses of web pages and associated multimedia files but they can also be virtual representations of things and places in the physical world, as well as abstract or intangible concepts and relationships, such as 'has capital city:', which DBpedia represents with the following URI:
`http://dbpedia.org/property/capital`

URIs are a collective term that include Uniform Resource Locators (URLs), as well as Uniform Resource Names (URNs), as shown in Figure 7. International Resource Identifiers IRIs are a generalization of URIs to support the use of extended character sets.

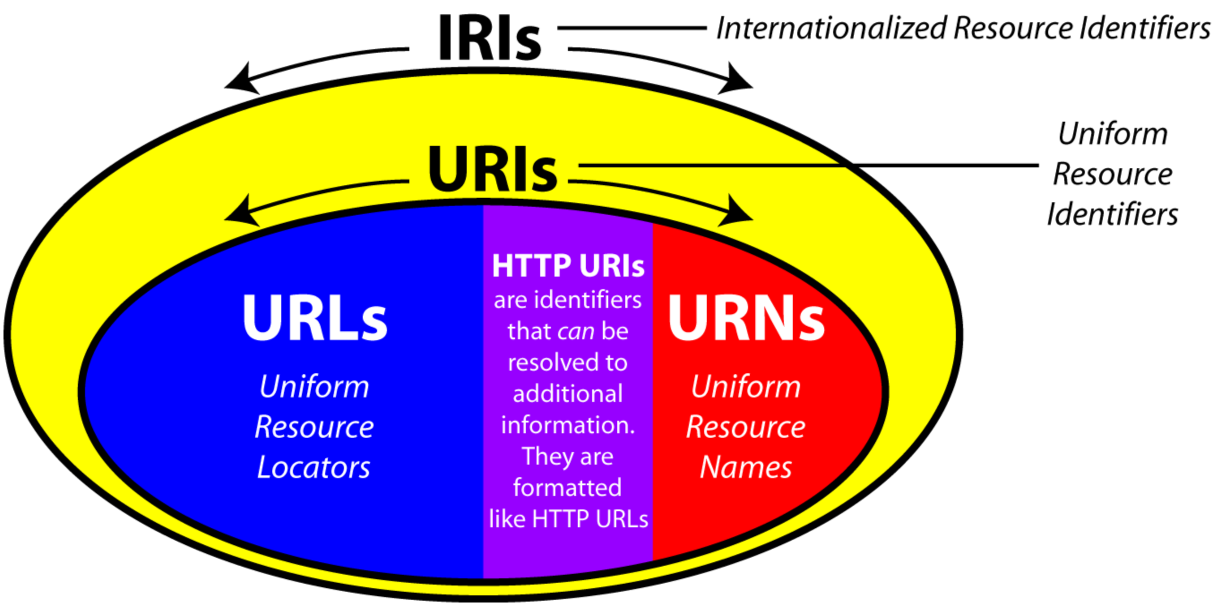


Fig. 7 Uniform Resource Identifiers (URIs) include Uniform Resource Locators (URLs), Uniform Resource Names (URNs) and HTTP URIs. Internationalized Resource Identifiers (IRIs) are a generalization of URIs that also support international characters.

An absolute URL or HTTP URI serves not only as a globally unique name or identifier, but also provides sufficient information for retrieving a representation of information about the thing that it identifies. The process of retrieving this additional information or representation is called 'resolution' or 'resolving' of a URI.

A URN only serves as a globally unambiguous name or identifier but does not provide a well-defined or obvious way of retrieving a representation of it. There may be some dedicated resolvers for particular families of URNs that share the same prefix - but at the present time, there is no established general-purpose global way of resolving every URN.

By creating HTTP URIs instead of URNs, it is very easy for any individual or organization to independently create an unlimited number of globally unambiguous identifiers that can *optionally* also be resolved to additional information representations such as RDF data. In practice this means that HTTP URIs can often be resolved to related information just by using a web browser to retrieve and display that related information, which might be a set of facts or a section of an ontology or vocabulary in which the meaning of the URI is defined.

An individual or organization has the authority to create HTTP URIs for domain names that they have registered and to link this to retrievable information about those identifiers. For example, the UK Ordnance Survey (the government agency responsible for official mapping of the United Kingdom) uses the following URI representation of the UK postcode CB3 0FS.

`http://data.ordnancesurvey.co.uk/id/postcodeunit/CB30FS`

However, that URI also resolves (and redirects) to information about that postcode.

Technical note on HTTP URIs, HTTP 302 Redirection and Content Negotiation:
The URI above is an example of an HTTP URI being used to name or identify a thing, in this case a physical location in Cambridge, UK identified by UK postcode CB3 0FS.

Because an HTTP web request cannot retrieve a physical object or physical location, web servers that serve Linked Data can be configured to automatically perform an HTTP 302 (See Other) redirection to a related HTTP URI that provides an information representation (i.e. data) about the thing or location. In the example above, the HTTP 302 redirection automatically redirects to the following HTTP URI:

`http://data.ordnancesurvey.co.uk/data/postcodeunit/CB30FS`

This HTTP URI returns a web page containing tables of data about the location with UK postcode CB3 0FS, including latitude and longitude co-ordinates, electoral wards, etc.
It is also possible to use HTTP Content Negotiation to request the underlying structured data, without the web page formatting. This is done by setting the appropriate value of the MIME type in the Accept: header of the HTTP request.

A web browser might send an HTTP header that specifies
`Accept: text/html, application/xhtml+xml`

If the web browser or other software instead sends an HTTP header of
`Accept: text/turtle` the RDF data may be returned in terse triple [Turtle] format.
If instead the following HTTP header is sent:
`Accept: application/rdf+xml` the RDF data may be returned in [RDF/XML] format.

Alternatively, if the following HTTP header is sent:
`Accept: application/ld+json` the RDF data may be returned in [JSON-LD] format.

Domain name registration can have very low registration costs and low annual renewal costs, making it very economical for anyone to create an unlimited number of globally unambiguous identifiers for their objects, places and the relationships that they would like to describe.

However, it is not always necessary to create new URIs. We might simply want to assert additional facts about a Subject or Object that is already identifiable through an existing URI or we might want to re-use the URI that another organization has already defined for a particular predicate (to identify a particular property or relationship) - for example re-using the Friend-Of-A-Friend [FOAF] predicate `http://xmlns.com/foaf/0.1/phone` to indicate an associated telephone number in international format. URIs are essentially the noun and verb 'words' of Linked Data and by re-using existing URIs where appropriate, we can avoid the need to invent a whole new language to express our facts, but instead align with existing terminology, where this is sufficient, just as our parents and teachers taught us how to use an established human language for communication.

GS1 currently uses URNs in many of its standards. For example, an Electronic Product Code (EPC) identifier is canonically expressed as a Pure Identity URI - in fact it is specifically this URN notation that is used when an EPC appears within event data within the EPC Information Services standard [EPCIS]. The Core Business Vocabulary standard [CBV], which is typically used in conjunction with EPCIS also uses URN notation for well-defined standard values of fields such as `bizStep`, `disposition`, `readPoint` and `bizLocation` as well as for transaction type identifiers and values. Furthermore, the technical work groups within GS1 who have developed standards such as EPCIS have often had the foresight to specify the data type `xsd:anyURI` for any place where an EPC or other URI or IRI is expected in the XML Schema Definition (XSD) files for such standards. This serves to future-proof the GS1 standards so that they can work not only with current GS1 identifiers in URI format but also much more broadly with *any* Linked Data identifiers formatted as URIs.

e. Resource Description Framework (RDF)

RDF notation is defined in [RDF] and allows us to assert 'facts' consisting of logical triples of the form: Subject – Predicate – Object. This is illustrated in Figure 8 and allows us to build up a directed graph of facts about the relationships between entities. If the Subject is the thing being described, then the Predicate can be one of its attributes and the Object can be the corresponding value of that attribute.

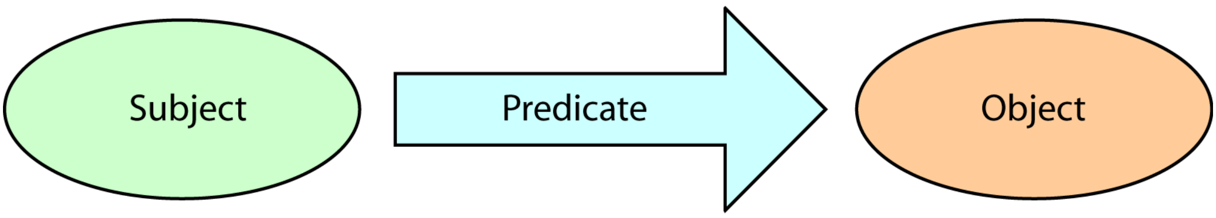


Fig. 8 RDF data consists of Subject - Predicate - Object triples that express 'facts'

There is an XML representation of RDF data [RDF/XML] but a more concise terse representation such as [Turtle] or N3 is typically much easier for humans to read.

RDF data can be embedded within existing web pages using RDF in annotations [RDFa], [RDFa Lite], JSON for Linked Data [JSON-LD] or using HTML5 Microformat notation

[Microdata]. These are just different serializations of RDF in much the same way that XML, ASN.1 and JSON could be considered as different serializations for a particular data model described abstractly in UML.

Sometimes the factual statement we wish to express is more complicated than a 'subject - predicate - object'. For example we might want to express an indirect object or a qualifying phrase. e.g. Retailer R has store S, which offers product P for sale at price X. This kind of statement would need to be broken down into a collection of subject - predicate - object triples, such as:

R has-store S

R makes-offering O

O includes P

O has-price-specification X

Values such as price specifications, dimensions, weight and volume are themselves not a single value. Instead they consist of a numeric value and a unit of measure or currency unit.

In the example above, we would need two further triples:

X has-currency-unit "GBP"

X has-value "0.58"^^xsd:float

This may seem somewhat convoluted, but it actually facilitates conversions between values expressed in different currency units or units of measure, since the unit is not embedded within the string. The “Quantities, Units, Dimensions and Time” [QUDT] ontology was developed by NASA as part of the Constellation Program. It provides a global reference for units of measure, conversion functions (e.g. from pounds to kilograms or feet to metres) as well as dimensionality checking (i.e. checking mathematical formulae to ensure the dimensions on the left-hand-side and right-hand-side are equal).

In the example above, we introduced an instance of an offering, O - but we might not be concerned about giving it a unique identifier. RDF also contains the idea of an un-named node ('blank node') so that we can use the blank node to make statements such as: Retailer R makes-offering that includes P and has price-specification X. Further information about blank nodes can be found in the RDF specifications [RDF].

RDF data can also be made openly available online on the web at an HTTP URI address. For example, http://data.ordnancesurvey.co.uk/doc/postcodeunit/CB30FS.ttl has RDF data provided by the Ordnance Survey (the agency responsible for official mapping of the UK) about the postcode 'CB3 0FS', which is the postcode of the Institute for Manufacturing in Cambridge, where our Auto-ID Lab at the University of Cambridge is located.

Sometimes organizations provide open data in RDF format using this approach. Sometimes, rather than explicitly publishing individual files of RDF data about a single 'thing', they provide a general-purpose SPARQL *endpoint* that can be queried, in order to selectively retrieve RDF data about *many* objects, things or places. For example, with the above approach of publishing an RDF data file for each postcode, we can use SPARQL to extract information about the corresponding latitude and longitude co-ordinates, the nearest city and county, as shown previously in Figure 6. By instead using a SPARQL *endpoint* of Ordnance Survey data at http://api.talis.com/stores/ordnance-survey/services/sparql we can use SPARQL to do additional queries *across* the data for *multiple* postcodes, such as starting with a city and finding the list of postcodes associated with that city or starting with latitude, longitude co-ordinates and obtaining the nearest postcode. We will discuss SPARQL in further detail in the following section.

Although some technology solution providers have developed and optimized 'RDF Triplestores', 'Quadstores' and 'Graph databases' for storing RDF data, RDF data does not need to be stored in a separate standalone database system. Indeed there are adapters (e.g. [D2RQ]) as well as W3C standards (e.g. [R2RML], [Direct Mapping] and [GRDDL]) that enable RDF data to be stored in existing relational databases or collections of XML documents and to be made accessible to SPARQL queries and available as dynamically generated RDF data.

f. SPARQL Protocol And Query Language (SPARQL)

SPARQL provides a mechanism for querying and retrieving RDF data ('facts') that match particular criteria. At first glance, SPARQL has a somewhat similar syntax to SQL that is used to query and update relational database systems. Like SQL, SPARQL has a **`SELECT`** command with a **`WHERE`** clause that is used to express the query matching criteria. However, unlike SQL, which operates on the named tables and columns of a relational database, SPARQL operates on RDF triples and is used to find 'facts' or triples that match specified graph patterns. A graph pattern is simply an RDF triple in which one or more of the Subject, Predicate or Object are not defined with fixed values but are instead expressed using a variable name as a placeholder. In SPARQL, variables are indicated by prefixing with a question mark, e.g. `?phone` is a variable name. We can find matches between subjects and their phone numbers by using a graph pattern such as

```
?subject <http://xmlns.com/foaf/0.1/phone> ?phone .
```

When a match is found, the corresponding value is said to be *bound* to that named variable. Note that it is *not* the name of the variable that determines the kind of data that matches - it is the graph pattern that is specified. In the example above, it is the use of the predicate `<http://xmlns.com/foaf/0.1/phone>` that causes the `?phone` variable to be bound to phone numbers. In SPARQL queries, it is quite common to begin with some **`PREFIX`** statements that declare a number of namespaces and associated URI prefixes. For example, if we include the following declaration:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

Then we can write the above graph pattern in our SPARQL query more concisely as:

```
?subject foaf:phone ?phone .
```

Later in some of the examples in this document, we use the prefix `ex:` to refer to a *fictitious* namespace or ontology within the domain `http://example.org`

In addition to the **SELECT** command, SPARQL has other commands such as **ASK**, **DESCRIBE** and **CONSTRUCT**. **CONSTRUCT** is particularly useful for constructing new RDF triples ('facts') that are derived from existing triples ('facts') according to graph matching patterns specified in the **WHERE** clause.

SPARQL 1.1 recommendation [SPARQL 1.1] (which is already widely implemented) has the capability to consider not only a single local source of RDF data but also *multiple sources* of local RDF data and remote RDF data that are available at specified URLs or via remote SPARQL endpoints. In this way, SPARQL 1.1 supports *federated queries*, which are very convenient for performing data 'mash-ups' that combine data from multiple sources and especially where we want to explore the relationship linkages across different data entities. The *predicate path* feature also allows us to navigate along a chain of relationship linkages to see if and how two things are connected. This is somewhat analogous to the idea of six degrees of separation that connect all humans alive today - i.e. that one person is likely to be a friend-of-a-friend-of-a-friend-of-a-friend-of-a-friend-of-a-friend of any other person. The Friend-Of-A-Friend ontology [FOAF] defines a predicate `foaf:knows` which can be used to indicate links to people you know. Social networks such as LinkedIn already display how their users are connected to other users, via one or more intermediate users.
In SPARQL, two individual people A and B might be connected directly or indirectly via a chain of mutual friends if A `foaf:knows+` B. The plus symbol (+) is used in predicate path notation to say '1 or more' of this predicate in series, i.e. A `foaf:knows` B directly or through a number of intermediate parties.

In a supply chain context, this predicate path feature of SPARQL 1.1 can be very useful for evaluating a chain of custody, to determine if somebody asking for information about an individual object was on its individual supply chain path and whether or not there is an unbroken chain of trust that connects them to the organization that has information about the object. We have experimented with this approach in the development of the access control framework for supply chain Discovery Services and continue to refine these ideas in the GS1 Event Based Traceability work group.

A concrete and highly practical example of the exploration of connectedness is in the life sciences domain, where SPARQL is being used to help with the discovery and development of new drugs to cure diseases. Academic literature typically consists of journal papers that report new theories and/or the results and conclusions of experimental work that reveals new fundamental science in understanding how the natural world works. An individual journal article might provide information about the link between a disease and a particular virus carried by a parasite, or the link between the surface chemistry of a specific virus and specific receptors within the human body, or the development of proteins that are produced by the virus and indicate its presence, or molecules that have been demonstrated to interfere with the binding of a virus particle to the receptors within the body. Unfortunately, it is often difficult for an individual journal paper to provide the entire end-to-end story about which biochemical compounds can be used to treat a particular disease or to anticipate all the potential future implications of the results, especially when missing links in the chain of research are only revealed in future discoveries and research. However, if the knowledge contained within each individual paper can be efficiently summarized and indexed, then we can perform a SPARQL query that can traverse a number of relationships in order to make a much more targeted approach to finding suitable candidate molecular compounds for the development of more effective treatment of diseases. Individual RDF triples can be regarded almost like jigsaw pieces with very specific shapes for their 'plugs' and 'sockets' and SPARQL can be viewed as a very efficient tool for solving such 'jigsaw puzzles' and making end-to-end connections across data and even across different datasets, through the *predicate path* and *federated query* capabilities introduced within the SPARQL 1.1 recommendation. An example of the use of linked data technology in the life sciences can be seen at [LinkedLifeData].

Returning to a supply chain context, in section 1 and Figure 4, we showed a 'jigsaw' illustration to show how we can navigate through several relationship linkages to start from a consumer's desire for a particular kind of product or service and various criteria (specifications, budget constraints etc.) to actionable information such as a marker on a map showing the nearest supplier. Although this could have been done previously by issuing a number of separate queries against different databases, SPARQL and linked data (RDF) make it possible to do this kind of end-to-end query within one or two queries that access data from multiple sources. In practice, members of the public will not be expected to learn SPARQL, although SPARQL could be used 'behind the scenes' by software applications and search engines.

SPARQL v1.1 introduces a number of additional enhancements beyond what was provided in SPARQL v1.0. Other important enhancements include support for insertion and deletion of data and service descriptions of SPARQL endpoints.

g. RDF Schema (RDFS)

RDF Schema [RDFS] and the Web Ontology Language [OWL] provide similar but complementary roles in expressing high-level constraints and entailments (what can be inferred) about RDF data. RDFS allows us to define classes and properties, subclasses and subproperties, as well as domain and ranges and human-readable labels. These will be discussed in the following sections. RDF Schema [RDFS] plays a very different role from XML Schema [XSD]. XML Schema files can be used to validate the format of an XML document and to constrain its structure and data types, by specifying the hierarchy of the XML elements and their permitted attributes, child elements and the data types (e.g. integer, string, date time) of any values - and to report an exception if a document or message is incorrectly formatted. RDF Schema [RDFS] does not serve to constrain the RDF triples that are asserted. Instead, both RDFS and OWL can be used to describe an abstract data model and its relations and can be used by automated logical reasoners to *infer* additional triples in addition to the triples that are explicitly stated in the data. This means that we can use ontologies expressed in RDFS and OWL to represent 'general knowledge' or 'domain-specific knowledge' and to discover new 'facts' in the data that are not explicitly stated in the original data. Unfortunately, many current implementations of SPARQL are not fully aware of RDFS and OWL, so SPARQL queries typically operate on explicitly stated triples, rather than the additional triples that can be inferred from the explicitly stated triples using the ontologies written in RDFS and OWL. However, some implementations of linked data repositories (e.g. [OWLIM] from Ontotext) use RDFS and OWL ontologies to generate all the possible additional triples ('facts') that can be inferred, at the time of loading the data, as well as removing them if the original facts are deleted or updated. In this kind of implementation, both the explicit 'fact' triples and the inferred 'fact' triples are then available to the SPARQL query.

h. RDFS - Classes and Properties, Subclasses and Subproperties

RDF Schema (RDFS) introduces basic concepts such as classes and properties, as well as subclasses and subproperties. A class is a set of zero or more individuals that share some common class-level features. A subclass is a subset of a class that inherit the features of the parent class but have some additional features that are specific to members of the subclass.

Properties are predicates that describe how a Subject is related to an Object. A Subproperty is an additional property relationship that specializes the parent property in some way.

It is probably easiest to explain these with some familiar concrete examples, as illustrated in Figure 9:

We might have a class '`ex:DairyProducts`' identified by the full URI `http://example.org/ns#DairyProducts` and within it two subclasses, '`ex:Cheeses`' and '`ex:Yoghurts`'. Note that the domain name `example.org` is fictitious and is used in examples for illustrative purposes. In practice, these URIs could be defined under any real domain name, such as gs1.org or any of its subdomains.

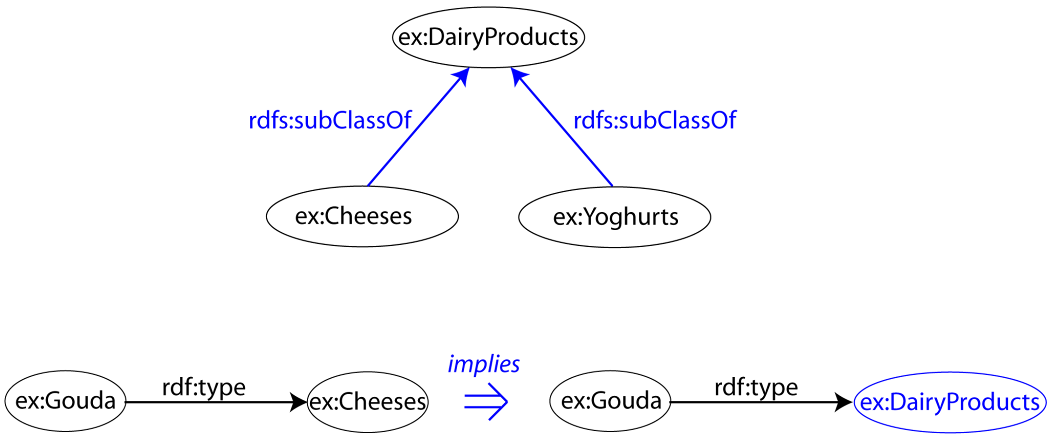


Fig 9. `rdfs:subClassOf` allows us to infer that an individual is also a member of the parent class.

If an individual is a member of a subclass, then it is also a member of the parent class.

For example, because our ontology might state that 'ex:Cheeses' is a subclass of 'ex:DairyProducts', if we have a statement that 'ex:Gouda' is a type of 'ex:Cheeses', we can also infer that 'ex:Gouda' is also a type of 'ex:DairyProducts'. There might be several layers of hierarchy, as in the GS1 Global Product Classification system [GS1 GPC]

As well as subclasses, RDFS can also be used to define subproperties. For example, we might have properties such as 'ex:hasFather' and 'ex:hasMother' and both of these might be defined to be a subproperty of 'ex:hasParent', as shown in Figure 10.

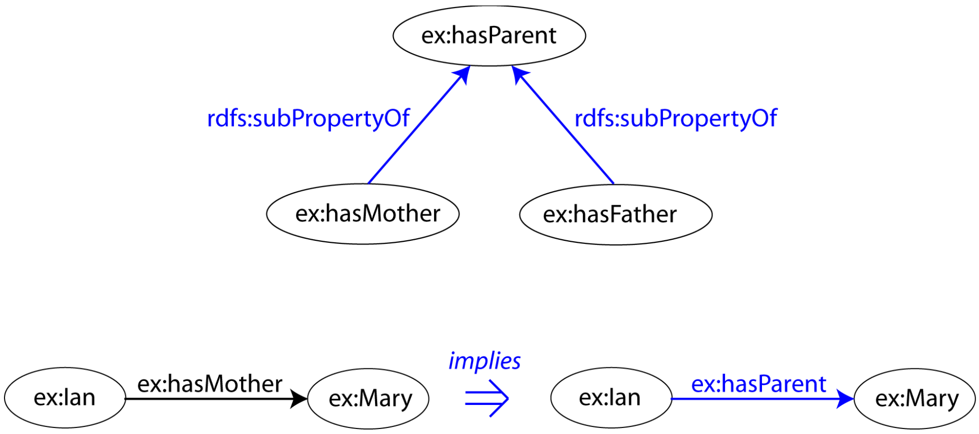


Fig. 10 rdfs:subPropertyOf allows us to infer that the subject and object are also related via the parent property.

e.g. if `ex:Ian ex:hasMother ex:Mary`, then we can infer that:

```
ex:Ian ex:hasParent ex:Mary
```

Classes and subclasses, properties and subproperties therefore allow us to represent domain knowledge using hierarchical data models and relationships. At first appearance, this may seem somewhat like UML Class Diagrams although the idea of 'inheritance' from Object-Oriented Programming does not correspond in a straightforward way to classes and subclasses, properties and subproperties. For example, `ex:hasParent` is a subproperty of `ex:hasAncestor` and although `ex:hasAncestor` is a *transitive* property (in the sense that an ancestor of an ancestor is itself an ancestor – see section m), the same cannot be said of the subproperties `ex:hasParent` and `ex:hasMother` – even though they are both subproperties of `ex:hasAncestor`, neither is a transitive property.

An example of a hierarchical classification scheme used within GS1 is the Global Product Classification (GPC) that is used to refer to a category of product by its functionality, without any reference to it being manufactured by a particular manufacturer. A GPC is therefore useful as a globally unique identifier for a 'yellow pages keyword' for sourcing products even when we don't know at that stage who manufactures such a product; it can also be useful for finding potential alternative products and suppliers for a product that has the same GPC code - they might not be identical matches in terms of technical specifications, ingredients or quality - but they are candidates to consider. GPC identifiers are constructed on a hierarchical basis, as shown in Figure 11.

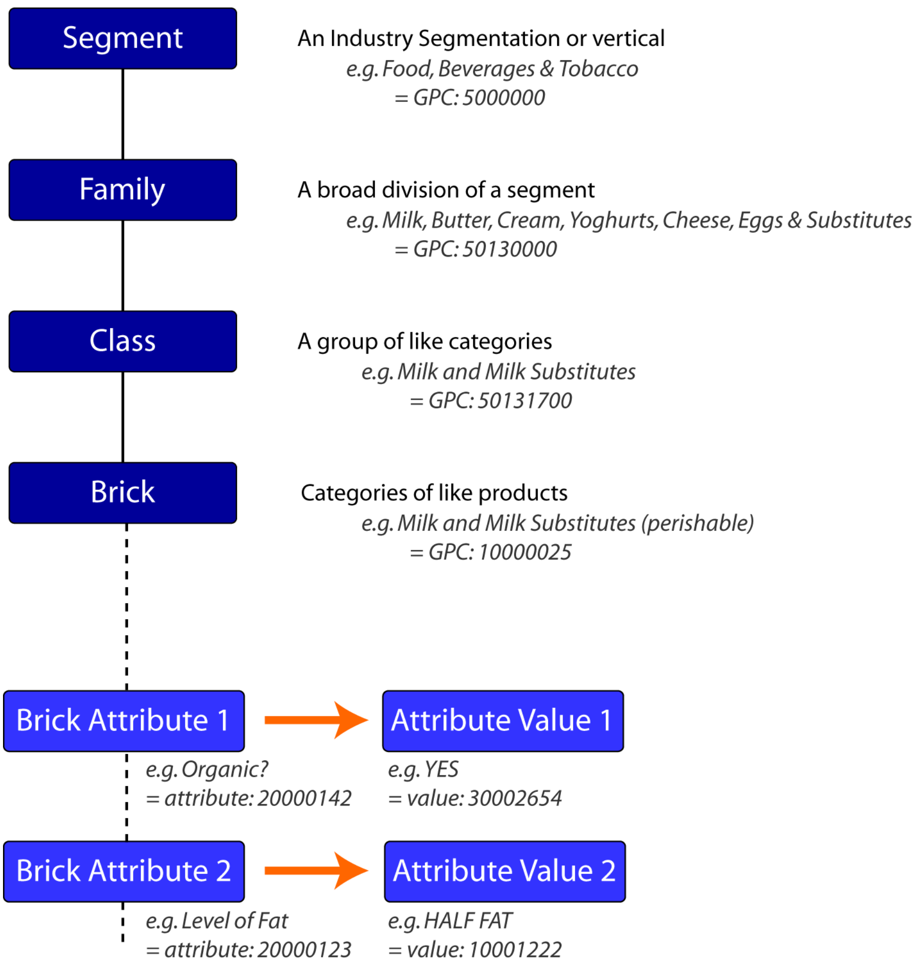


Fig. 11. GS1 Global Product Classification hierarchy, with examples [GS1 GPC]

Each product class GTIN corresponds to only one GPC 'brick', with further refinement through the use of attributes and values. Several different GTIN identifiers (of functionally similar products from different manufacturers) might share the same GPC 'brick' but differ in some of the brick attributes, which can express claims such as whether a product is from animal origin, low-fat or organic or from a specific geographic origin / region of production.

A GPC identifier such as '`10000025`' identifies products that are Milk / Milk Substitutes (Perishable) and includes a definition as well as a note about related but distinct categories (i.e. see also, different from). Within the GPC hierarchy, '`10000025`' lies within the class '`50131700`' (Milk/Milk Substitutes') within the family '`50130000`' ('Milk/Butter/Cream/Yogurts/Cheese/Eggs/Substitutes') within the segment '`50000000`' ('Food/Beverage/Tobacco'), as shown in Figure 11.

Currently, GS1 provide a web-based GPC browser at http://www.gs1.org/1/productssolutions/gdsn/gpc/browser/ for exploring the GPC hierarchy. However, to make this hierarchical classification available using Linked Data tools, it would be logical to consider defining an HTTP URI scheme for all GPC identifiers, as well as GPC attributes and GPC attribute values and to additionally publish the GPC classification hierarchy as a formal OWL ontology, using concepts such as OWL subclasses (e.g. each GPC Brick is a subclass of the corresponding GPC Class, which is a subclass of the corresponding GPC Family, which is a subclass of the corresponding GPC Segment), to use **`rdfs:label`** to publish a human-readable name for that segment/family/class/brick in various human languages (indicated by the ISO-639 standard 2-letter language suffix, e.g. **`@fr`**, **`@de`**), as well as using **`rdfs:isDefinedBy`** to point to the authoritative definition, **`rdfs:abstract`** to provide a description and **`rdfs:seeAlso`** to refer to related classifications, including related categories in other classification systems such as [UNSPSC], Common Procurement Vocabulary [CPV] and [eCl@ss]. The cMap project [cMap] has developed a mapping across these specific classification systems.

Figure 12 provides an illustration of how a small extract of the GS1 Global Product Classification [GS1 GPC] system could be expressed using linked data techniques such as RDF, RDFS. At the time of writing, GS1 has not yet defined the HTTP URI formats it might use for Linked Data to represent thing like GS1 identification keys, GPC bricks, attributes and values or other vocabulary elements and ontologies.

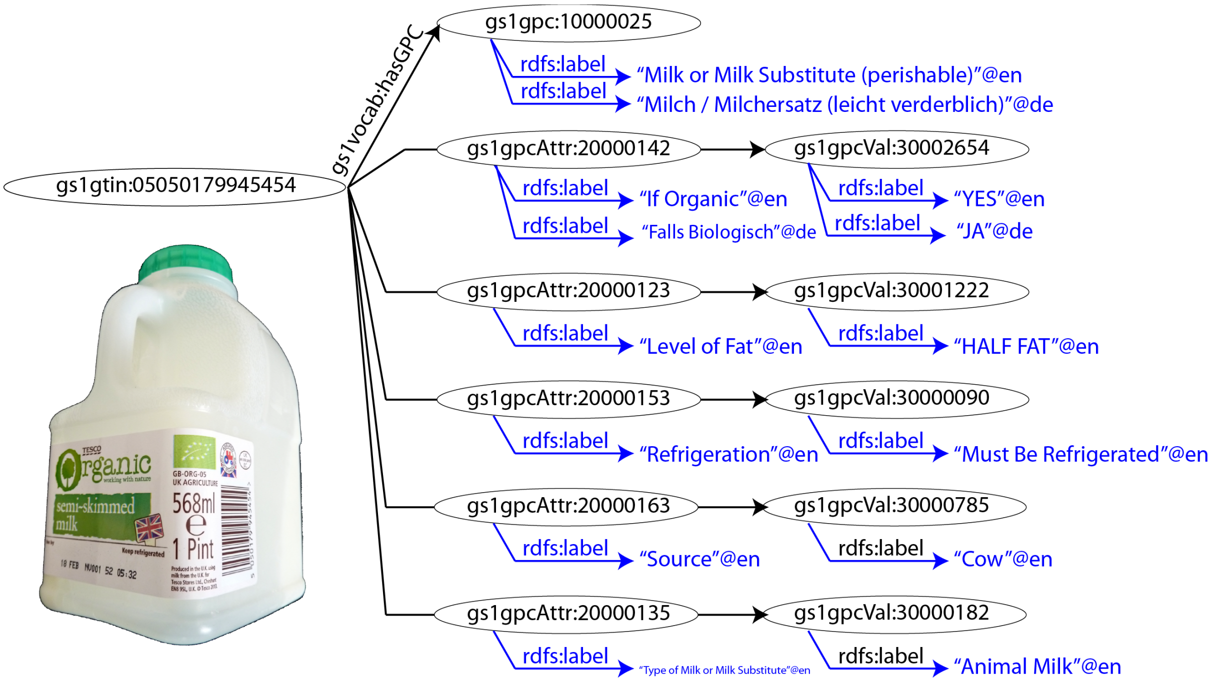


Fig. 12 An illustration of the potential representation of GTIN identifiers, GPC bricks and GPC attribute-value pairs as Linked Data.

i. RDFS - Ranges and Domains

RDF Schema also allows us to infer information about the class of the Subject or the Object.

In RDFS, we can say that '`ex:isSonOf`' has a *domain* of '`ex:MalePerson`' by writing the following triple in the ontology that defines the '`ex:hasMother`' relationship:

```
ex:isSonOf rdfs:domain ex:MalePerson .
```

So, if, `ex:Ian ex:isSonOf ex:Mary`, we can then infer that `ex:Ian rdf:type ex:MalePerson` because *domain* specifies a class of the *Subject*.

In RDFS, we can say that '`ex:hasMother`' has a *range* of '`ex:FemalePerson`' by writing the following triple in the ontology that defines the '`ex:hasMother`' relationship:

```
ex:hasMother rdfs:range ex:FemalePerson .
```

From our previous example, `ex:Ian ex:hasMother ex:Mary`, we can then infer that `ex:Mary rdf:type ex:FemalePerson` because *range* specifies a class of the *Object*.

In Figure 13, we illustrate these inferences using domain and range.

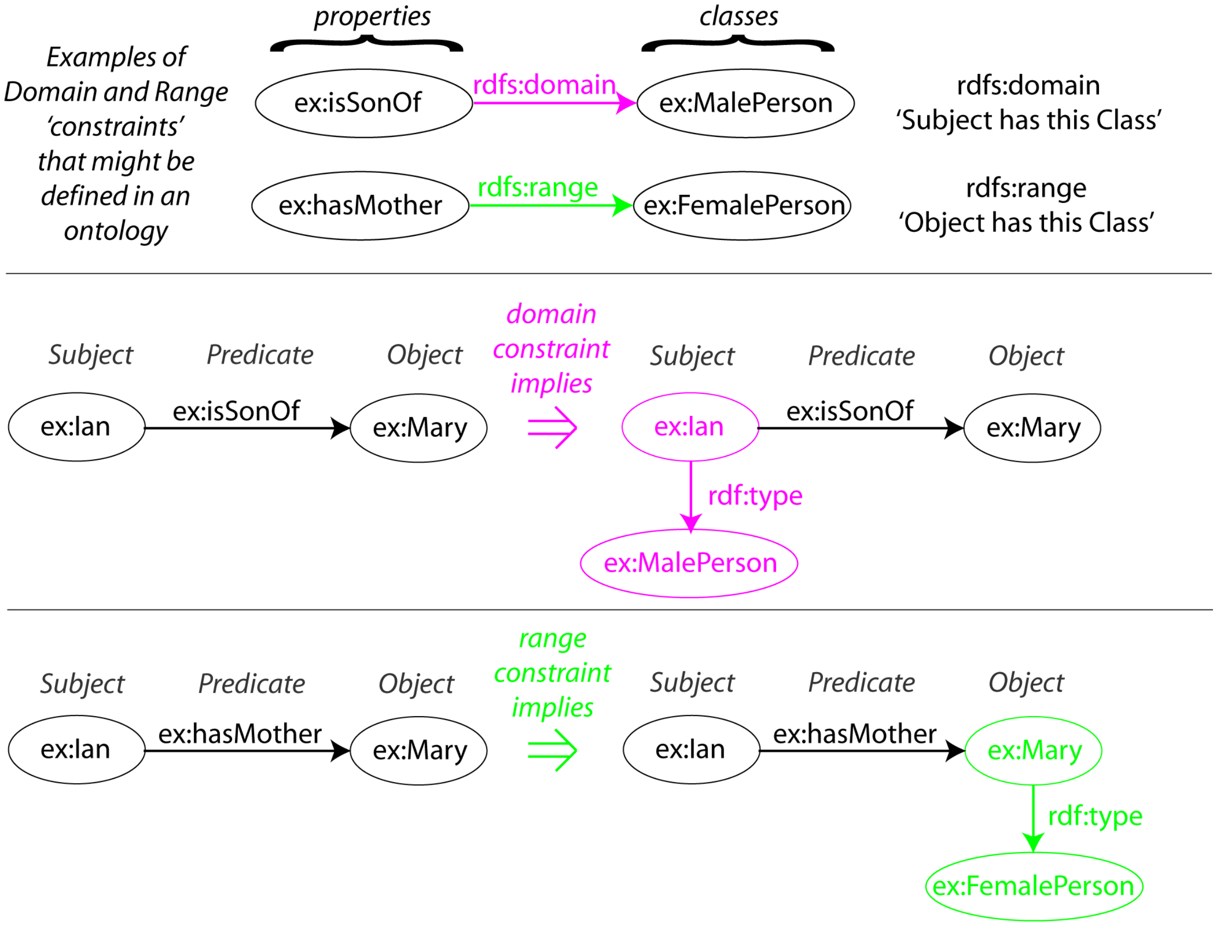


Fig. 13 Illustration of inferences or entailments that rely upon RDFS domain and range constraints

The *domain* and *range* relationships that can be expressed in RDFS allow us to further refine our data models and to use existing facts about properties to infer additional facts about the classes of the Subject and Object, respectively. However, note that any property is defined independently of any class. Furthermore, a resource may be a member of multiple classes.

j. RDFS - Labels and language tags

URIs provide a globally unique identifier for properties, classes and individuals. However, sometimes we want to have human-readable *labels* of those URIs in different human languages. RDFS allows us to associate one or more labels with a particular URI, so that when the data is presented (or requested) in a particular language, the label can be shown as a human-readable string, rather than showing the machine-readable URI.

For example, the RDF data at DBpedia about `http://dbpedia.org/resource/Paris` could contain several label values in different languages,

```
http://dbpedia.org/resource/Paris  rdfs:label   "Paris"@en .
http://dbpedia.org/resource/Paris  rdfs:label   "Parigi"@it .
```

Labels allow us to present human-readable names for URIs in different human languages, while keeping a single global URI or IRI to represent the class, property or individual - i.e. we don't need to translate the URI into multiple versions for each language - we simply look up the corresponding RDFS *label* in our desired language.

The obvious implication for GS1 is that GPC attributes and values, as well as terms and code lists appearing in the GS1 Global Data Dictionary [GDD] should have a URI / IRI representation for each name of a data field and each permitted value within a controlled vocabulary, while multiple human readable translations of these can be provided in different human languages, using the RDFS *label* technique and the language tags, e.g. '**@en**', '**@it**', in the examples above.

The web-based GPC browser already provides the GPC hierarchy in multiple languages, so it should be quite straightforward to make this information also accessible as multi-lingual Linked Data.

The GS1 Core Business Vocabulary already uses URIs (currently URNs) to define standard global terms for values of EPCIS event data fields, e.g. for the bizStep field, a URN value of `urn:epcglobal:cbv:bizstep:shipping` is defined, but additional human-readable label and definition strings in various languages could be associated with this when the GS1 Core Business Vocabulary is published as a Linked Data ontology. When publishing as Linked Data, GS1 may wish to consider defining HTTP URI aliases for each of its existing URNs. For example (purely speculative and fictitious), the URN above could be given an HTTP URI alias such as `http://vocab.gs1.org/epcglobal/cbv/bizstep/shipping` and doing so, would allow a page of RDF linked data to be available at that address, providing definitions and labels in multiple human languages.

k. Web Ontology Language (OWL)

Web Ontology Language is more expressive than RDF Schema [RDFS] and allows us to model the following:

- intersections, union and complements of classes
- inverse properties, transitive properties, symmetric properties
- functional properties, inverse functional properties
- equivalent classes or equivalent properties
- chaining of properties using `owl:propertyChain`

OWL also allows us to declare whether two individuals are the same or different.

l. OWL - Inverse properties

If a property P relates a Subject S to an object O, then the inverse property P' relates the object O to the subject S, i.e.

```
S P O  =>  O P' S
```

As an example of inverse properties, `ex:hasFather` and `ex:isFatherOf` are inverse properties of each other, as illustrated in Figure 14.

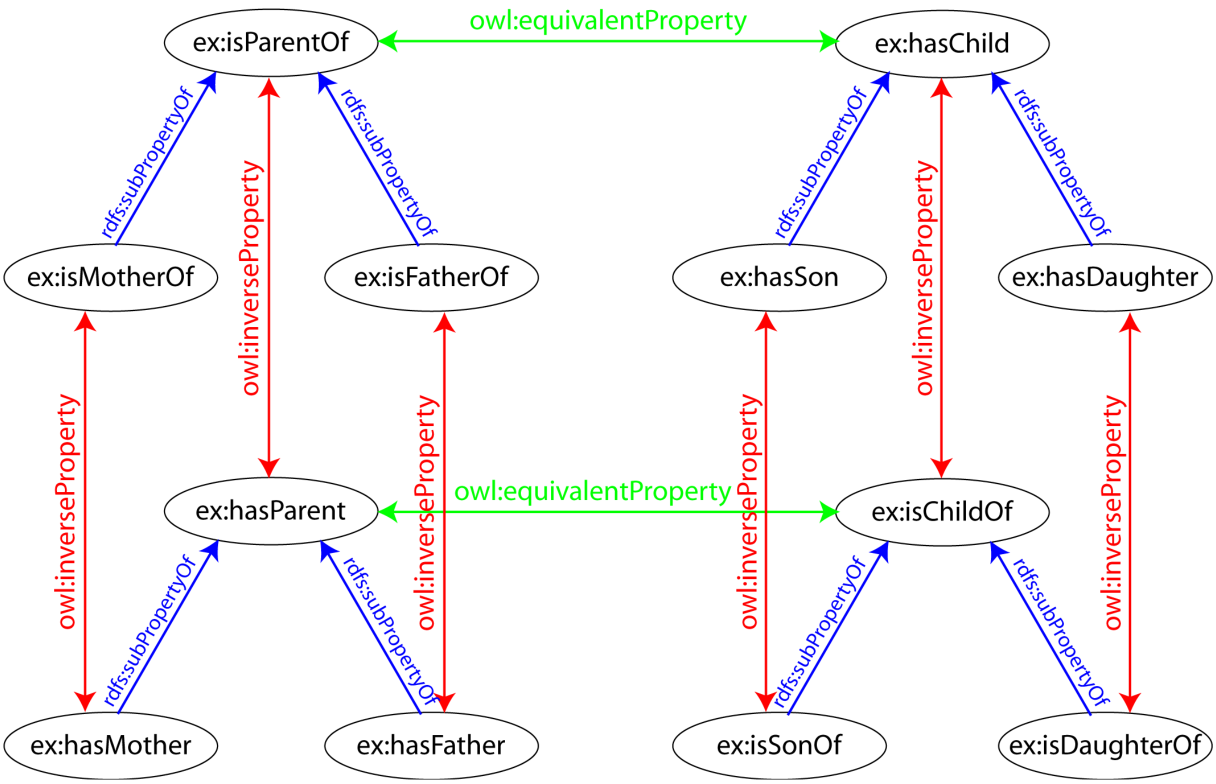


Fig. 14 Use of RDFS subPropertyOf and OWL inverseProperty and OWL equivalentProperty to represent some domain knowledge at an abstract level, in this case for family relationships

m. OWL - Transitive properties

A property `P` is transitive if `x P y` and `y P z` implies `x P z`.

Some examples of transitive properties are `ex:hasAncestor` , `ex:hasDescendant` , `ex:contains` , `ex:isContainedInside` .

Transitive properties are therefore useful for doing reasoning about containment hierarchies and spatial hierarchies. In a logistics context, if we know that an item I is contained within a case C and that the case C is contained within pallet P, we can infer that item I is contained within pallet P.

n. OWL - Symmetric properties

Symmetric properties are self-inverses, i.e. `S P O => O P S`

A couple of familiar examples of symmetric properties are ex:hasSibling or ex:isSiblingOf

Note that ex:hasBrother is not symmetric if the subject is female, whereas ex:hasSibling is symmetric. Sometimes even if a property is not symmetric, we can often find a symmetric broader property or parent property of which it is a sub-property. This is illustrated in Figure 15.

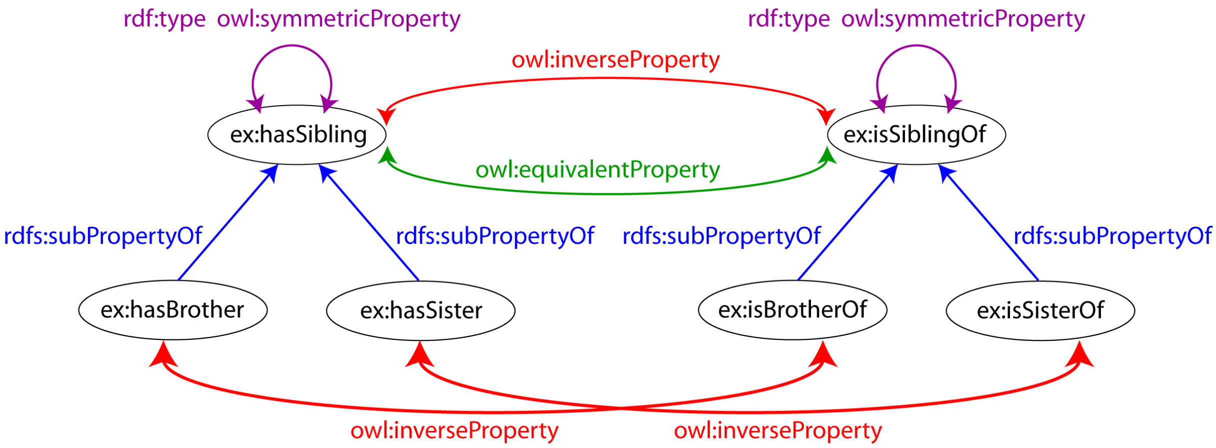


Fig. 15 Symmetric properties are self-inverses

o. OWL - Functional properties

A property P is a functional property if x P y and if there is only one possible value of y for a particular x. An everyday example of a functional property is 'has date of birth', since nobody can have more than one date of birth. In a GS1 context, 'has GPC brick' is a functional property because any given Global Trade Item Number (GTIN) that identifies a product class from a specific manufacturer can be assigned to at most *one* Global Product Classification (GPC) brick value that classifies the type of product that it is. Functional properties therefore imply a many-to-1 cardinality restriction.

p. OWL - Equivalent Classes and Equivalent Properties

Equivalent classes are useful for stating that two or more classes are semantically equivalent. Equivalent properties are useful for stating that two or more properties are semantically equivalent. These are particularly useful for mappings between ontologies that have been developed independently but which have some semantic overlap of some classes or properties.

For example, both the VCard ontology and the schema.org ontology have a property that relates a Subject to its postcode and we can use OWL **equivalentProperty** to say:

```
http://www.w3.org/2006/vcard/ns#postal-code
http://www.w3.org/2002/07/owl#equivalentProperty
http://schema.org/postalCode .
```

q. OWL - sameAs and differentFrom

Because two or more URIs might refer to the same individual, OWL also has the predicates **sameAs** and **differentFrom** to make it possible to explcitly state whether two individuals are the same or are distinct.

For example, DBpedia has RDF facts about Paris, the capital of France - but it also links to a resource at GeoNames, which also has RDF facts about Paris. DBpedia uses the OWL **sameAs** predicate to make this link between the two datasets to say that they refer to the same resource. This means that facts (RDF triples) from DBpedia and GeoNames can be merged because they both refer to the same entity – in this case Paris, capital of France.

```
http://dbpedia.org/resource/Paris
http://www.w3.org/2002/07/owl#sameAs
http:/sws.geonames.org/2988507 .
```

r. RDFS - seeAlso

Sometimes two classes, two properties or two individuals are semantically closely related but not strictly identical or equivalent. RDFS has the **rdfs:seeAlso** property to express that another resource might have some additional information, without implying that the other resource is strictly equivalent or identical.

Using RDF Schema (RDFS) and Web Ontology Language (OWL), we can not only formally describe our knowledge representation or data model about classes and properties. We can also cross-reference across different ontologies developed by different organizations. Because we use URIs to unambiguously represent classes, properties and individuals, it is perfectly acceptable to use a mixture of ontologies within RDF data that we publish; if one ontology does not adequately define a property or class that we need, we can simply use the property or class defined in another ontology – the use of URIs avoid any ambiguity about the namespace (or which ontology a particular term came from). A number of core ontologies are widely used - see Table 2. A more extensive list can be found at [Ontologies].

rdf	RDF Vocabulary http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	RDF Schema (RDFS) Vocabulary

	http://www.w3.org/2000/01/rdf-schema#
owl	Web Ontology Language (OWL) http://www.w3.org/2002/07/owl#
xsd	XML Schema http://www.w3.org/2001/XMLSchema#
dc	Dublin Core ontology http://purl.org/dc/elements/1.1/
dcterms	Dublin Core Metadata Terms http://purl.org/dc/terms/
skos	Simple Knowledge Organization System (SKOS) http://www.w3.org/2004/02/skos/core#
foaf	Friend Of A Friend (FOAF) http://xmlns.com/foaf/0.1/
vcard	vCard ontology (personal contact information including address) http://www.w3.org/2001/vcard-rdf/3.0#
geo	Basic Geo (WGS 84 latitude, longitude) vocabulary http://www.w3.org/2003/01/geo/wgs84_pos#
georss	GeoRSS ontology http://www.georss.org/georss/
ical	iCal (RDF Calendar) http://www.w3.org/2002/12/cal/ical#
geonames	GeoNames ontology http://www.geonames.org/ontology#
gr	GoodRelations ontology http://purl.org/goodrelations/v1/
dbp	DBpedia resource http://dbpedia.org/resource/
dbpedia-owl	DBpedia ontology http://dbpedia.org/ontology/
dbpprop	DBpedia property http://dbpedia.org/property/

Table 2. Some frequently used Linked Data ontologies and their usual compact URI prefixes

s. Simple Knowledge Organization System [SKOS]

SKOS is a W3C Recommendation that provides a way to represent knowledge organization systems, including controlled vocabularies, as distributed linked data.

Whereas RDFS provides **rdfs:label** to indicate a human-readable label for a resource URI, SKOS defines three sub-properties of **rdfs:label**, namely **skos:prefLabel**, **skos:altLabel** and **skos:hiddenLabel**.
Like **rdfs:label**, **skos:prefLabel**, **skos:altLabel** and **skos:hiddenLabel** can include values in various human languages, using the language tags, eg. **@fr**, **@de**.
skos:prefLabel is used to indicate the preferred human-readable label for a resource. A resource should have no more than one preferred label per language tag.
skos:altLabel is used to indicate an alternative human-readable label, e.g. a synonym for the resource.
skos:hiddenLabel is used to indicate a human-readable label that should not be displayed, but which might be used for the purposes of matching an input text string to the resource. For example, if **skos:prefLabel** or **skos:altLabel** use accented characters, the corresponding non-accented characters might be used within the value for **skos:hiddenLabel**, to cater for users who have difficulty in entering the accented characters through their keyboard - or who fail to notice the accents.
SKOS also defines a number of semantic relations, such as **skos:broader**, **skos:narrower** and **skos:related**.
skos:narrower points to a narrower (more specialized) resource or class of objects.
skos:broader points to a broader (more generalized) resource or class of objects.
For GPC Global Product Classification (GPC), we might use **skos:broader** and **skos:narrower** to indicate the relationship between specific GPC Bricks, GPC Classes, GPC Families and GPC Segments, as illustrated in Figure 16. The figure also shows possible use of **skos:altLabel** to provide synonyms and **skos:hiddenLabel** to show labels without accented characters.

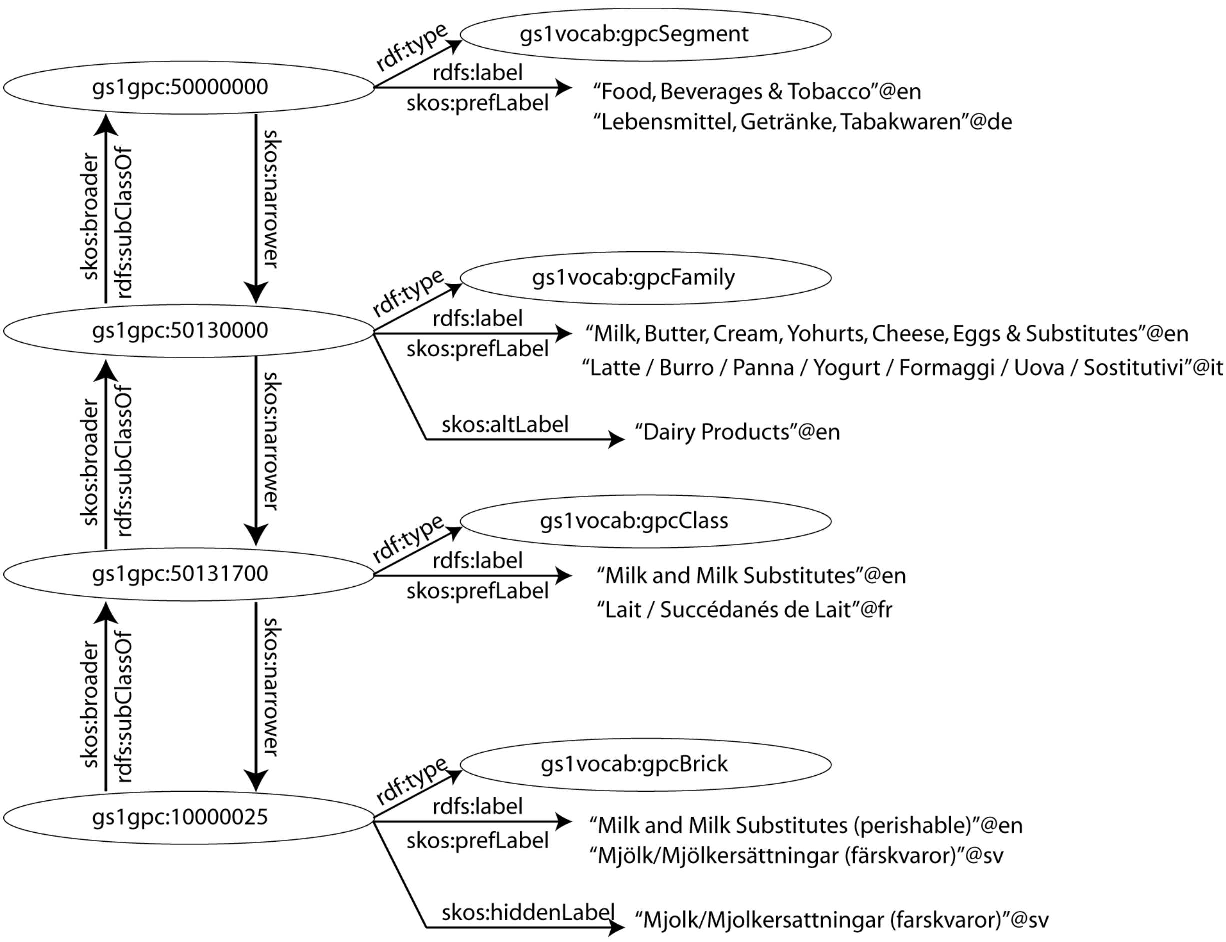


Fig. 16 Example of potential uses of SKOS terms for describing the GPC hierarchy

SKOS also defines terms such as **`skos:exactMatch`**, **`skos:narrowMatch`**, **`skos:broadMatch`** and **`skos:closeMatch`**, which are particularly useful for indicating relationships between concepts and terms defined in different controlled vocabularies. These could be useful for expressing the relationship between categories within different product classification schemes, such as eCl@ss, UNSPSC, CPV, or tariff codes used by customs agencies where the alignment with GPC might not always be exact or 1-1. These terms in SKOS generally have much less inference-based semantics than similar terms in RDFS or OWL, such as **`rdfs:subClassOf`**, **`owl:sameAs`**, **`owl:inverseOf`**.
SKOS also includes the idea of concept schemes and defines predicates such as **`skos:inScheme`** (to point to the scheme of concepts of which the subject is a member), as well as **`skos:hasTopConcept`** (to point to the highest-level (most general) concept of the scheme) and its inverse, **`skos:topConceptOf`**.
The SKOS W3C Recommendation [SKOS] and the [SKOS Primer] provide additional information about SKOS. SKOS is being used to express various vocabularies, including the United Nations Agriculture Vocabulary [AGROVOC], the US Library of Congress Subject

Headings and [EUROVOC], a multi-lingual and multi-disciplinary thesaurus covering the activities of the EU and European Parliament.

References:

[AGROVOC]

AGROVOC vocabulary of the Food and Agriculture Organization of the United Nations

http://www.fao.org/agrovoc/

[AJAX]

Asynchronous Javascript and XML

See http://www.w3.org/TR/XMLHttpRequest/ (W3C Working Draft)

[Apple Siri]

http://www.apple.com/ios/siri/

[cMap]

The cMap project

http://www.cmap.eu

[CPV]

Common Procurement Vocabulary

http://simap.europa.eu/codes-and-nomenclatures/codes-cpv/codes-cpv_en.htm

[D2RQ]

The D2RQ Platform - Accessing Relational Databases as Virtual RDF Graphs

http://d2rq.org

[data.gov]

USA Government Open Data initiative

http://www.data.gov/

[data.gov.uk]

UK Government Open Data initiative

http://data.gov.uk/

http://data.gov.uk/linked-data/who-is-doing-what

[DBpedia]

http://dbpedia.org/

http://dbpedia.org/sparql

[Direct Mapping]

A Direct Mapping of Relational Data to RDF

http://www.w3.org/TR/rdb-direct-mapping/

[eCl@ss]

eCl@ss - Classification and Product Description

http://www.eclass.de

[EU Public Data]

Europe's Public Data

http://publicdata.eu/

[EU Core Business Vocabulary]

EU Core Business Vocabulary | Joinup

https://joinup.ec.europa.eu/asset/core_business/description

[EUROVOC]

EUROVOC - the multilingual thesaurus of the European Union

http://eurovoc.europa.eu

[FOAF]

Friend Of A Friend (FOAF) Vocabulary Specification

http://xmlns.com/foaf/spec/

[GeoNames]

http://www.geonames.org/

[GeoSPARQL]

GeoSPARQL - A Geographic Query Language for RDF Data | OGC(R)

http://www.opengeospatial.org/standards/geosparql

http://geosparql.org/

[GoodRelations]

GoodRelations: The Professional Web Vocabulary for E-Commerce

http://purl.org/goodrelations/

http://www.heppnetz.de/ontologies/goodrelations/v1

http://wiki.goodrelations-vocabulary.org/GoodRelations_and_schema.org

http://wiki.goodrelations-vocabulary.org/GoodRelations_for_Semantic_SEO

[Google Now]

http://www.google.com/landing/now/

[GRDDL]

GRDDL Primer

http://www.w3.org/TR/grddl-primer/

[GS1 B2C]

GS1 Business-to-consumer (B2C)

http://www.gs1.org/b2c

http://www.gs1.org/gsmp/kc/b2c

[GS1 CBV]

GS1 Core Business Vocabulary

http://www.gs1.org/gsmp/kc/epcglobal/cbv

[GS1 EPCIS]

GS1 EPC Information Services

http://www.gs1.org/gsmp/kc/epcglobal/epcis

[GS1 GDSN]

GS1 Global Data Synchronization Network

http://www.gs1.org/gsmp/kc/ecom/xml/gdsn_grid

[GS1 GPC]

GS1 Global Product Classification

http://www.gs1.org/gdsn/gpc/what

http://www.gs1.org/1/productssolutions/gdsn/gpc/browser/index.html

[GS1 Source]

http://www.gs1.org/source

[GS1 TSD]

GS1 Trusted Source of Data (TSD) 1.0 Standard

http://www.gs1.org/source/standards

[GS1 UK TrueSource]

http://www.gs1uk.org/what-we-do/truesource/Pages/default.aspx

[GS1 US DataDriver]

http://www.gs1us.org/resources/tools/data-driver

[HTML]

HyperText Markup Language

http://www.w3.org/MarkUp

[HTTP]

RFC 2616 - Hypertext Transfer Protocol -- HTTP/1.1

http://tools.ietf.org/html/rfc2616

See also http://www.w3.org/Protocols/

[IRI]

RFC 3987 - Internationalized Resource Identifiers (IRIs)

http://tools.ietf.org/html/rfc3987

[JSON]

JavaScript Object Notation

http://www.json.org

[JSON-LD]

A JSON-based Serialization for Linked Data

http://www.w3.org/TR/2014/REC-json-ld-20140116/

http://json-ld.org/

[Learning SPARQL]

Learning SPARQL

Bob DuCharme
O'Reilly Media, July 2011
ISBN: 978-1-4493-0659-5

http://www.learningsparql.com

[LinkedLifeData]

Linked Life Data - A Semantic Data Integration Platform for the Biomedical Domain

http://linkedlifedata.com/

http://linkedlifedata.com/sparql

[Linked Data]

Linked Data: Evolving the Web into a Global Data Space (1st edition)

Tom Heath and Christian Bizer (2011)
Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136.
Morgan & Claypool.
ISBN: 9781608454303 (paperback)
DOI: 10.2200/S00334ED1V01Y201102WBE001

http://linkeddatabook.com/editions/1.0/

[Linking Enterprise Data]

Linking Enterprise Data

David Wood (Editor)
Springer-Verlag, 1st Edition (2010)
ISBN: 978-1-4419-7664-2

http://3roundstones.com/linking-enterprise-data/

[LinkedModel]

Linked Models - OWL and SKOS specifications of Industry and Government Standards

http://www.linkedmodel.org/

[LOD Cloud]

The Linking Open Data cloud diagram

http://lod-cloud.net/

[Microdata]

HTML Microdata

http://www.w3.org/TR/microdata/

[Ontologies]

Common Ontologies for Semantic Web Domain Modeling - 3kbo

http://blog.3kbo.com/2008/10/11/common-ontologies-for-semantic-web-models/

Predefined Namespace Prefixes at SINDICE SPARQL interface

http://sparql.sindice.com/sparql?nsdecl

Namespace lookup for RDF developers

http://prefix.cc

[OWL]

OWL 2 Web Ontology Language Document Overview (Second Edition)

http://www.w3.org/TR/owl-overview/

OWL Web Ontology Language Reference

http://www.w3.org/TR/owl-ref/

[OWLIM]

OWLIM | Ontotext

http://www.ontotext.com/owlim

[QUDT]

Quantities, Units, Dimensions and Time ontology

http://www.qudt.org

[R2RML]

R2RML: RDB to RDF Mapping Language

http://www.w3.org/TR/r2rml/

[REST]

Architectural Styles and the Design of Network-based Software Architectures

Roy Thomas Fielding

PhD dissertation, University of California, Irvine

http://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf

[RDF]

Resource Description Framework

RDF - Semantic Web Standards

http://www.w3.org/RDF/

RDF 1.1 Concepts and Abstract Syntax

http://www.w3.org/TR/rdf11-concepts/

RDF Primer

http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

[RDFa]

RDFa Core 1.1

http://www.w3.org/TR/rdfa-syntax/

RDF annotations - RDFa 1.1 Primer

http://www.w3.org/TR/xhtml-rdfa-primer/

[RDFa Lite]

RDFa Lite 1.1

http://www.w3.org/TR/rdfa-lite/

[RDFS]

RDF Vocabulary Description Language 1.0: RDF Schema

http://www.w3.org/TR/rdf-schema/

[RDF/XML]

RDF/XML Syntax Specification (Revised)

http://www.w3.org/TR/rdf-syntax-grammar/

[RichSnippets]

Google Webmaster Tools - Rich Snippets Testing Tool

http://www.google.com/webmasters/tools/richsnippets

[schema.org]

http://schema.org/

http://schema.org/docs/schemas.html

[SKOS]

http://www.w3.org/TR/skos-reference/

[SKOS Primer]

http://www.w3.org/TR/skos-primer/

[SPARQL]

SPARQL Query Language for RDF

http://www.w3.org/TR/rdf-sparql-query/

[SPARQL 1.1]

SPARQL 1.1 Query Language for RDF

http://www.w3.org/TR/sparql11-query/

[StructuredData]

Linked Data – Structured data on the Web

David Wood, Marsha Zaidman, Luke Ruth, and Michael Hausenblas

Manning Publications Co., 1st Edition (December 2013)

ISBN: 9781617290398

http://www.manning.com/dwood/

[Turtle]

http://www.w3.org/TR/2013/CR-turtle-20130219/

[Unicode]

http://www.unicode.org

[UNSPSC]

United Nations Standard Products and Services Code

http://www.unspsc.org

[URIs]

Uniform Resource Identifiers

IETF RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax

http://tools.ietf.org/html/rfc3986

See also http://www.w3.org/Addressing/

URIs, URLs, and URNs: Clarifications and Recommendations 1.0

http://www.w3.org/TR/uri-clarification/

[URL]

Uniform Resource Locators

IETF RFC 1738 - Uniform Resource Locators

http://tools.ietf.org/html/rfc1738

[vCard]

Representing vCard objects in RDF

http://www.w3.org/TR/vcard-rdf/

[W3C SW]

W3C Semantic Web Activity Homepage

http://www.w3.org/2001/sw/

[W3C Reg Org Vocab]

W3C Registered Organization Vocabulary (W3C Working Draft)

http://www.w3.org/TR/vocab-regorg/

[Working Ontologist]

Semantic Web for the Working Ontologist

Dean Allemang & Jim Hendler

AUTO-ID LABS

Morgan Kaufmann (20 June 2008)
ISBN: 978-0123735560

http://workingontologist.org

[XML]

XML - Extensible Markup Language

http://www.w3.org/XML

[XPath]

http://www.w3.org/TR/xpath

[XQuery]

http://www.w3.org/XML/Query

[XSD]

W3C XML Schema

http://www.w3.org/XML/Schema

[XSLT]

W3C XML Transformations

http://www.w3.org/TR/xslt