



UNIVERSITY OF
CAMBRIDGE

Linked Data and Open Data

How can they help your business?

MARK HARRISON

Director, Auto-ID Lab, University of Cambridge

www.autoidlabs.org.uk/GS1Digital

6-8 November 2013

operational excellence

london conference

I'd like to thank the Consumer Goods Forum for this opportunity to tell you about Open Data and Linked Data and how they can help your business.

Later today we'll hear a talk about Big Data and how to extract meaningful insight from the huge amounts of data that your organization collects every day.

Outline

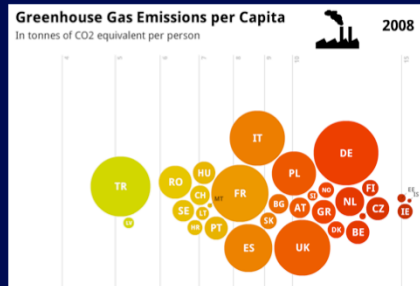
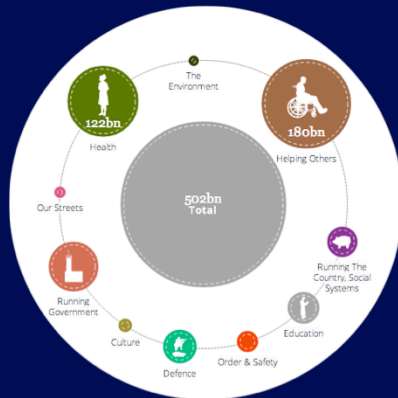
- Open Data sources
- Linked Data technology
- Improved Search Result Listings
- Enabling a product-centric ecosystem

operational excellence

london conference

Open Data and Linked Data are somewhat complementary to Big Data - but both can be very helpful to your company. I'm going to give you a brief overview about Open Data, then explain what is Linked Data and how you can use these two approaches together not only to improve your search engine results but also enable a new product-centric ecosystem of new services and applications that consumers can access, e.g. using their smartphones.

Open Data



Thursday	Friday
14°	13°



11	Ardrossan C &	Due
11	Ardrossan Cha	16:35
11	Ardrossan Cha	16:39
12	Ardrossan Sta	16:42
11	Ardrossan C &	12min
11	Ardrossan C &	19min



operational excellence

london conference

Many governments and other organisations are making their datasets freely available for others to re-use.

Examples include data about the economy and government spending - by department and by region, socio-economic data and aggregated census data, crime statistics, geographic datasets, weather forecasts, even information about planned roadworks and diversions as well as real-time traffic information, updated every 10 minutes.

Finding Open Data



Government spending
Environmental data
Public Transport
Road traffic
Roadworks, Congestion
Accidents
Population / demographics
Local socio-economic data
Economy / Industry
Health
Geographic / Postcode

operational excellence

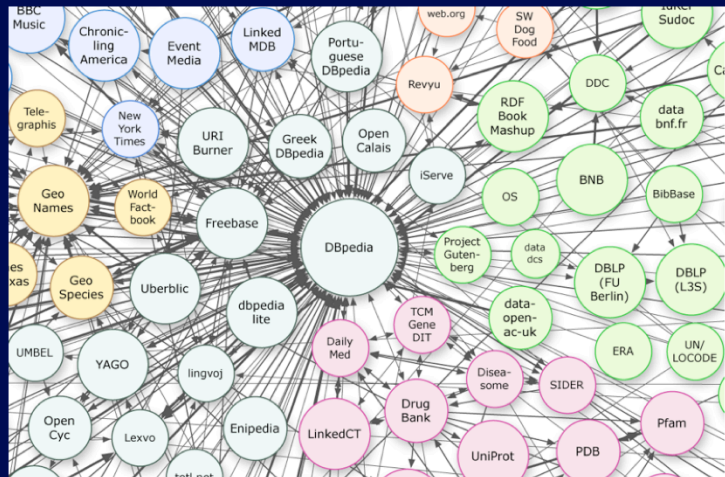
london conference

All this open data is already available to your business right now, absolutely free of charge - you can find it at websites such as data.gov , data.gov.uk , publicdata.eu and you can also find some great examples of data mash-ups that let you visualise the data.

Linked Open Data



London	
	
Area	
• City	1,572.00 km ² (606.95 sq mi)
• Urban	1,737.9 km ² (671.0 sq mi)
• Metro	8,382.00 km ² (3,236.31 sq mi)
Elevation ^[1]	24 m (79 ft)
Population (2012) ^{[2][3][4]}	
• City	8,308,369
• Density	5,285/km ² (13,690/sq mi)
• Urban	9,787,426
• Urban zone	11,805,500
• Metro	15,010,295

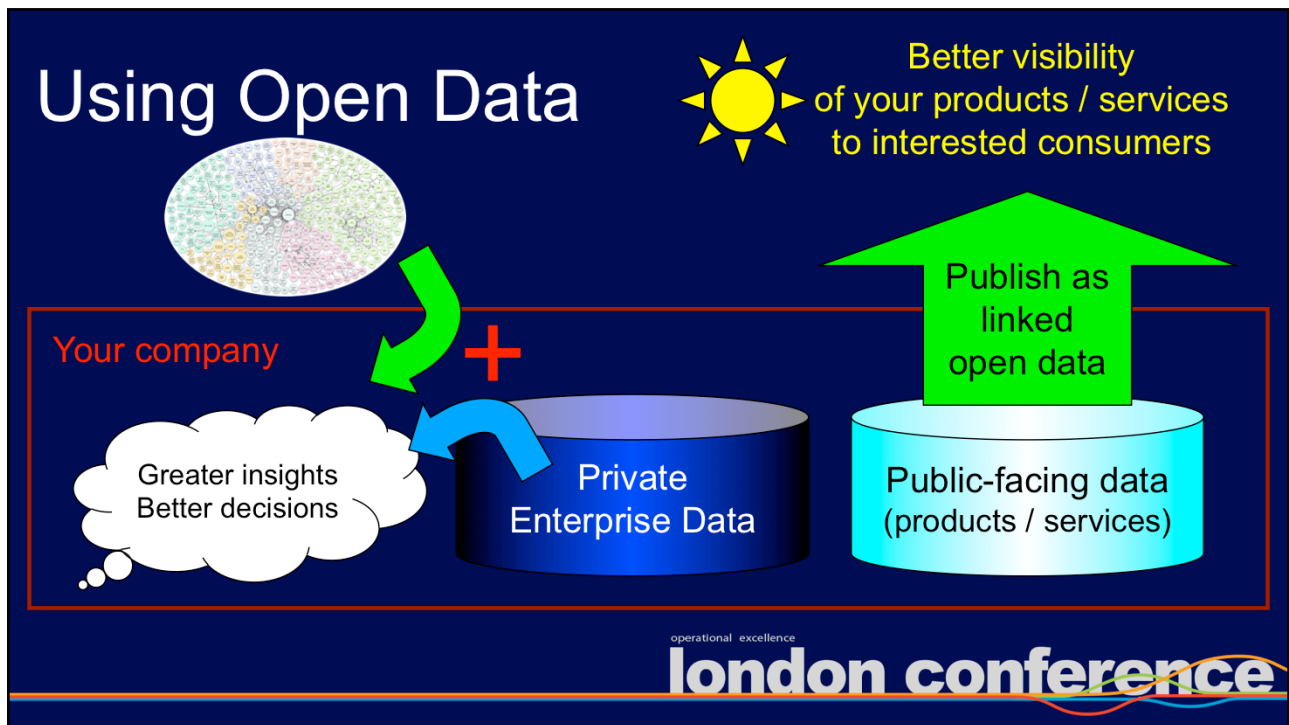


operational excellence

london conference

Here is a picture of the Linked Open Data cloud diagram. It represents an increasing collection of open datasets in various sectors, including geographic information, government data, information feeds from the media, cultural information about music and films, as well as open data in life-sciences where Open Data combined with Linked Data technology is transforming research into new cures for diseases by connecting together in a logical way the separate pieces of new discoveries reported by biochemical researchers around the world. You can try it out at LinkedLifeData.com/refinder

At the centre of the Linked Open Data cloud is DBpedia. We've all used Wikipedia, the multi-lingual global crowd-sourced encyclopaedia that any of us can read and update. Well, all those info-boxes of facts and figures that appear in Wikipedia have been converted into data that is available via DBpedia - and there are so many cross-references between DBpedia and other open datasets, meaning that the collections of facts contained in DBpedia are linked to other collections of facts found elsewhere.

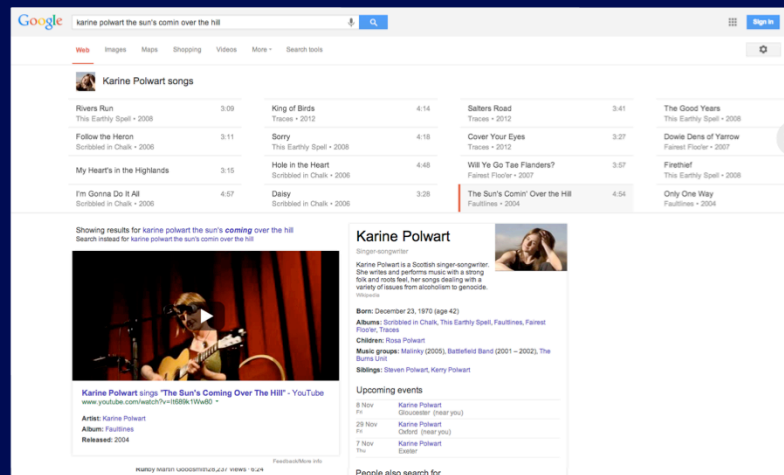


Your company has some information that is private and for internal use - and some that is intended to be open - for example the information you use to advertise the products and services that you offer and their specifications.

You can use public open data sets to enhance your company's own private internal data to give it more context. For example, your company might have a database of customers or suppliers, together with their postcodes. There are datasets like Code-Point Open (from the UK Ordnance Survey) that make the link from a postcode to geographic co-ordinates (latitude, longitude) - so you can then visualise your lists of customers or suppliers as geographic clusters on a map. That could be very useful if your company is planning where to open a new retail store or a new distribution centre. You could even mix in some socio-economic data or demographic data for that local area to help plan which range of goods or services are most appropriate to offer at that new store - or when the new store opens, use open weather forecast data to plan whether you need to increase stocks of extra barbecue equipment or snow shovels for the next weekend.

As well as consuming existing open data, your business can use it to improve the visibility of the products and services it offers - i.e. to improve your search listing results.

Search is changing!



operational excellence

london conference

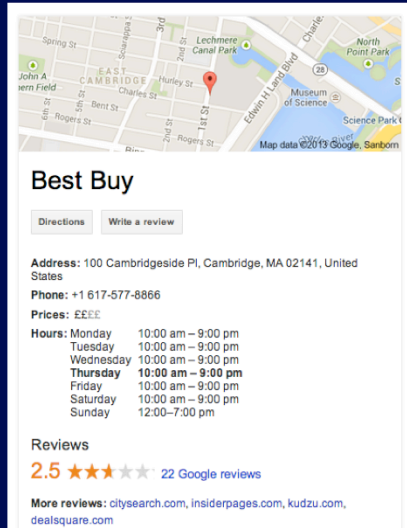
Remember when search results looked like this? (basic search results shown as a list of websites matching a keyword)

Well, they don't any more. Search engines are getting much smarter about understanding the meaning of words and recognising Named Entities. Named Entities include concepts, cities, musicians and bands, movie actors - but also companies, brand names and products.

Some of you will have seen search results like this (shows basic Google Rich Snippet on the right-hand side, with map, store location, opening hours) - and you maybe wondered how that other company got the special treatment - did they pay Google for an ad? Actually, they didn't. They got it completely free, just by making some of their public-facing data available openly in a more structured way that is easier for search engines or smartphone apps to extract.

It doesn't stop there. Just a couple of weeks ago, I noticed that Google are still developing their Rich Snippets even further - so now if you search for songs by a particular musician (example of Google search results for Karine Polwart songs) you will see the search results page being transformed into a full page dashboard of relevant information about the musician and her songs - every song title is a button.

Companies already find benefits!



30% increase in web traffic
to store pages since
introducing Linked Data

operational excellence

london conference

Now imagine that your team could do the same for your company's products and services. Using Linked Data technology you can – and some companies have already started doing so.

In the USA, BestBuy - a major retailer of consumer electronics used Linked Data markup in their website and saw a 30% increase in click-through rates to their store pages.

Linked Data Technology



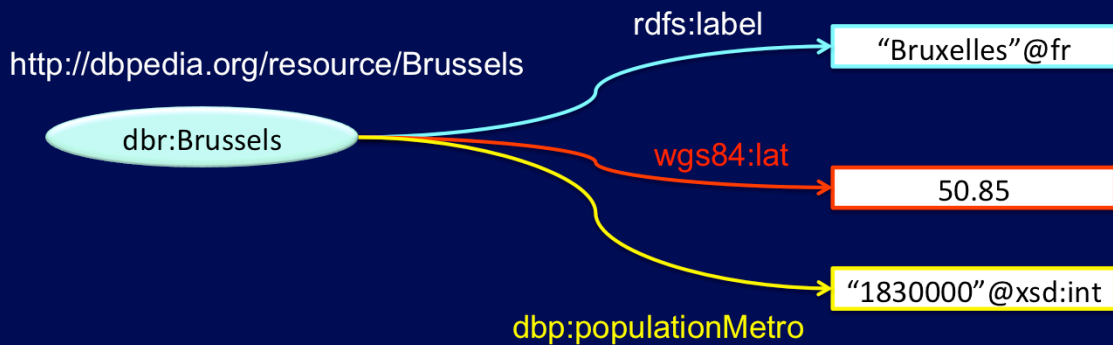
operational excellence

london conference

Linked Data technology (also known as the Semantic Web) is about transforming the Web from a web of documents into a web of facts – to form a massive global distributed database of facts that we can all use. That doesn't mean your company has to publish all its data - just the parts that it wants to expose. Just like having a company website doesn't mean you publish your internal production data and sales data. Your company keeps control of what it publishes - but you can use Linked Data technology as the way to make it available in a structured way so that search engines and mobile apps can extract the facts and figures easily and understand the meaning.

Linked Data technology is based on open standards from the World Wide Web Consortium (W3C) - the folks who gave us XML and HTML. Well, they developed another data format called RDF (Resource Description Framework). RDF serves as a Lowest Common Denominator format for representing factual claims as very simple logical sentences.

Using Linked Data (RDF) to write facts



operational excellence

london conference

Many of us are familiar with mind-maps for representing the relationships between things - well a mind-map is really just a graph of data and RDF is really just a way of collapsing a complicated mind-map into a set of simple logical sentences, each consisting of a Subject, Property and a Value.

Here are some examples of facts expressed as RDF triples:

The French name for Brussels is Bruxelles.

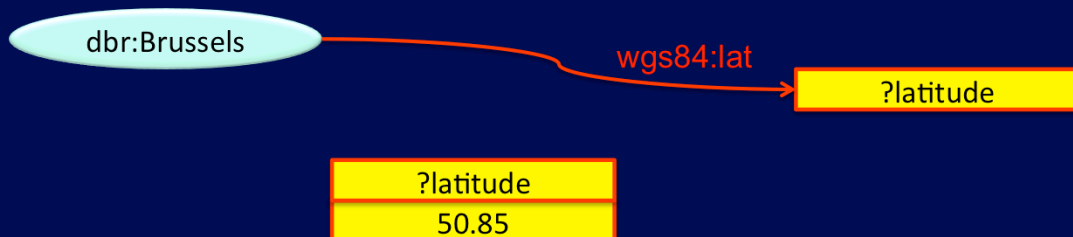
Brussels lies at a latitude of 50.85 degrees North

Brussels has a metropolitan population of 1.83 Million

All of this is machine-readable and available from DBpedia.

Using SPARQL to query facts

```
SELECT ?latitude WHERE { dbr:Brussels wgs84:lat ?latitude . }
```



operational excellence

london conference

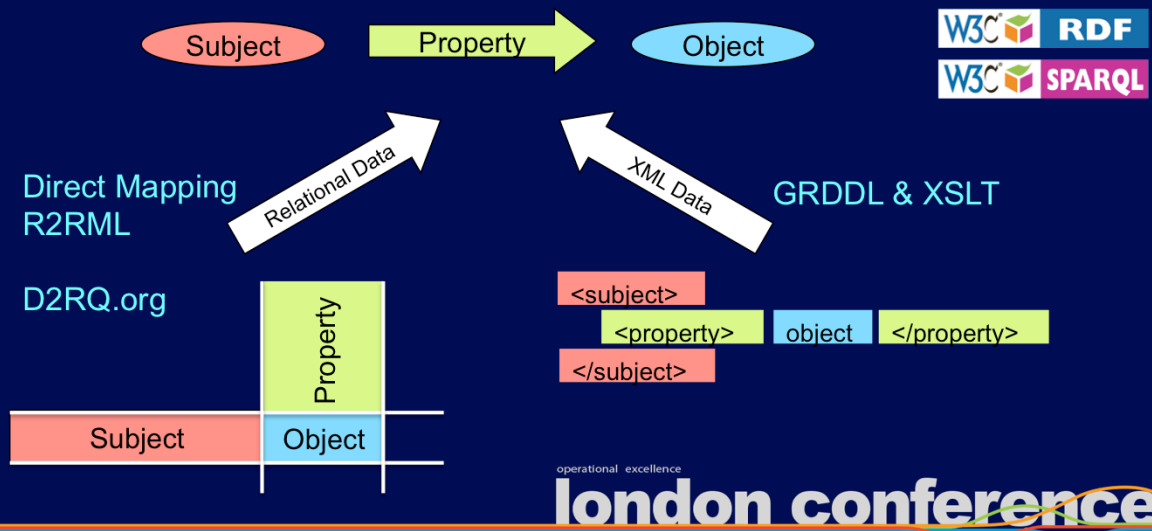
There's also a query language called SPARQL, which provides a way to look for patterns in a graph of data and extract the values where the pattern matches the data. You can even use SPARQL to combine data from multiple sources, combining some of that public open data with your company's internal data - and even using it to define rules for creating new facts that can be inferred from the combination of existing facts – that is to say that you can use it to do some reasoning.

Here's a very simple example of a SPARQL query:

```
SELECT ?latitude WHERE { dbpedia:Brussels geo:lat ?latitude . }
```

Answer: 50.85

Your existing data as Linked Data



But I suspect you're now asking yourselves how this works with your existing data - because that's in XML and relational databases etc. Well, there are standards and toolkits to help you access your existing data as RDF or query it using SPARQL - and you can even use Linked Data technology purely internally to make it much easier to do data integration across different systems.

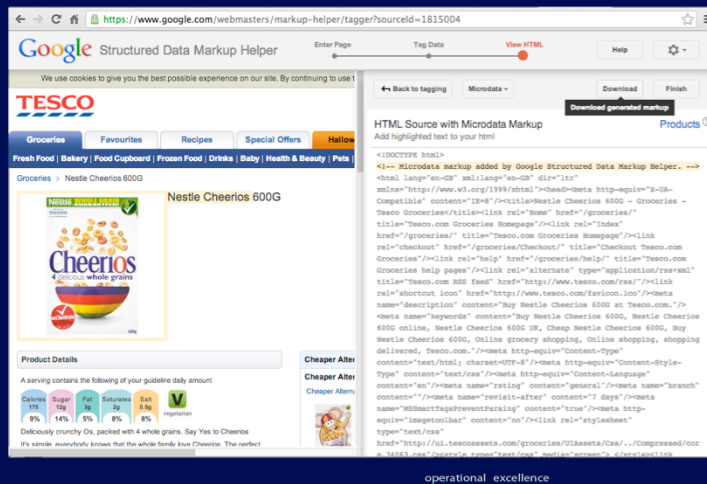
Linked Data vocabularies



So, how do we use Linked Data to make some Google Rich Snippets for your business, to advertise its products and services? Well, your technical team just puts some extra markup inside your company's web pages to tell the search engines where to find specific information and what it means.

Your technical team can use various terms from Linked Data vocabularies such as schema.org and GoodRelations to describe your company and the products and services it offers.

Adding Linked Data to your web pages

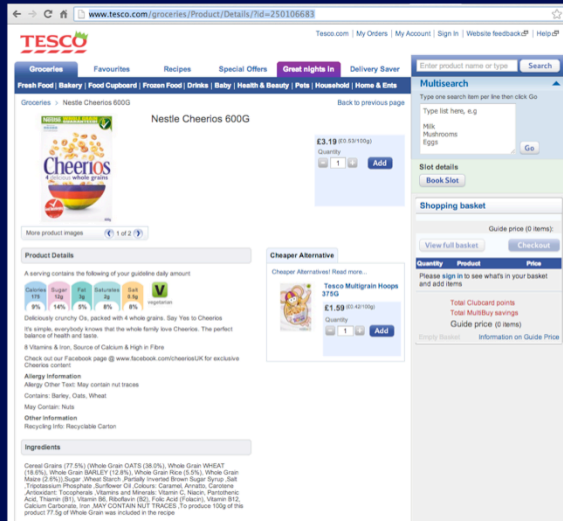


operational excellence

london conference

To get started, they can try tools like the Google Structured Data Markup Tool, which lets them point and click to add annotations to your web page - and it generates the extra markup for your web page and exports that back modified HTML

Example: web page with Linked Data



Original web page has no Linked Data

Our modified version has...

<http://tinyurl.com/rdfa-test>

operational excellence

london conference

Here's an example of applying Linked Data markup to product pages. We took a Tesco web page for a Nestle product and used tools such as the Google Structured Data Markup Tool to add some basic semantic annotations to the web page. We then added some further markup for the nutritional information, the product category and some further attributes and values.

Example: web page with Linked Data



In a web browser, the original web page and the page with the extra markup look identical. However, when we view them with tools that extract the structured data, we see a big difference. For the original web page, it figured out that it's a web page - no surprise there. For the modified page with the extra markup, it has extracted all kinds of structured information about the product, its ingredients, nutritional information and the price offer.

Enabling new consumer interactions



Organic?	☒
Skimmed	☐
Half Fat	☒
Full Fat	☐
Max. Price	
How Urgent?	1hr
How Far?	1km

operational excellence

london conference

That information is then available to search engines and smartphone apps so that people can find the product they are really looking for. We can use Linked Data with product classification systems such as GS1 GPC (Global Product Classification) to enable contextual search and refinement by attributes and values.

A customer could scan a product and be informed if it contains a particular allergen or an ingredient that is incompatible with their dietary preferences - and advise them of an alternative product that is a better match for them.

Enabling new consumer interactions



URL (user manual)

URL (user forum)

URL (product support)

URL (firmware updates)

URL (driver software)

URL (warranty registration)

operational excellence

london conference

Of course we can also take into account ethical and environmental accreditation of products, ratings and reviews - and for post-sale services, we could link to instruction manuals and recipes, a user forum for feedback or troubleshooting, warranty registration or spare parts and refill cartridges.

Furthermore, descriptions, product categories and attributes can be provided using multi-lingual labels, so that even if consumers are travelling in a country where they doesn't speak the local language, their smartphone app can help them find the kind of product that they are looking for.



GS1 Digital initiative: Get Involved!

Using technology such as Linked Open Data
to make trustworthy data about products and services
available on the web

To improve visibility of products and product offerings

To enable enhanced consumer interactivity with products and
support new services and applications

Contact: robert.beideman@gs1.org

operational excellence

london conference

GS1 has recently launched its GS1 Digital initiative and the GTIN+ on the Web project, focusing on ways to help companies use Linked Data technology to make their products and services much more discoverable on the web. We're setting up some pilot activities right now, so if you are interested, please contact my colleague Robert Beideman - or find him at the GS1 Digital booth during the break. Ultimately we're trying to use Linked Open Data technology to enable a whole new ecosystem for product-centric services.

There's much more detail I could tell you if I had more time here - but we're making an extended version of this presentation available at <http://www.autoidlabs.org.uk/GS1Digital>, together with a couple of online examples that your technical team can try - and some links to additional resources they might find useful.

Summary

- Use Open Data sources to enrich your internal data and to inform strategic and tactical decisions
- Linked Data technology is transforming the Web into a massive globally distributed database of facts
- Search is becoming smarter and more semantic
Start reaping the benefits now and don't be left behind!

More info: www.autoidlabs.org.uk/GS1Digital

operational excellence

london conference

So, to summarise what I've told you so far:

- 1) There are many source of open data that you can already start using to enhance your existing internal data and to help you make better decisions, both strategic and tactical/operational
- 2) Linked Data technology transforms the Web into a massive globally distributed database of facts - and provides tools for searching across those facts.
- 3) Search engines are getting much smarter about understanding the meaning of words - and they are providing enhanced search listings for organisations that include Linked Data markup in their web pages. Your company can get involved in the GS1 Digital pilot activities to learn more and gain first-hand experience of the benefits.