



UNIVERSITY OF
CAMBRIDGE

Linked Data and Open Data

How can they help your business?

MARK HARRISON

Director, Auto-ID Lab, University of Cambridge

6-8 November 2013

operational excellence

london conference

Outline

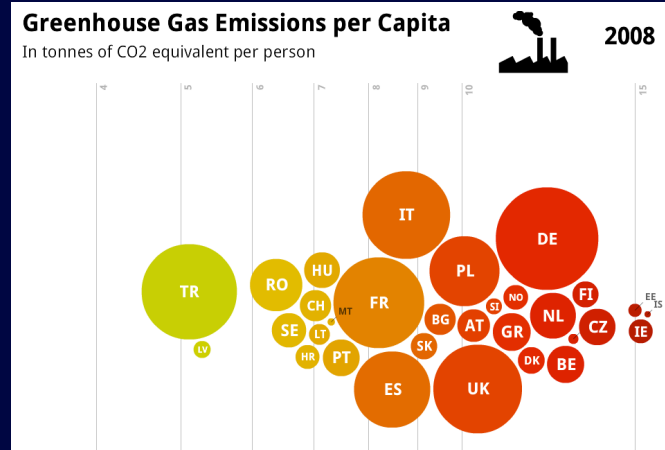
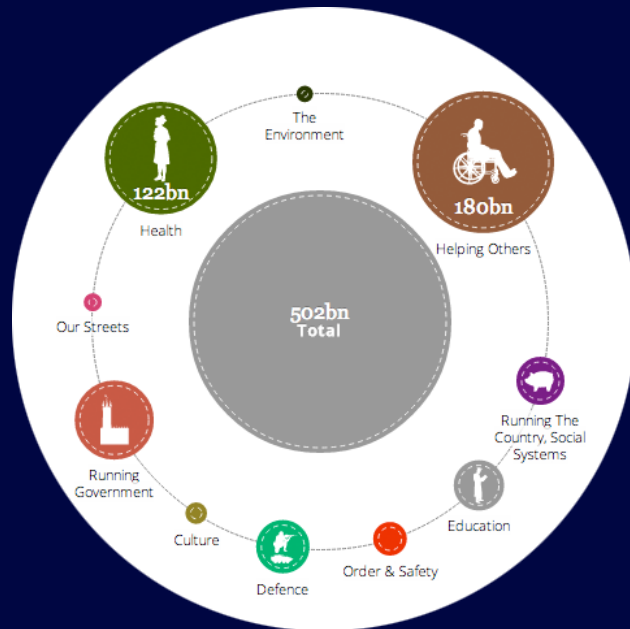
- Open Data sources
- Linked Data technology
- Improved Search Result Listings
- Enabling a product-centric ecosystem

operational excellence

london conference



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

Open Data relevant for logistics



Live Traffic Information:

Current Roadworks	10 mins	1Mb
Current Planned Events	10 mins	0.01Mb
Future Roadworks	6 hours	2Mb
Future Planned Events	6 hours	0.2Mb
Unplanned Events	10 mins	0.05Mb
All traffic data	10 mins	30Mb
Journey Time Data	10 mins	0.5Mb
Variable Message Signs	10 mins	3Mb
Matrix Signals	10 mins	9Mb

<http://data.gov.uk/publisher/highways-agency>

Data is provided in XML Datex2 format (<http://www.datex2.eu>) - not yet as linked data

Open Data for weather



Weather forecast for London (as web page) [http://www.yr.no/place/](http://www.yr.no/place/United_Kingdom/England/London/)

[United_Kingdom/England/London/](http://www.yr.no/place/United_Kingdom/England/London/)

http://www.yr.no/place/United_Kingdom/England/London/hour_by_hour

http://www.yr.no/place/United_Kingdom/England/London/long

Weather forecast for London (as XML data)

http://www.yr.no/place/United_Kingdom/England/London/forecast.xml

http://www.yr.no/place/United_Kingdom/England/London/forecast_hour_by_hour.xml

Provided by the Norwegian Meteorological Institute and the Norwegian Broadcasting Corporation (NRK)

Very good fine-grained coverage worldwide.

Javascript and PHP script examples are also provided.

Free yr.no mobile apps also available for iOS and Android.

<http://om.yr.no/verdata/free-weather-data/>

operational excellence

london conference

Open Data for UK postcodes

Code-Point® Open (<http://data.gov.uk//dataset/code-point-open>)

postcode data downloads freely available upon e-mail registration
(reply e-mail provides links to text file (CSV) data with the following

PostCode, OS Easting, OS Northing, (6 figures, 1m resolution)

(Online conversion to latitude, longitude available at:

<http://www.bgs.ac.uk/data/webservices/convertForm.cfm>)

Ordnance Survey data and Talis SPARQL endpoint

RDF data per postcode, e.g. <http://data.ordnancesurvey.co.uk/doc/postcodeunit/CB30FS.ttl>

<http://api.talis.com/stores/ordnance-survey/services/sparql>

operational excellence

london conference



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

5 stars of Open Data – Sir Tim Berners-Lee



Data available on the web in any format, using an **open licence** (PDDL, ODC-by, CC0)



Availability as **structured data**(e.g. spreadsheet vs bitmap image of chart)



Use non-proprietary formats (e.g. CSV vs .xls)



Use URIs to denote things, so other people can point to your data



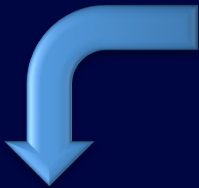
Link your data to other data to provide context

<http://5stardata.info>

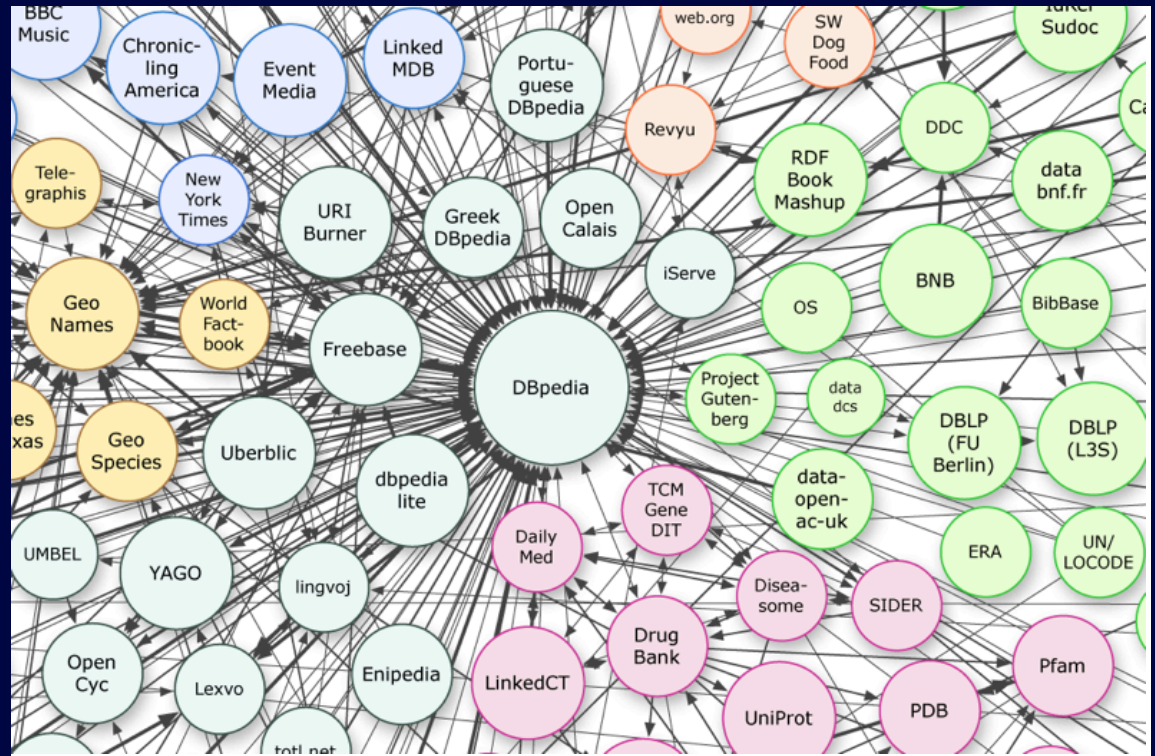
operational excellence

london conference

Linked Open Data



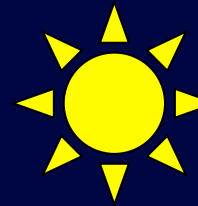
London	
	
Area	
• City	1,572.00 km ² (606.95 sq mi)
• Urban	1,737.9 km ² (671.0 sq mi)
• Metro	8,382.00 km ² (3,236.31 sq mi)
Elevation ^[1]	24 m (79 ft)
Population (2012) ^{[2][3][4]}	
• City	8,308,369
• Density	5,285/km ² (13,690/sq mi)
• Urban	9,787,426
• Urban zone	11,905,500
• Metro	15,010,295



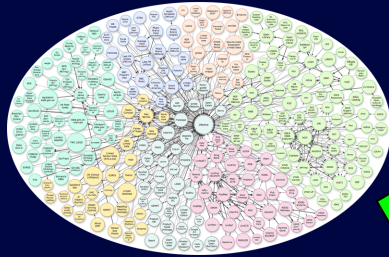
operational excellence

london conference

Using Open Data



Better visibility
of your products / services
to interested consumers



Your company



Greater insights
Better decisions

+

Private
Enterprise Data

Publish as
linked
open data

Public-facing data
(products / services)

operational excellence

london conference

Search is changing!

Google karine polwart the sun's comin over the hill

Web Images Maps Shopping Videos More Search tools

Karine Polwart songs

Rivers Run This Earthly Spell • 2008	3:09	King of Birds Traces • 2012	4:14	Salters Road Traces • 2012	3:41	The Good Years This Earthly Spell • 2008
Follow the Heron Scribbled in Chalk • 2006	3:11	Sorry This Earthly Spell • 2008	4:18	Cover Your Eyes Traces • 2012	3:27	Dowie Dens of Yarrow Fairrest Floo'er • 2007
My Heart's in the Highlands	3:15	Hole in the Heart Scribbled in Chalk • 2006	4:48	Will Ye Go Tae Flanders? Fairrest Floo'er • 2007	3:57	Firethief This Earthly Spell • 2008
I'm Gonna Do It All Scribbled in Chalk • 2006	4:57	Daisy Scribbled in Chalk • 2006	3:28	The Sun's Comin' Over the Hill Faultlines • 2004	4:54	Only One Way Faultlines • 2004

Showing results for karine polwart the sun's **coming over the hill**
Search instead for karine polwart the sun's comin over the hill

Karine Polwart
Singer-songwriter

Karine Polwart is a Scottish singer-songwriter. She writes and performs music with a strong folk and roots feel, her songs dealing with a variety of issues from alcoholism to genocide.

Wikipedia

Born: December 23, 1970 (age 42)
Albums: Scribbled in Chalk, This Earthly Spell, Faultlines, Fairrest Floo'er, Traces
Children: Rosa Polwart
Music groups: Malinky (2005), Battlefield Band (2001 – 2002), The Burns Unit
Siblings: Steven Polwart, Kerry Polwart

Upcoming events

8 Nov Fri	Karine Polwart Gloucester (near you)
29 Nov Fri	Karine Polwart Oxford (near you)
7 Nov Thu	Karine Polwart Exeter

Karine Polwart sings "The Sun's Coming Over The Hill" - YouTube
www.youtube.com/watch?v=It689k1Ww80

Artist: Karine Polwart
Album: Faultlines
Released: 2004

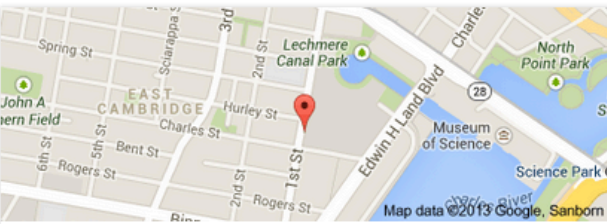
Feedback/More info

People also search for

operational excellence

london conference

Companies already find benefits!



Best Buy

[Directions](#) [Write a review](#)

Address: 100 Cambridgeside Pl, Cambridge, MA 02141, United States

Phone: +1 617-577-8866

Prices: ££££

Hours:

Monday	10:00 am – 9:00 pm
Tuesday	10:00 am – 9:00 pm
Wednesday	10:00 am – 9:00 pm
Thursday	10:00 am – 9:00 pm
Friday	10:00 am – 9:00 pm
Saturday	10:00 am – 9:00 pm
Sunday	12:00–7:00 pm

Reviews

2.5 ★★★★★ 22 Google reviews

More reviews: [citysearch.com](#), [insiderpages.com](#), [kudzu.com](#), [dealsquare.com](#)



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

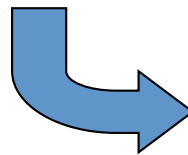
operational excellence

london conference

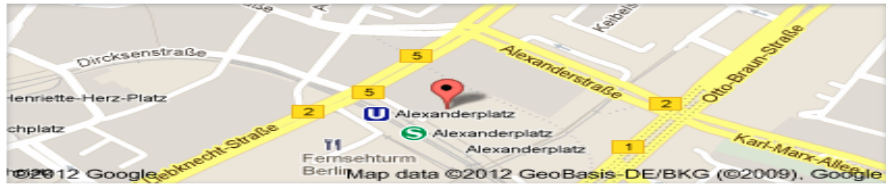
Why is this important now?

- Web search engines are making use of semantic markup, especially for helping consumers to find products and services
- Using semantic markup makes it easier for search engines to index content accurately and websites that use semantic markup are being rewarded with better search engine rankings as well as more prominent enhanced presentation in web search results, e.g. Google Rich Snippets

[GALERIA Kaufhof - Filialfinder - Berlin Alexanderplatz - Aktuelles](#)
[www.galeria-kaufhof.de/.../berlin-alexanderplatz/](#) - Translate this page
Öffnungszeiten. Zu diesen Zeiten können Sie uns in unserer Filiale GALERIA Kaufhof Berlin Alexanderplatz besuchen. Mo.-Mi.:09:30 - 20:00 Uhr Do.-Sa.:09:30 ...



Galeria Kaufhof Berlin Alexanderplatz [Directions](#)



©2012 Google Map data ©2012 GeoBasis-DE/BKG (©2009), Google



SCORE OVERALL **18** [Sign in for full scores and summary](#)

42 Google reviews

Address: Alexanderplatz 9, 10178 Berlin, Germany
Phone: 030 247430
Hours: Mon-Wed 9:30am-8pm
Thu-Sat 9:30am-10pm
Sun Closed
Transit: [Berlin Alexanderplatz](#)
More reviews: [qype.com](#), [prinz.de](#), [cityguide.com](#), [meinstadt.de](#)

[Feedback](#)

How BestBuy use Schema.org & GoodRelations

BestBuy's store pagese.g. <http://stores.bestbuy.com/1033>

Uses RDFa 1.0 (RDF annotations) and ontologies: **GoodRelations**, **vCard**, **FOAF**, **Geo (WGS84)**, to embed: **Latitude**, **Longitude**, **Street Address**, **City**, **State**, **Postcode**, **Phone No**, **Opening Hours**, link to each store's Facebook page

BestBuy's product-specific pages Use HTML5 microdata (itemscope, itemtype, itemprop etc.) and **schema.org** ontology to embed:

Link to product image, **Availability**, **Price**, **Manufacturer**,
Mfr Model Number, **Retailer's SKU**, **Description**,
Aggregate Rating (1-5), **Review Count**



Best Buy saw a **30%** increase in traffic to individual store websites

http://semanticweb.com/semtechbiz-keynote-jay-myers-discusses-linked-data-at-best-buy_b29622

operational excellence

london conference

Benefits of publishing linked data

Semantic SEO

If you want to improve your search engine ranking and prevent the search engines from indexing your site, you need to publish your data in a structured way.

Automotive

Many car manufacturers can now provide their customers with a personalized experience by publishing their data in a structured way.

Real Estate

Public real estate agencies can provide their clients with a personalized experience by publishing their data in a structured way.

Restaurants

Send your open specifications to mobile browsers and improve your online presence.

Tickets

Search for flights and book your tickets online.

Mobile commerce

Send your product information directly to your smartphone.

In short: Use rich data for Web marketing

The Web makes your offers globally visible. With GoodRelations, you can add an unprecedented level of precision to your communication with potential customers over the Web.

Sales DataBase

TimeKey
CustomerKey
ShipperKey
ProductKey
EmployeeKey
ItemFreight
ItemQuantity
ItemAll
Discount

Address

City
Region
PCode
Country
Phone
Mphone
Fax
Email

Product

ProductID
CategoryID

Time

Source:



Linked Data Technology



operational excellence

london conference

What is the Semantic Web / Linked Data?

- **World Wide Web**

- a global network of **linked documents** (web pages), primarily intended for human consumption (reading, understanding)
- information-rich but almost no machine-readable meaning of content
- HTML originally focused on presentation of information content for display within web browsers
- **Relies on human beings to read and understand**, then follow links or search

- **Semantic Web / Linked Data**

- builds on web technologies to achieve a global network of **linked data** at web scale
- enables unified **federated queries** of data **across multiple distributed data sources**
- can **ease data integration across different types of databases**
- enables **automated logical deductions** using this data (additional inferred information)
- supports the use of multiple distributed datasets and multiple ontologies (data dictionaries + logic) within queries

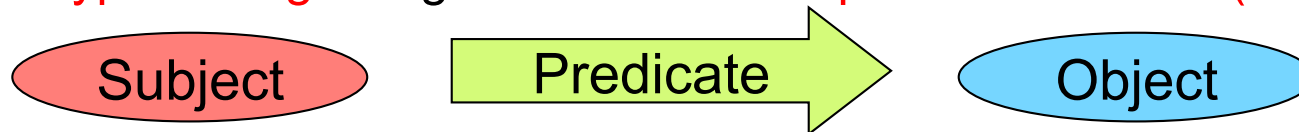
operational excellence

london conference



Linked Data essentials

- Identify all kinds of things (resources) - not only web pages but also places, people, products, companies, *also relationships* using Uniform Resource Identifiers (URIs)
- Transform a web of documents to a web of 'facts'
Publish data 'facts' about things identified by URIs
Fact-level hyperlinking using Resource Description Framework (RDF)



- Combine data 'facts' from multiple sources to enrich our understanding of the data we have - and find what we're really looking for!
 - using a query language for RDF data, (SPARQL)

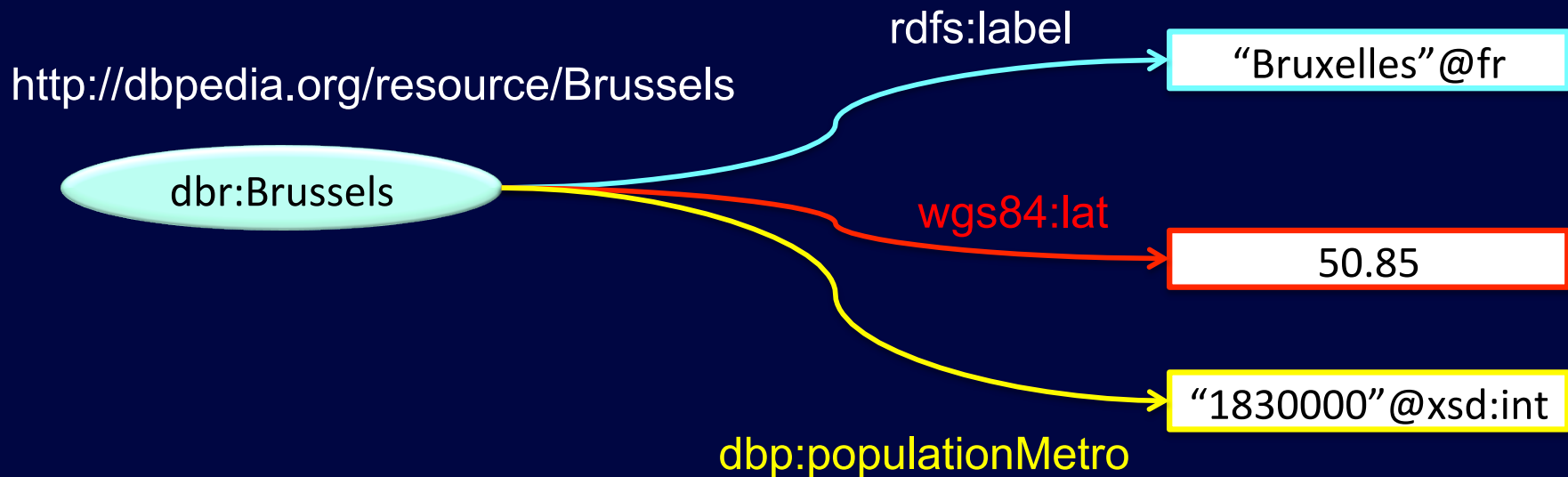
operational excellence

london conference

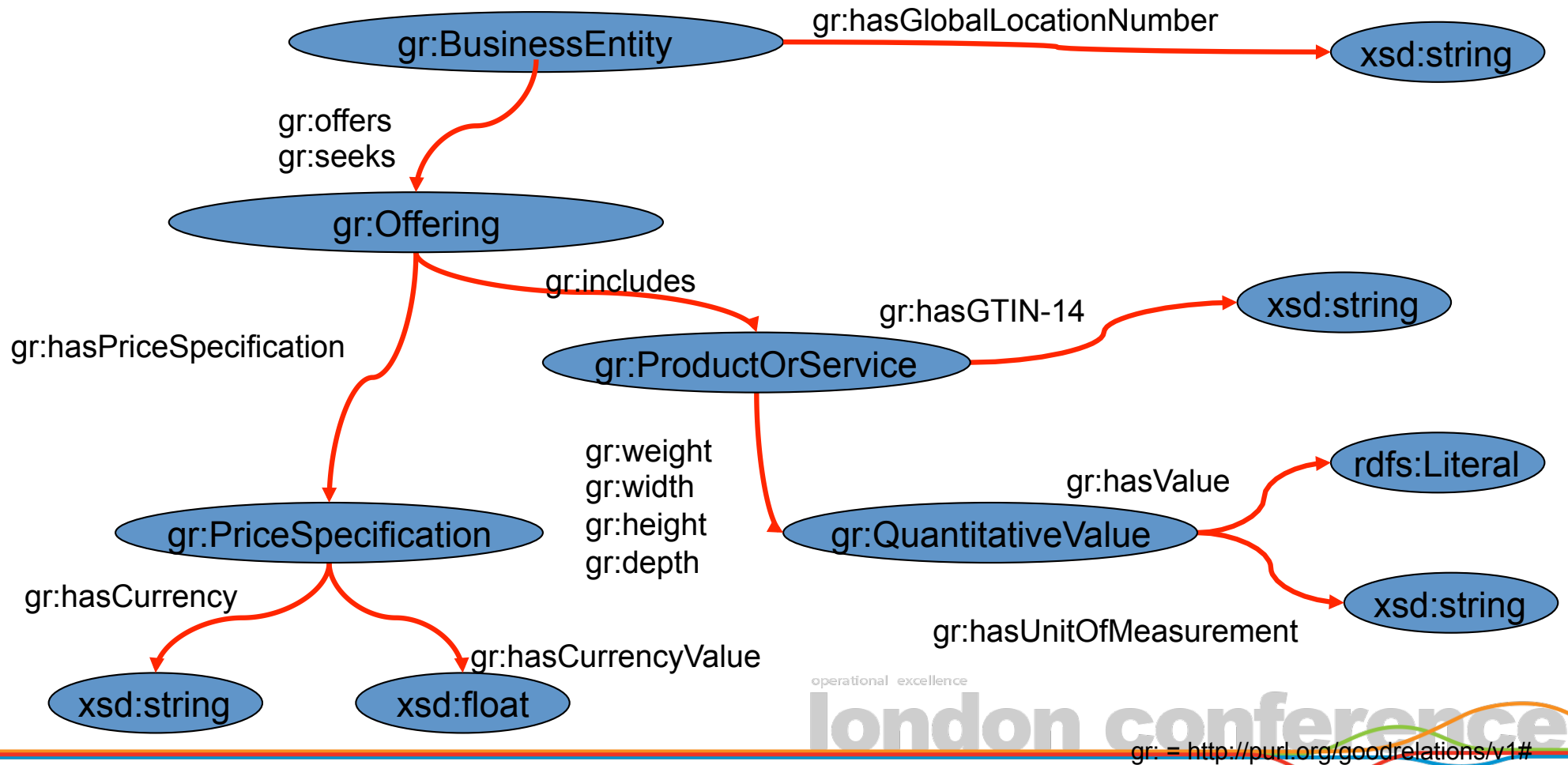
Core Linked Data Technologies

- **Uniform Resource Identifiers (URIs)** used to identify not only documents but also concepts (people, places, things, abstract/intangible concepts) and properties / data relationships
- **Resource Description Framework (RDF)** provides a W3C standard way to write simple logical statements about relationships.
- **Ontologies** are like data dictionaries with additional logical annotations (to say how properties and resources are related)
Multiple ontologies (for different domains) can co-exist and be used in parallel.
It's also easy to cross-reference between them.
- **SPARQL** query language enables a query to combine machine-readable data from multiple sources and also allows new data relationships to be constructed (*inferred*) from existing data.

Using Linked Data (RDF) to write facts



RDF examples using GoodRelations ontology



Linked Data & Inferences – RDFS & OWL

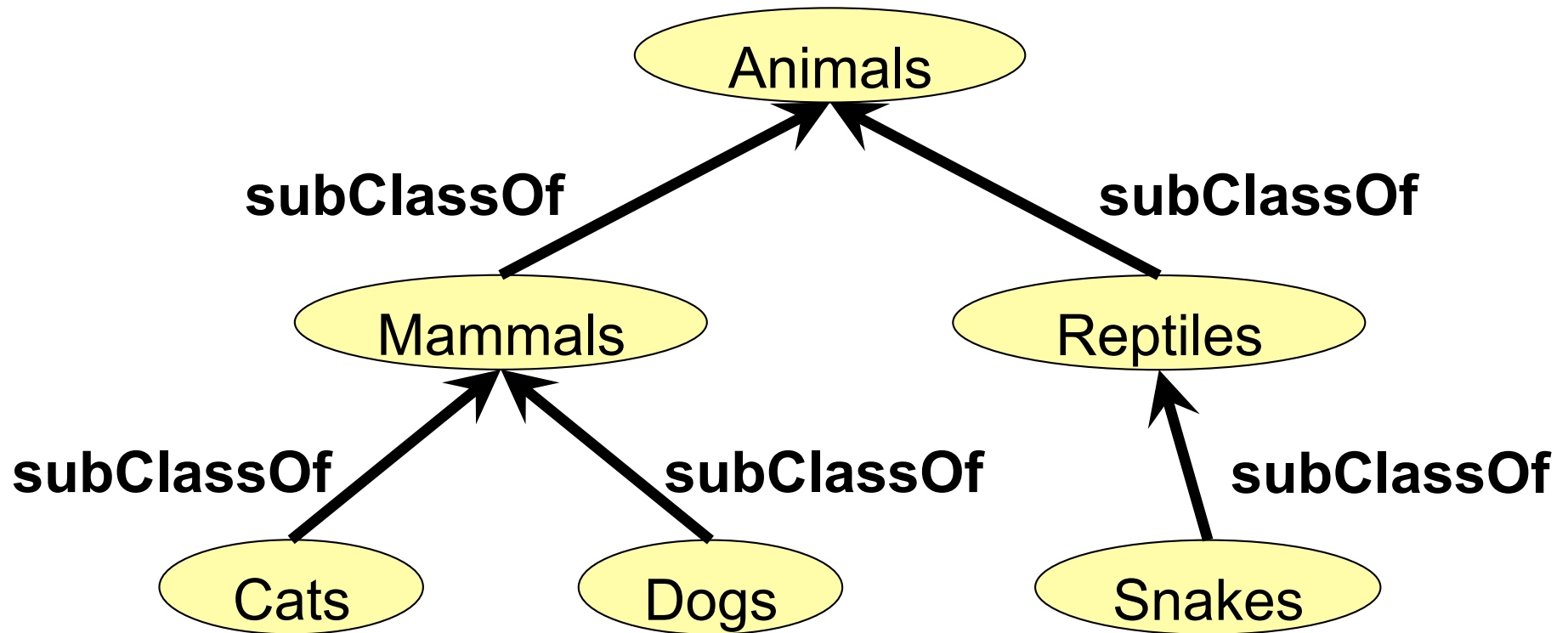
- Going beyond simply using RDF to state facts, we can use RDF Schema (RDFS) and Web Ontology Language (OWL) to represent knowledge and data models and relationship constraints / inferences
 - Classes and Sub-Classes
 - Disjoint Classes, Unions, Intersections
 - Properties and Sub-Properties
 - Inverse Properties, Symmetric Properties
 - Transitive Properties, Functional Properties
 - Domain and Range constraints on properties
- and much more...

operational excellence

london conference



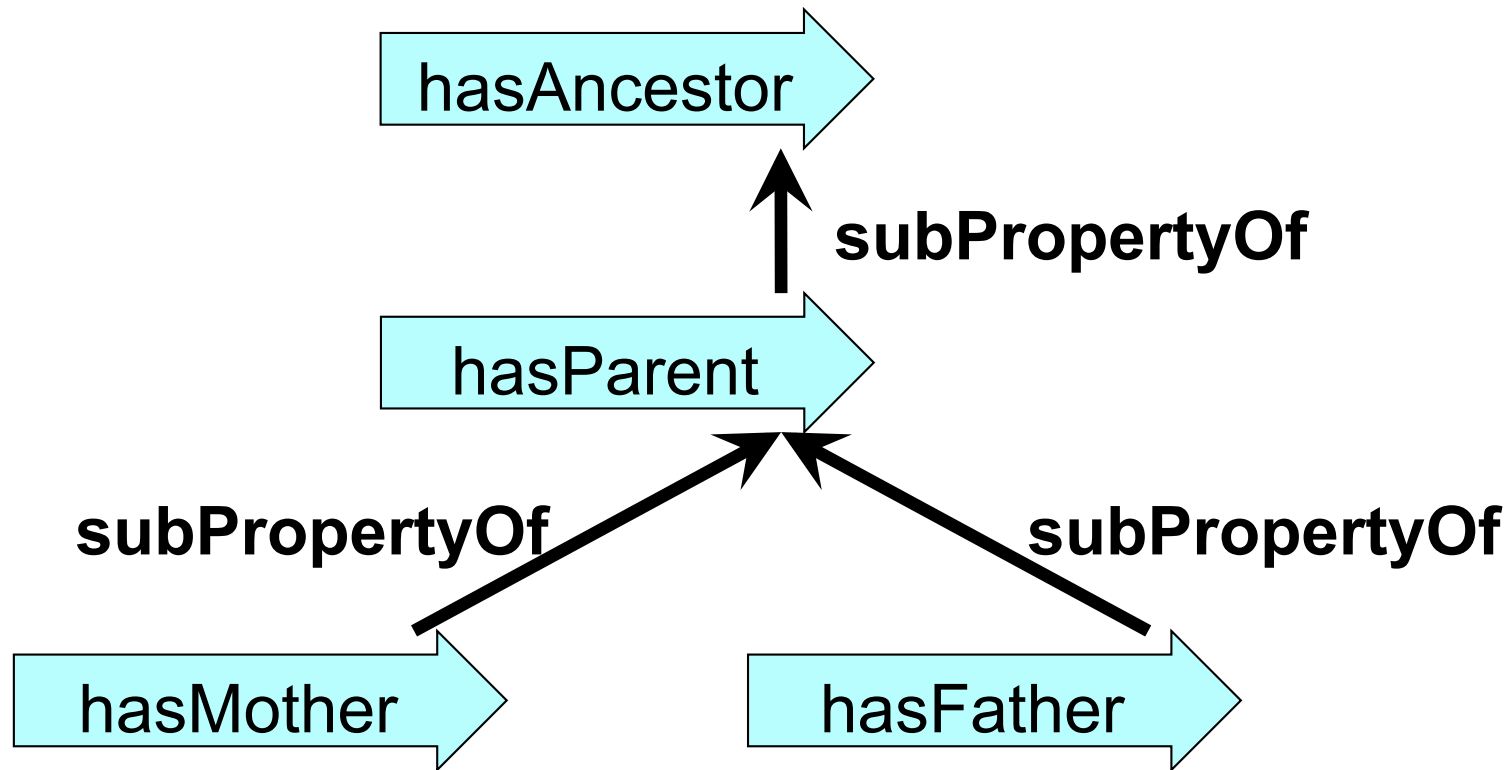
Classes and Sub-Classes



operational excellence

london conference

Properties and Sub-Properties

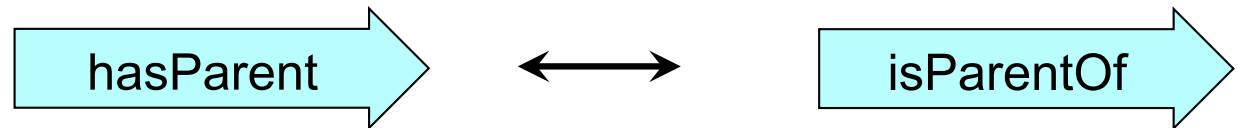


operational excellence

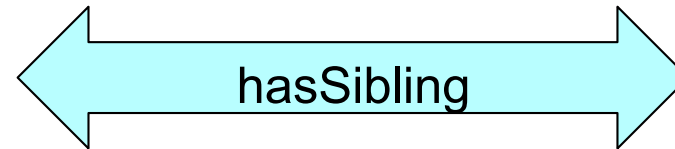
london conference

Inverse, Symmetric, Transitive, Functional Properties

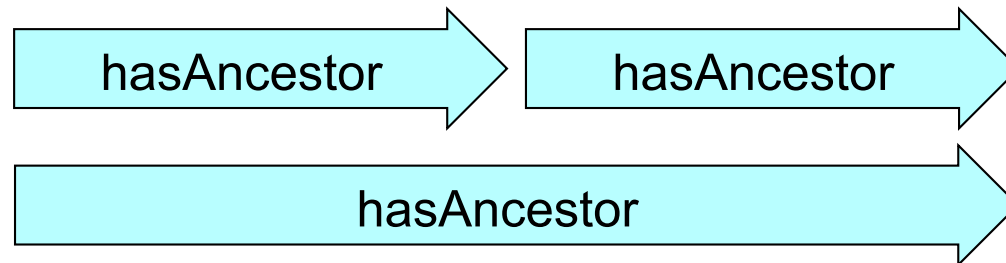
INVERSE



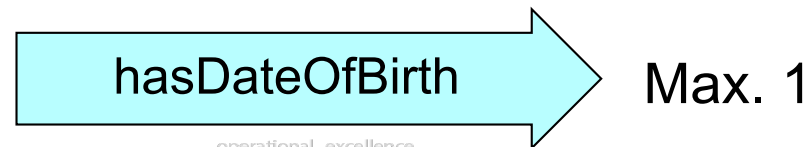
SYMMETRIC



TRANSITIVE



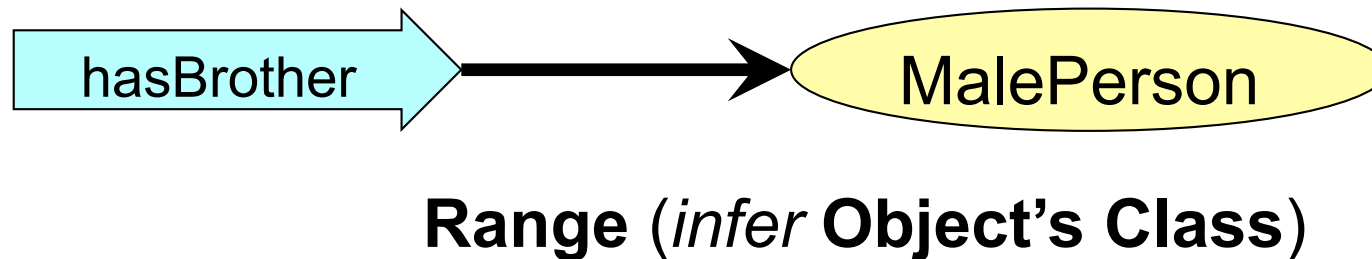
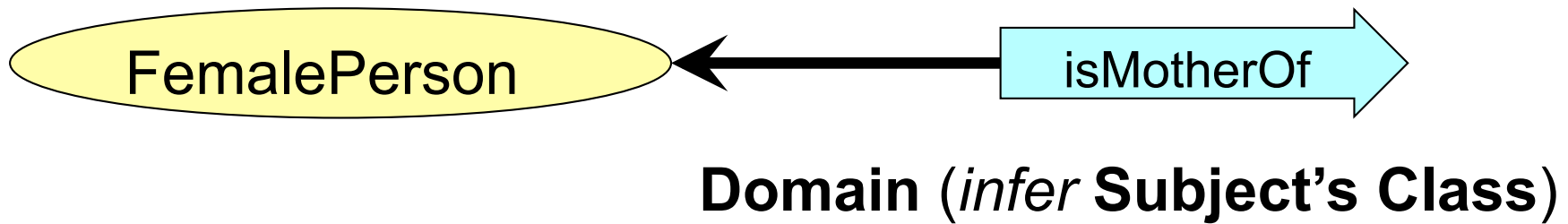
FUNCTIONAL



operational excellence

london conference

Domain and Range of Properties

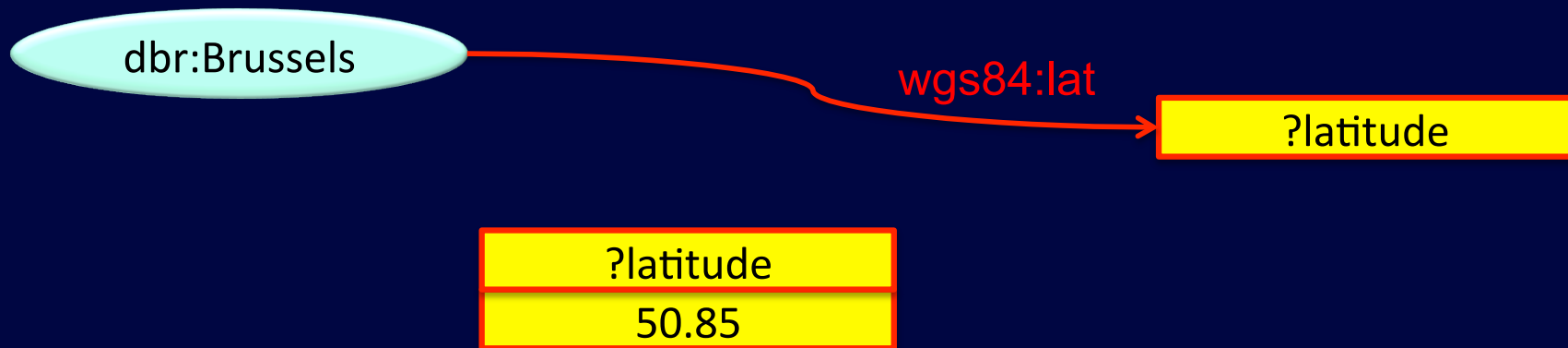


operational excellence

london conference

Using SPARQL to query facts

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



SPARQL Query Language

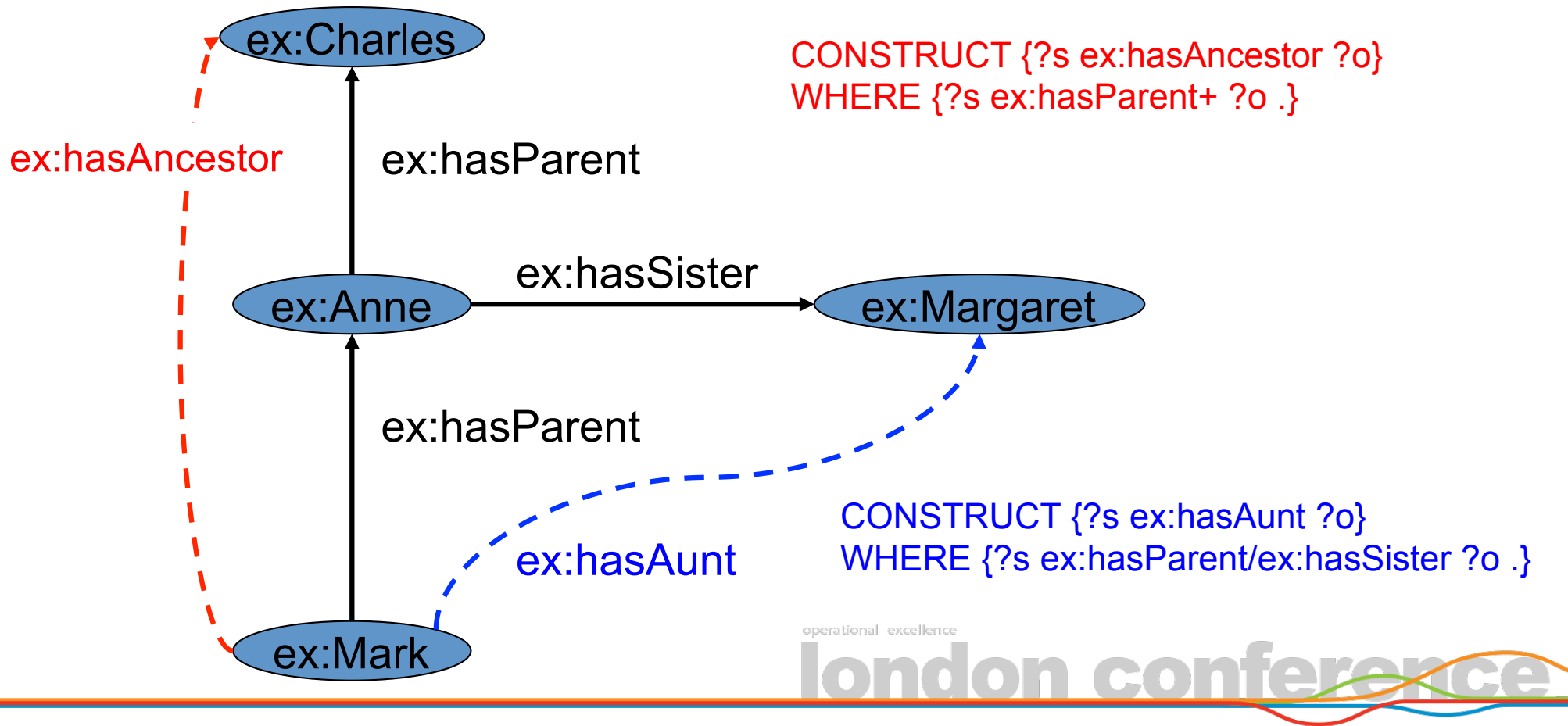
- W3C standard RDF query language
www.w3.org/TR/rdf-sparql-query (SPARQL 1.0 W3C recommendation)
www.w3.org/TR/sparql11-query (SPARQL 1.1 W3C recommendation)
- Enables **federated queries** to be made **across multiple RDF data sets** and SPARQL service endpoints
- Can use this **within** the enterprise to do mash-ups of enterprise data *with* open public linked data (e.g. mapping data, demographic data, traffic data or weather data)
- Can **CONSTRUCT** new RDF data (logical inferences) from existing RDF data **WHERE** it matches particular constraints / criteria specified in the SPARQL queries

operational excellence

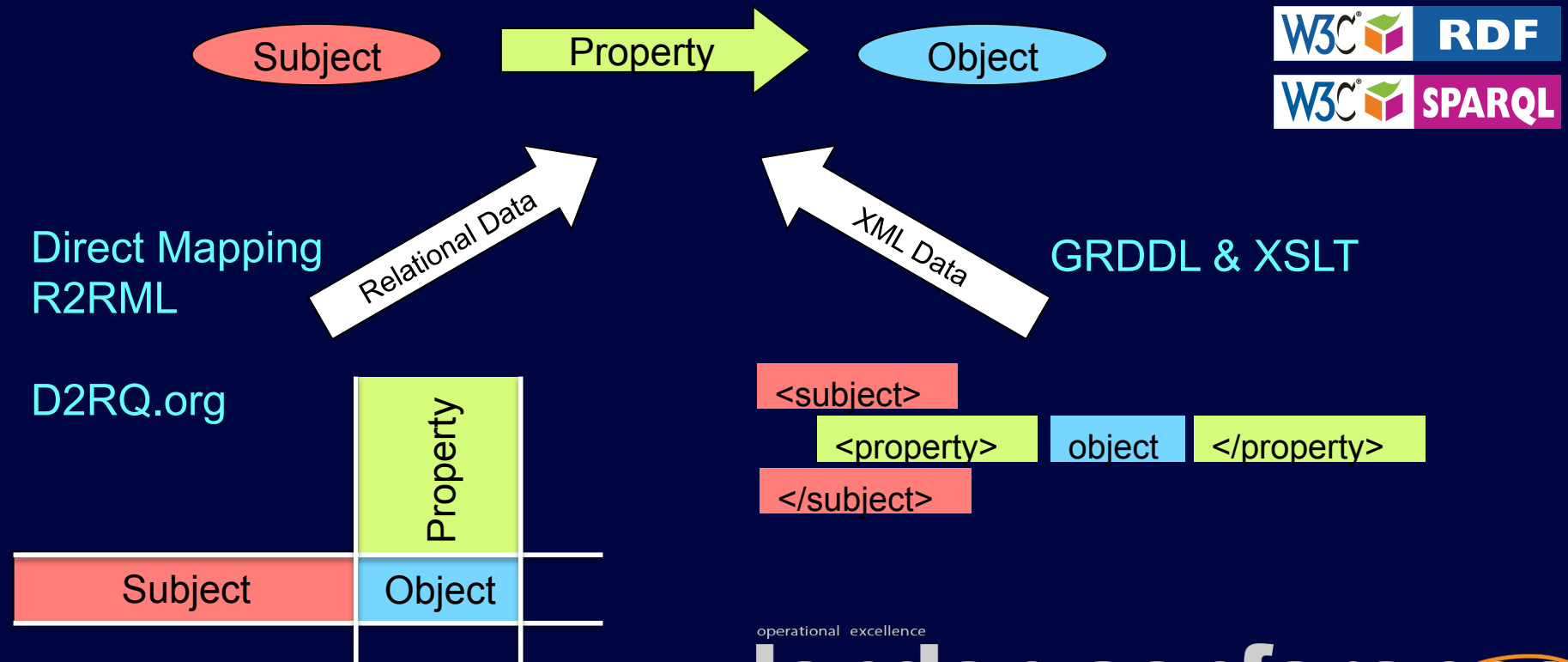
london conference



User-Defined Rules via SPARQL CONSTRUCT



Your existing data as Linked Data



operational excellence

london conference

Linked Data vocabularies



schema.org



operational excellence

london conference

Adding Linked Data to your web pages

The screenshot shows the Google Structured Data Markup Helper interface. The left pane displays a Tesco product page for 'Nestle Cheerios 600G'. The right pane shows the generated HTML source code with microdata markup. The URL in the browser is <https://www.google.com/webmasters/markup-helper/tagger?sourceid=1815004>.

Product Details:

	Calories	Sugar	Fat	Saturates	Salt
	175	12g	3g	2g	0.5g
%	9%	14%	5%	8%	8%

Deliciously crunchy Os, packed with 4 whole grains. Say Yes to Cheerios
It's simple, everybody knows that the whole family love Cheerios. The perfect

HTML Source with Microdata Markup:

```
<!-- Microdata markup added by Google Structured Data Markup Helper. -->
<html lang="en-GB" xml:lang="en-GB" dir="ltr"
xmlns="http://www.w3.org/1999/xhtml"><head><meta http-equiv="X-UA-
Compatible" content="IE=8"/><title>Nestle Cheerios 600G - Groceries -
Tesco Groceries</title><link rel="Home" href="/groceries/"
title="Tesco.com Groceries Homepage"/><link rel="Index"
href="/groceries/" title="Tesco.com Groceries Homepage"/><link
rel="checkout" href="/groceries/checkout/" title="Checkout Tesco.com
Groceries"/><link rel="help" href="/groceries/help/" title="Tesco.com
Groceries help pages"/><link rel="alternate" type="application/rss+xml"
title="Tesco.com RSS feed" href="http://www.tesco.com/rss/"><link
rel="shortcut icon" href="http://www.tesco.com/favicon.ico"/><meta
name="description" content="Buy Nestle Cheerios 600G at Tesco.com."/>
<meta name="keywords" content="Buy Nestle Cheerios 600G, Nestle Cheerios
600G online, Nestle Cheerios 600G UK, Cheap Nestle Cheerios 600G, Buy
Nestle Cheerios 600G, Online grocery shopping, Online shopping, shopping
delivered, Tesco.com."/><meta http-equiv="Content-Type"
content="text/html; charset=UTF-8"/><meta http-equiv="Content-Style-
Type" content="text/css"/><meta http-equiv="Content-Language"
content="en"/><meta name="rating" content="general"/><meta name="branch"
content=""><meta name="revisit-after" content="7 days"/><meta
name="MSSmartTagsPreventParsing" content="true"/><meta http-
equiv="imagetoolbar" content="no"/><link rel="stylesheet"
type="text/css"
href="http://ui.tescoassets.com/groceries/UIAssets/Css/.../Compressed/cor
e_34063.css"/><style type="text/css" media="screen"> </style><link
```

operational excellence

london conference

Example: web page with Linked Data

The screenshot shows the original Tesco website interface. The URL is www.tesco.com/groceries/Product/Details/?id=250106683. The page displays the product details for Nestle Cheerios 600G. The price is £3.19 (£0.53/100g). The page includes a search bar, navigation menu, product image, and a 'Cheaper Alternative' section for Tesco Multigrain Hoops. The page also features nutritional information, allergen details, and ingredients.

Original web page has no Linked Data

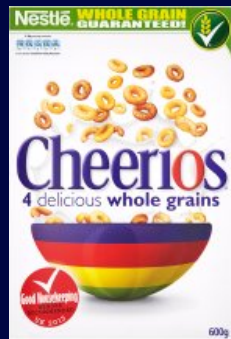
Our modified version has...

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

operational excellence

london conference

Example: web page with Linked Data



Web Page

05011476100885

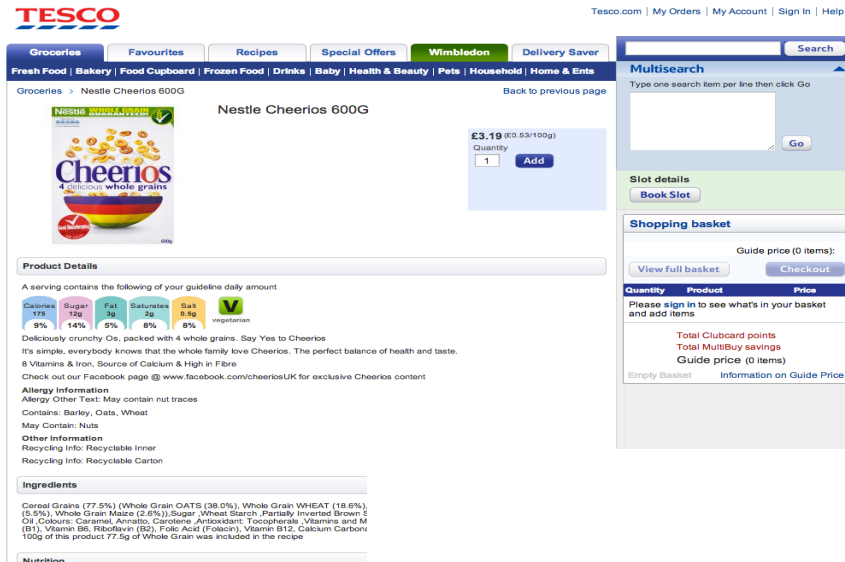


<http://tinyurl.com/rdfa-test>

Visualisation by
<http://rdfa.info/play>

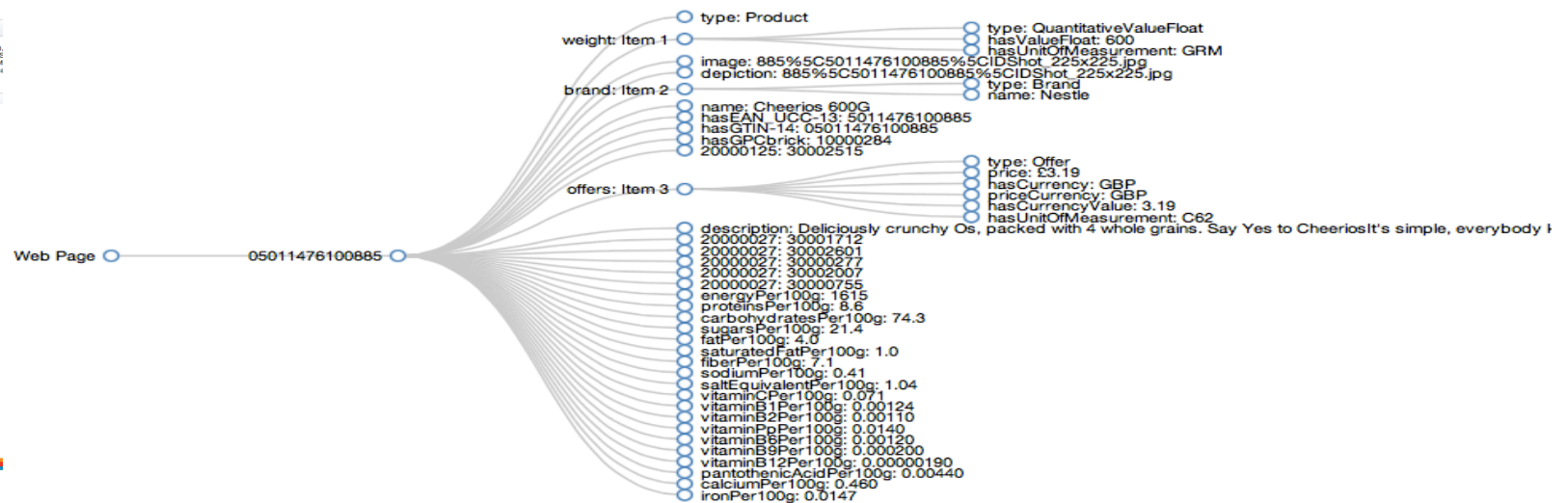
operational excellence
london conference

Example of using RDFa markup



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

View Page Source and paste into
<http://rdfa.info/play>



Enabling new consumer interactions



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

operational excellence

london conference

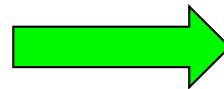
Business-to-Consumer example



*Map human-readable keyword(s) to
Product category identifier (GPC)*

*Contextual filters are shown
for the product category*

User constraints are specified



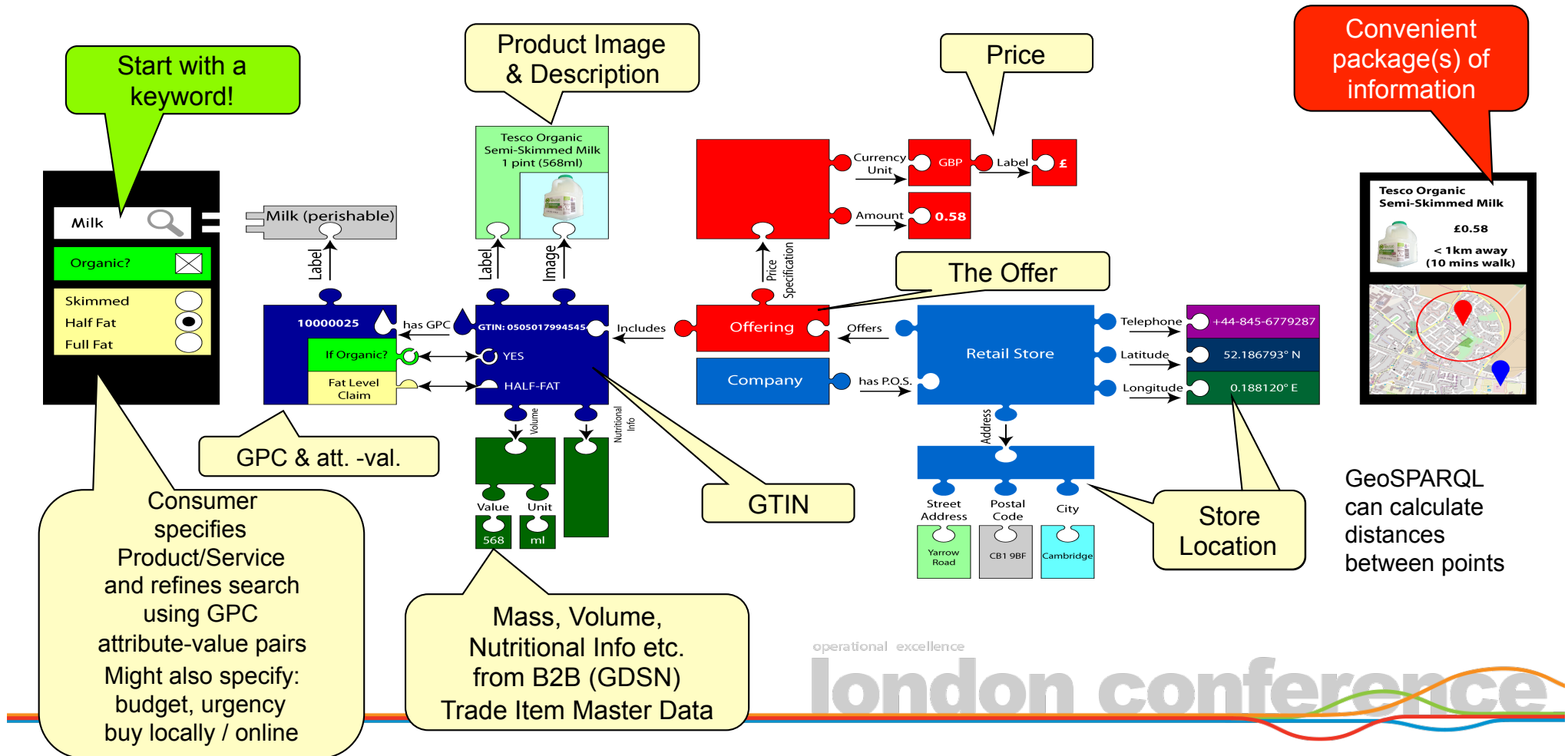
*Information about matching
products and services*



operational excellence

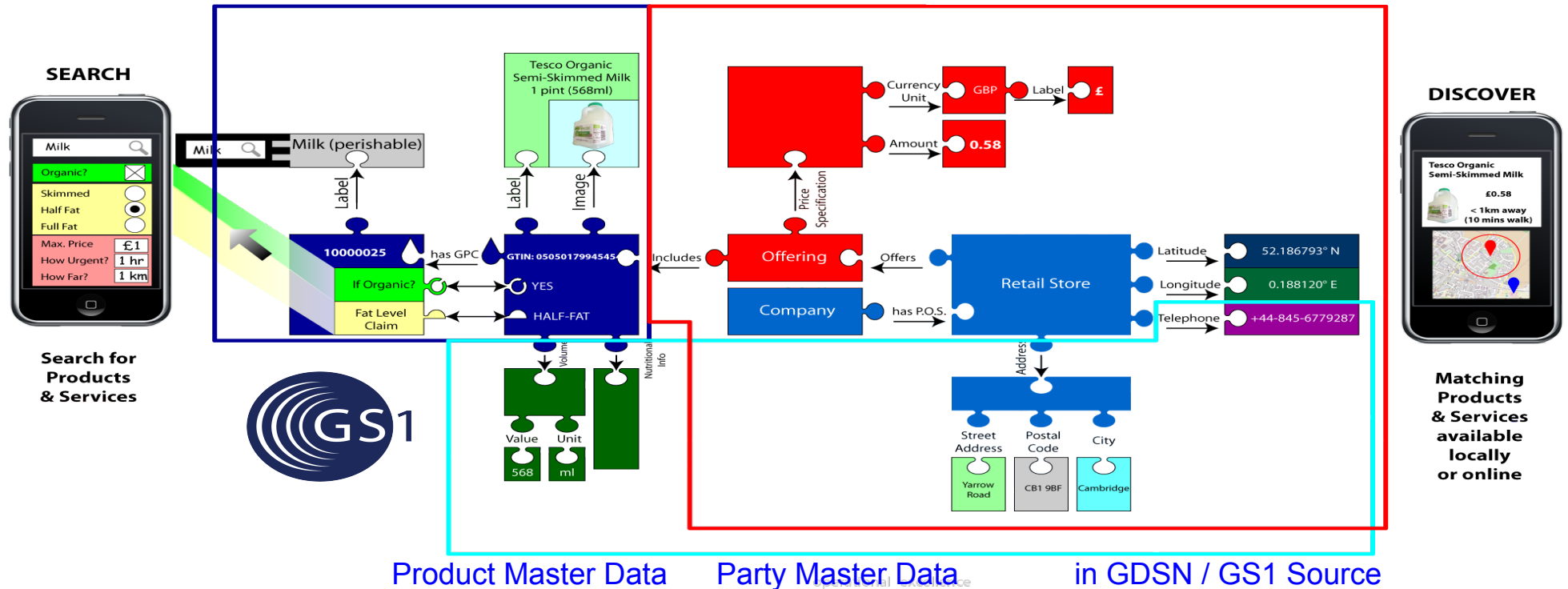
london conference

Consumer to Business example



Data Linkages

GS1 Global Product Classification
(GPC) + attribute - value pairs



Product Master Data

Party Master Data

in GDSN / GS1 Source

london conference

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



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

Why should this be important to consumers?

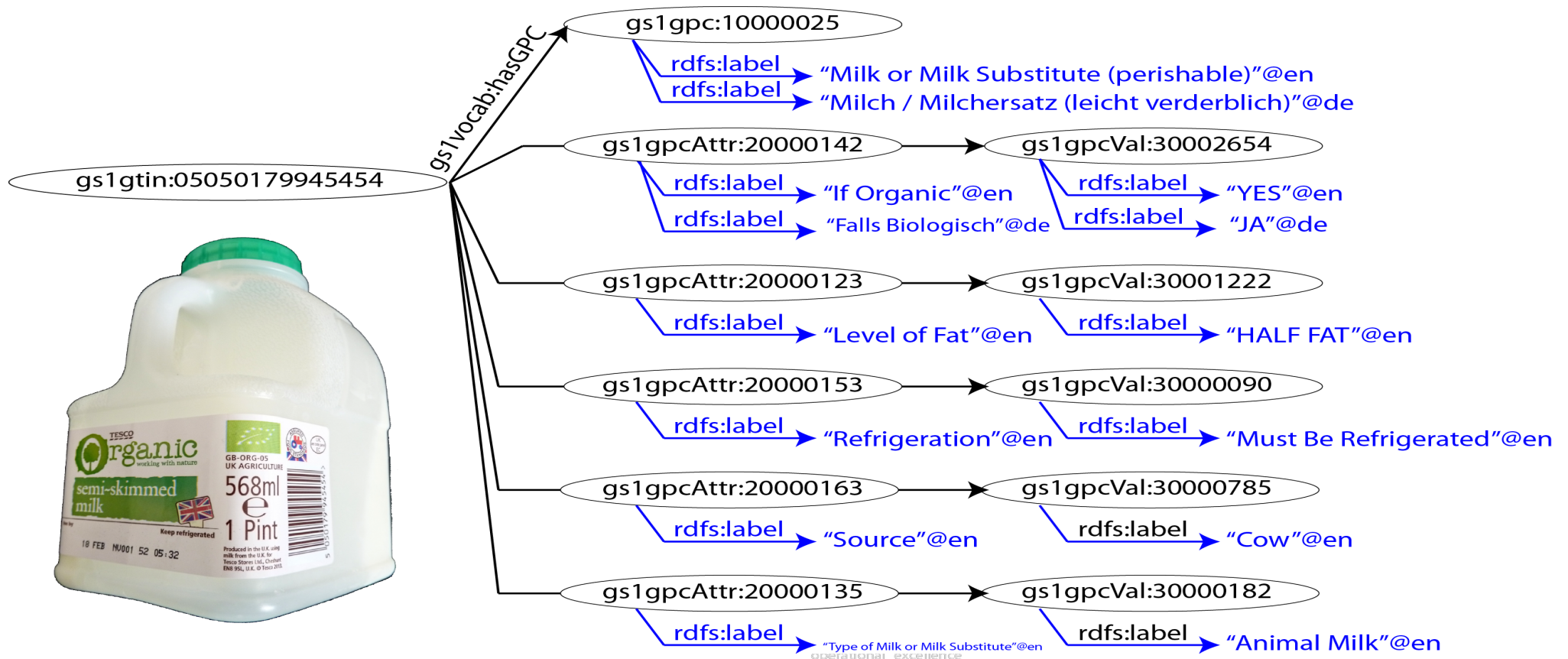
- Consumers can **more easily find the products and services that match** their needs and preferences:
- **Less time** *actively* trawling the web for specifications, price comparison, ratings, reviews, checking availability etc.
- **Smarter search engines** on the web / **search agents** in the cloud:
 - Enter a keyword and it attempts to **understand the context**,
 - Providing the user with (**contextual**) relevant ways of filtering their search
 - Technical specifications (e.g. for consumer electronics products)
 - Ingredients, nutritional information and potential allergens (food, pharmaceuticals)
 - Accreditation (Fair Trade, Marine Stewardship Council, Organic/Bio, Free Range etc.)
 - Measures of through-life environmental footprint (e.g. for electrical appliances, food)
 - Price (unit item price, delivery charges) and promotional offers
 - Ratings and recommendations from other consumers
 - Proximity of local availability (GLN → Street Address → Latitude/Longitude)
 - Lead time of remote availability
- Infrastructure for automated shopping agents and travel planning agents that gather the relevant information on behalf of consumers (and their preferences / needs), presenting them with options for bespoke tailor-made packages (all relevant info collected coherently), **Buy Now** *in fewer clicks*.



operational excellence

london conference

Global Product Classification as linked open data



Ways to embed structured data

RDFa (Resource Description Framework in attributes)

<http://www.w3.org/TR/rdfa-primer/>

HTML5 Microdata

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

JSON-LD (JavaScript Object Notation for Linking Data)

<http://json-ld.org/>

operational excellence

london conference



RDFa vs HTML5 Microdata vs JSON-LD

RDFa and HTML5 Microdata both use attributes within HTML elements to include structured data facts inline, close to the visible markup

RDFa is easier to use with multiple vocabularies and also supports explicit declarations of datatypes, e.g. `xsd:datetime`

RDFa Lite 1.1 is functionally equivalent to HTML5 Microdata but is also upwards-compatible to RDFa, for advanced features.

JSON is a lightweight way of exchanging structured data (alternative to XML). JSON-LD uses Linked Data principles (URIs etc.) to avoid ambiguities and to ensure semantic mappings across JSON datasets

JSON-LD is not really intended for inline markup within HTML tags.

operational excellence

london conference



Tools for creating Linked Data markup

GoodRelations Snippet Generator

<http://www.ebusiness-unibw.org/tools/grsnippetgen/>

Google Structured Data Markup Helper

<https://www.google.com/webmasters/markup-helper/>

Further tools for RDFa listed at:

<http://www.w3.org/2001/sw/wiki/RDFa>

operational excellence

london conference



Tools for extracting / verifying Linked Data markup

RDFa Play

<http://rdfa.info/play/>

Apache Any23 (Anything To Triples)

<http://any23.org/>

<http://any23.apache.org/>

W3C RDFa 1.1 Distiller and Parser

<http://www.w3.org/2012/pyRdfa/Overview.html>

Google Structured Data Testing Tool

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

operational excellence

london conference



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



References and further reading

W3C Semantic Web Activity (presentations, specifications - RDF, RDFS, OWL, SPARQL)

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

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.

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

Linking Enterprise Data

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

http://3roundstones.com/led_book/led-contents.html

Semantic Web Concepts - presentation by Sir Tim Berners-Lee

<http://www.w3.org/2005/Talks/0517-boit-tbl/>

operational excellence

london conference

