Semantic Web technologies
Introduction & relevance to the supply chain

Mark Harrison

Auto-ID Labs,
University of Cambridge

mark.harrison@cantab.net

AUTO-ID LABS

UNIVERSITY OF CAMBRIDGE

Outline

- What is the Semantic Web?

- Core concepts
  - Linked Data
  - Resource Description Framework (RDF)
  - Ontologies and Web Ontology Language (OWL)
  - SPARQL Query Language

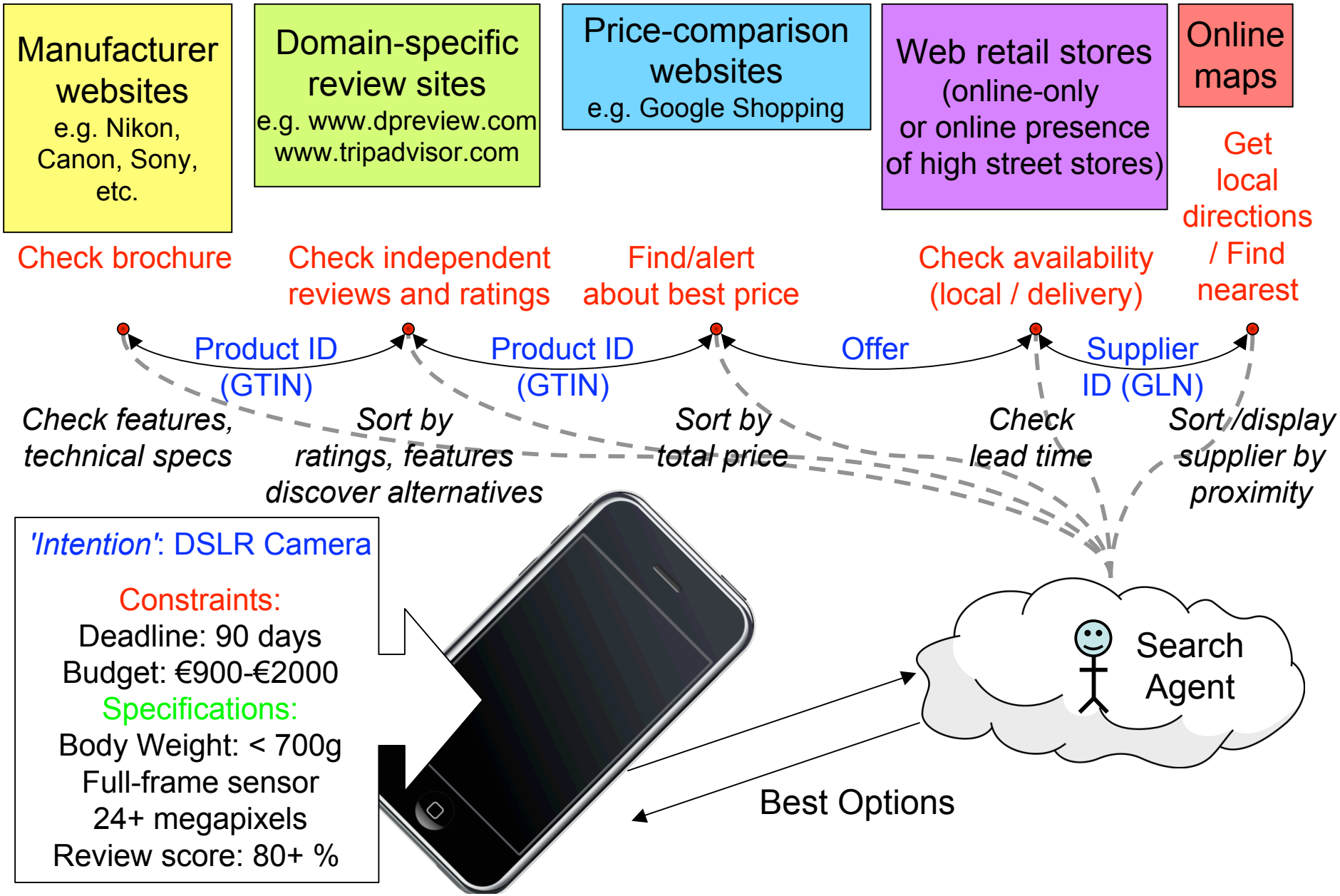
- Why is this relevant to manufacturers and retailers?
  - Helping consumers find products, product information and offerings (from retailers)
  - Relation to current GS1 standards initiatives on extended packaging, trusted source of data
  - Joining all the dots from 'Intent' to the decision to Buy Now!

What is the Semantic Web?

- **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

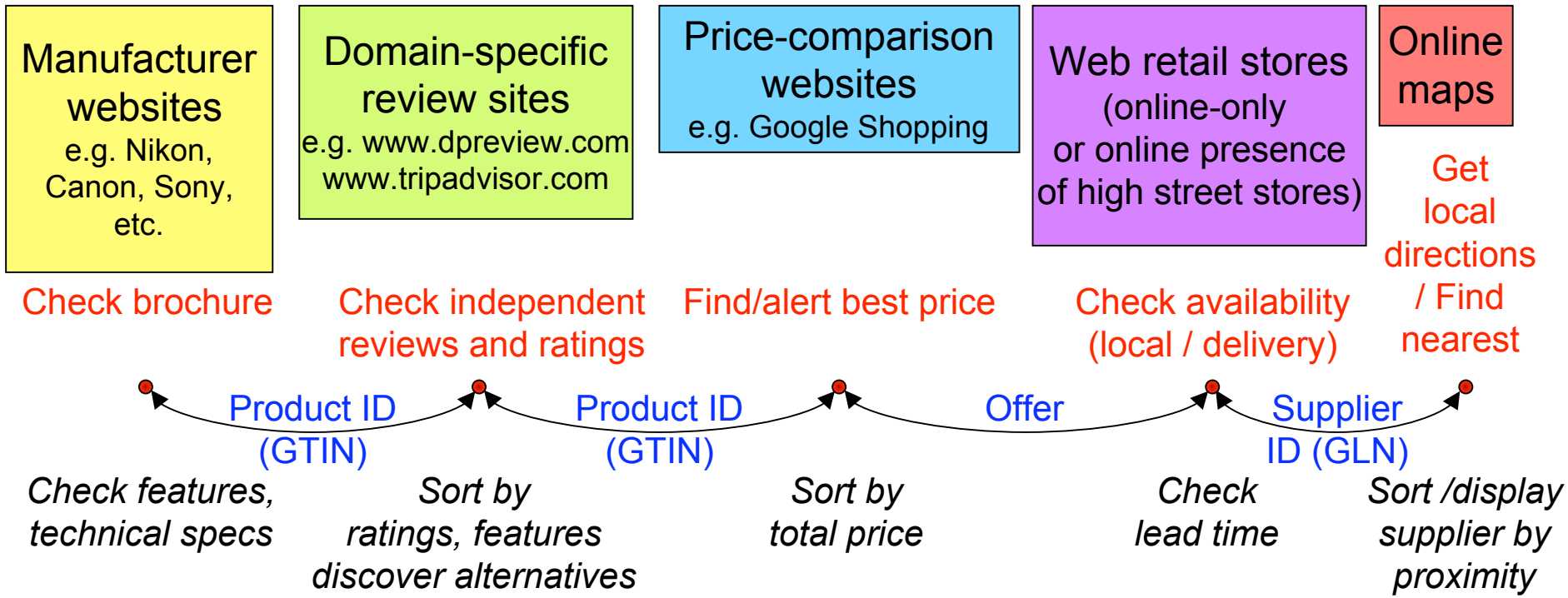
- **Semantic Web / Linked Data**
  - builds on web technologies to achieve a global network of **linked data** at web scale
  - enables unified federated queries of data across multiple distributed data sources
  - can ease data integration across different types of databases
  - enables automated logical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

Background / Passive Search



Joining the dots - from 'Intent' to Buy Now!
-using semantic web technology

Application scenarios that involve querying ***multiple*** data sources (data 'islands') that are linked in some way (via Product ID, Supplier ID, Offer, many other relationships)

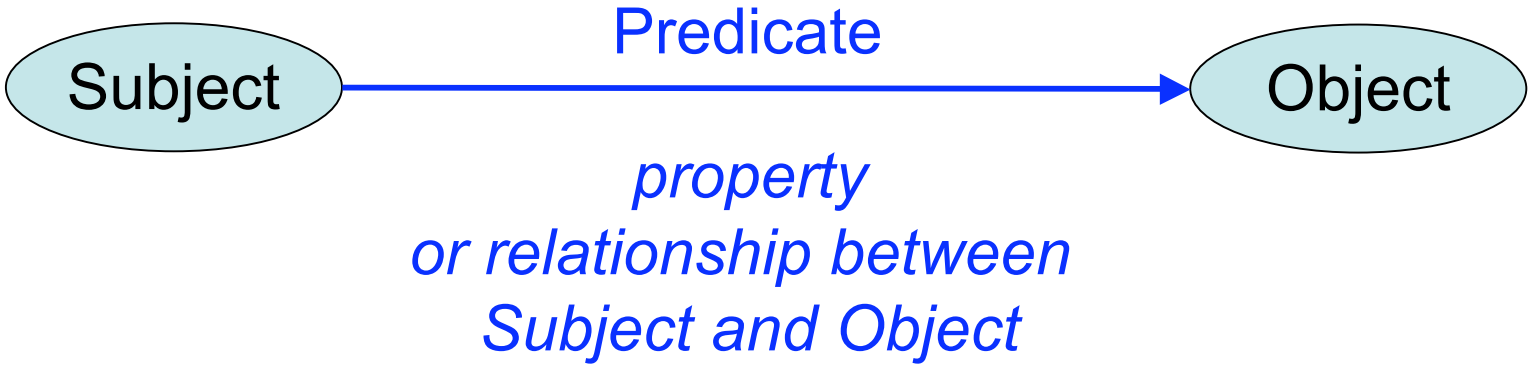


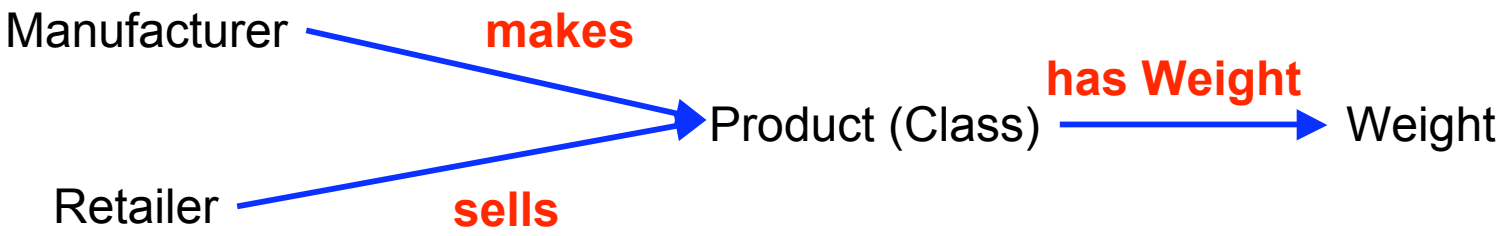
*(Equally applicable to **automating** the internal sourcing of your suppliers)*

What is the Semantic Web?

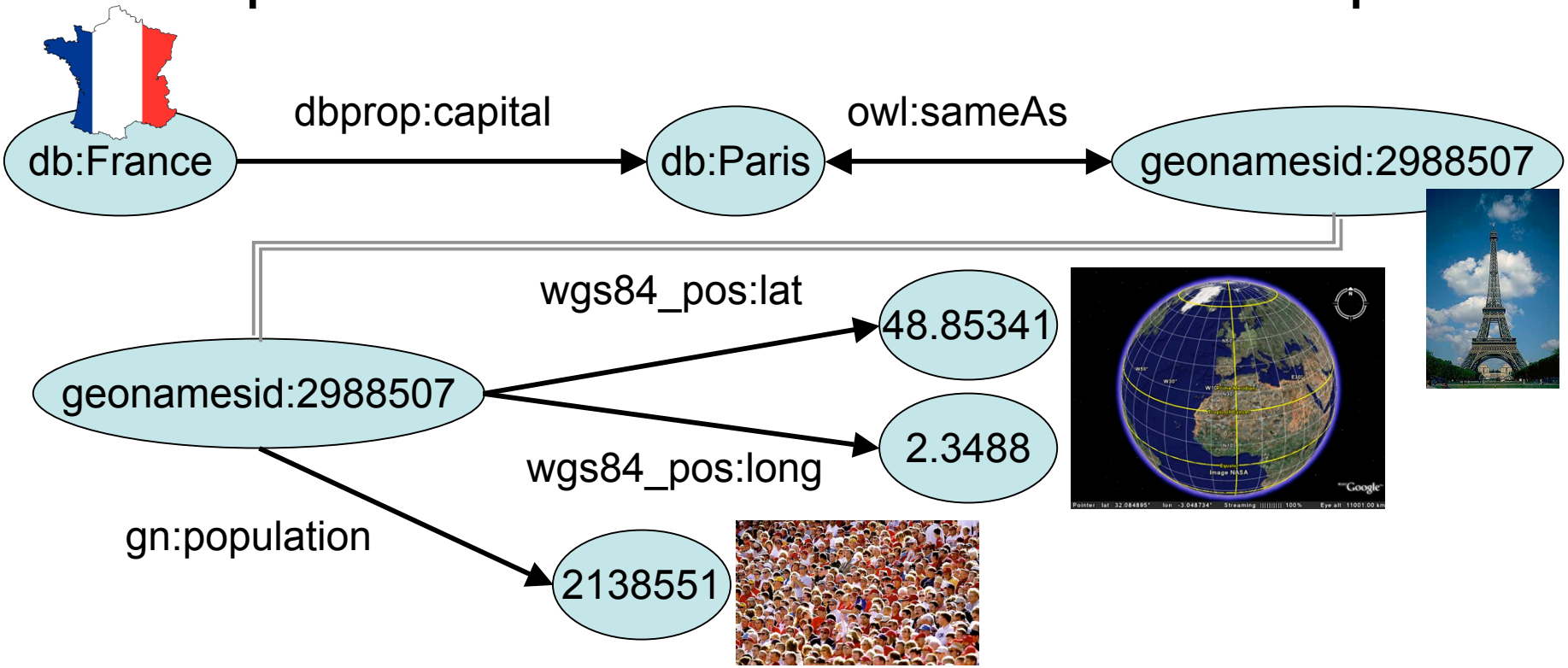
The Semantic Web builds on web technologies to achieve a global network of **linked data** to enable **unified global queries of data** and **logical deduction** (additional information can be inferred).

hyperlinks* + *meaning





Example of semantic web data - RDF triples

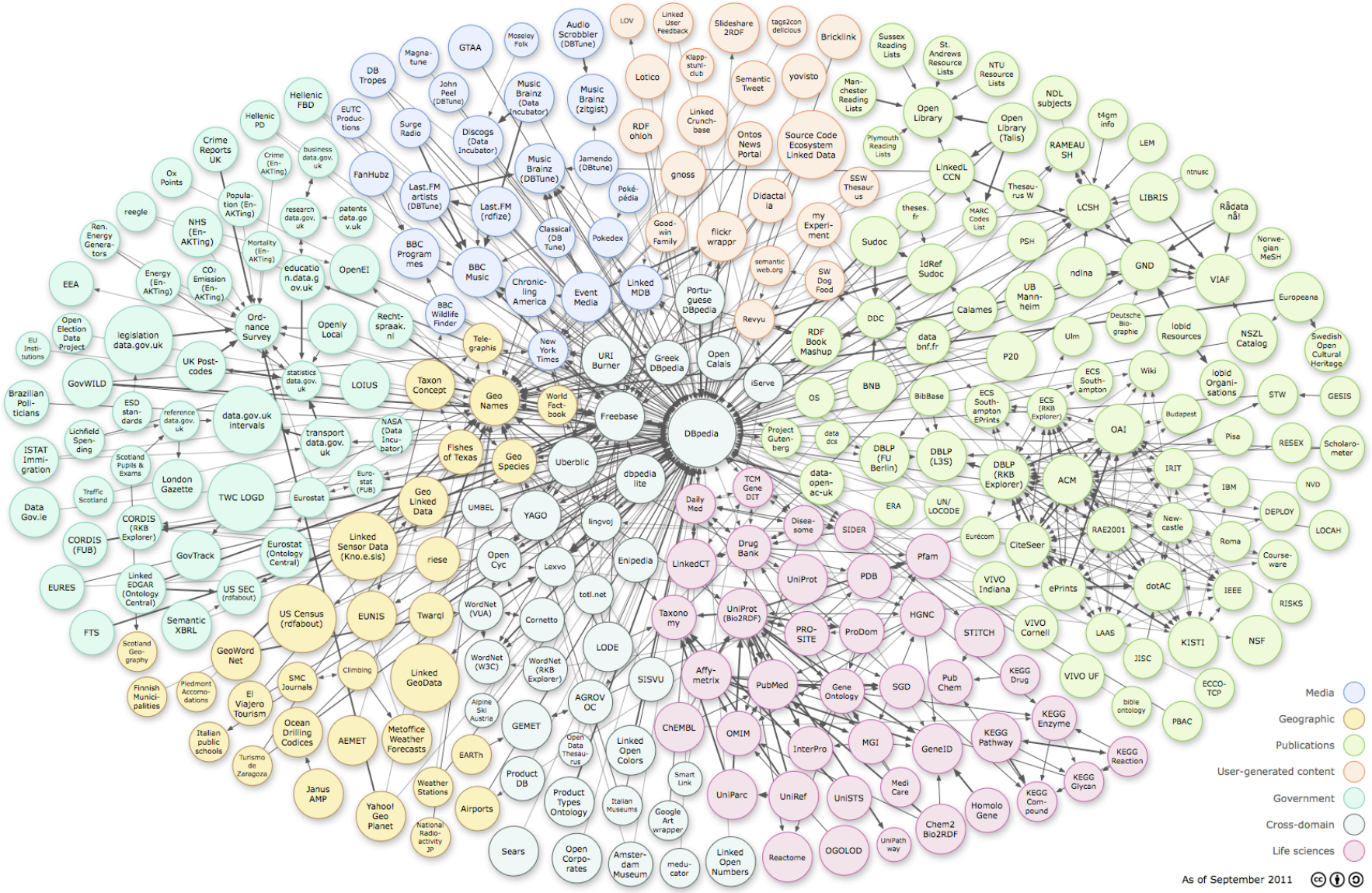


db:France	**dbprop:capital**	**db:Paris**	.
db:Paris	**owl:sameAs**	**geonamesid:2988507/**	.
geonamesid:2988507/	**gn:population**	**2138551**	.
geonamesid:2988507/	**wgs84_pos:lat**	**48.85341**	.
geonamesid:2988507/	**wgs84_pos:long**	**2.3488**	.

db: = http://dbpedia.org/resource/
dbprop: = http://dbpedia.org/property/
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#

Linked Open Data cloud of datasets



Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net

Core Semantic Web 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 **construct**ed (*inferred*) from existing data.

URIs as identifiers for everything

- **Uniform Resource Identifiers (URIs)** used to identify not only documents but also concepts (people, places, things, abstract/intangible concepts) and properties / data relationships

  http://dbpedia.org/resource/Brussels

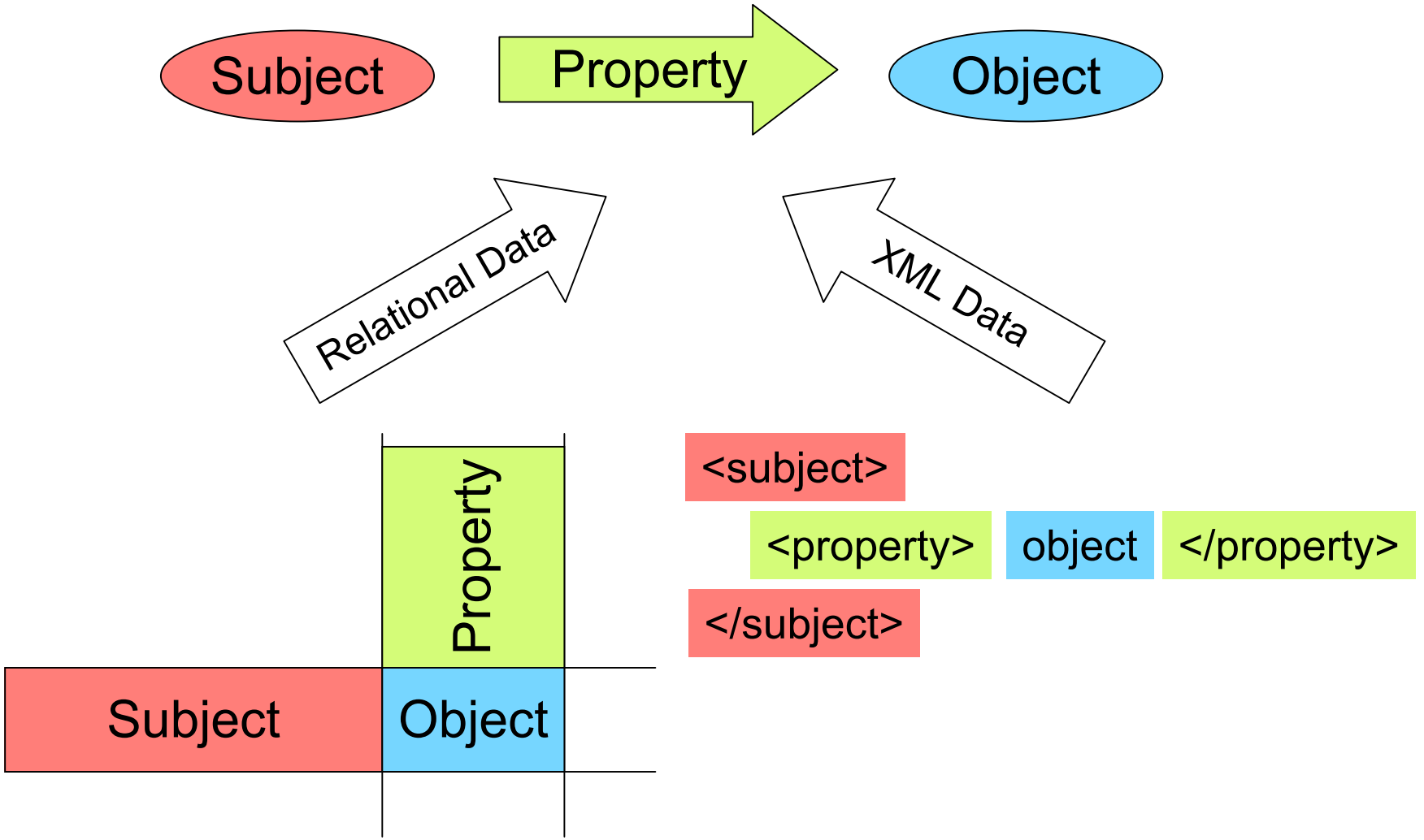
  http://purl.org/goodrelations/v1#hasGTIN-14

- GS1 standards already uses URIs
  - EPCs are canonically expressed as URIs
    urn:epc:id:sscc:0614141.1234567890
  - EPCIS Core Business Vocabulary uses URIs for values of: businessStep, disposition, readPoint, businessLocation, transaction type and identifiers.

    urn:epcglobal:cbv:bizstep:shipping

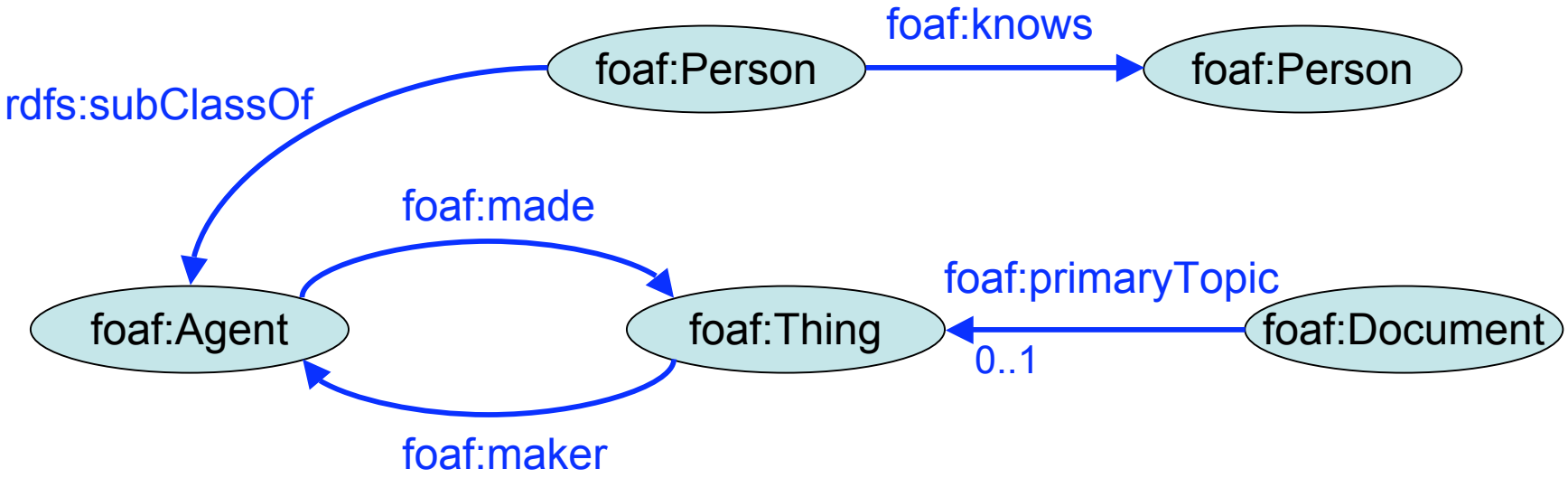
Resource Description Framework (RDF)

- provides a W3C standard way to write simple logical statements in a 'lowest common demoninator' format:



Ontologies, RDF Schema (RDFS) and OWL

- Ontologies are data dictionaries with additional annotations about how various properties (predicates) and classes of resources are related to each other **(also at an abstract level / data model)**



- Ontologies exist for multiple domains of interest
- Ontologies can be used together and also cross-referenced e.g. owl:sameAs , owl:equivalentClass, owl:equivalentProperty
- Some 'core' ontologies include FOAF, DublinCore, GeoNames...

RDF Schema (RDFS) and OWL

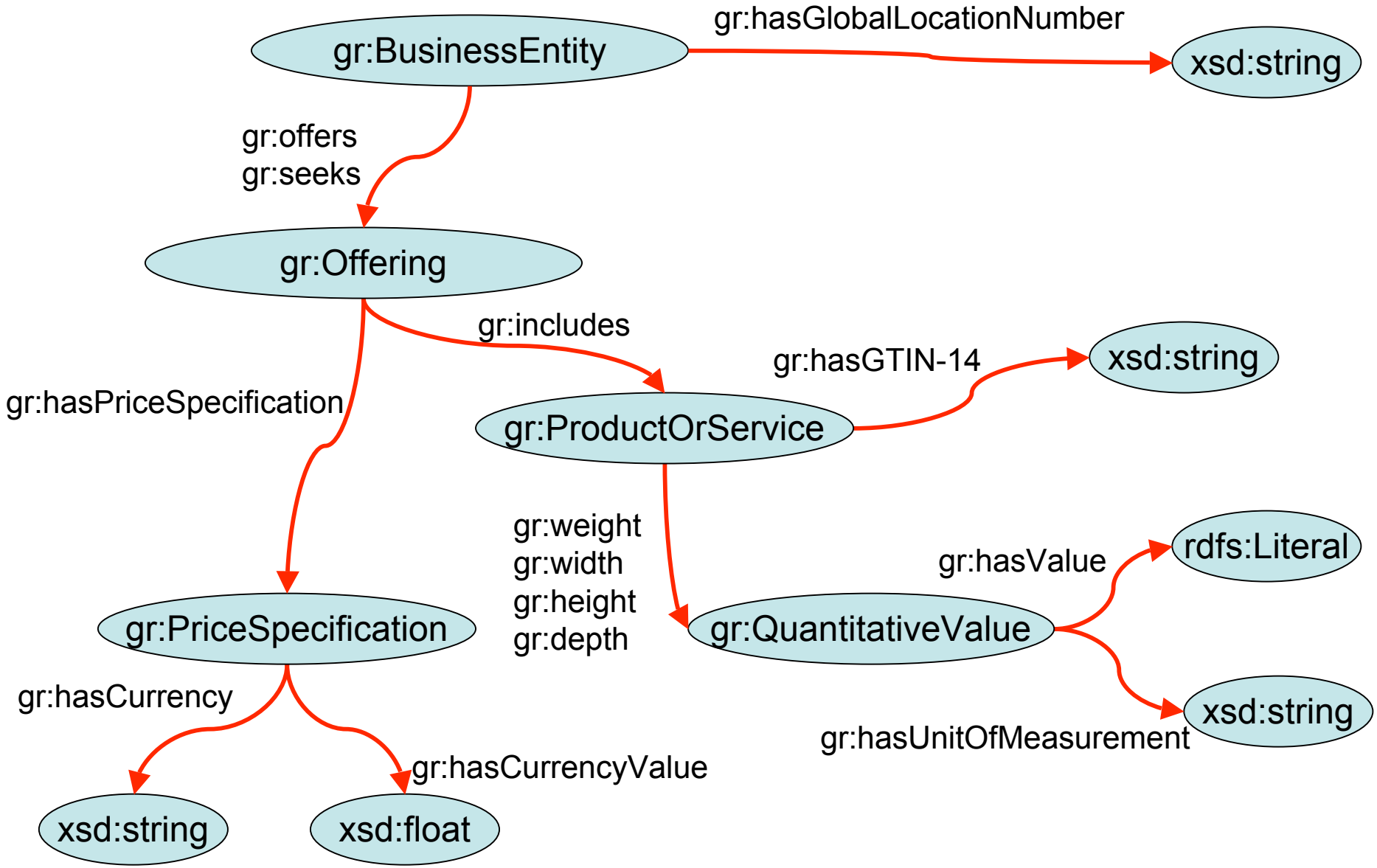
- **RDF Schema (RDFS)** introduces basic concepts such as:
  classes and properties, subclasses and subproperties,
  human-readable labels in various languages (more readable than URIs)
  ranges (what can be inferred about the object's class) and
  domains (what can be inferred about the subject's class)

  www.w3.org/TR/rdf-schema

- **Web Ontology Language (OWL)** is more expressive, including:
  - intersection, union and complement of sets
  - inverse properties, transitive properties, symmetric properties
  - equivalent classes or equivalent properties
  - whether two individuals are the same or different
  - chaining of properties using owl:propertyChain

  www.w3.org/TR/owl-ref

RDF examples using GoodRelations ontology



gr: = http://purl.org/goodrelations/v1#

SPARQL Query Language

- W3C standard RDF query language

  www.w3.org/TR/rdf-sparql-query (SPARQL 1.0 W3C recommendation)

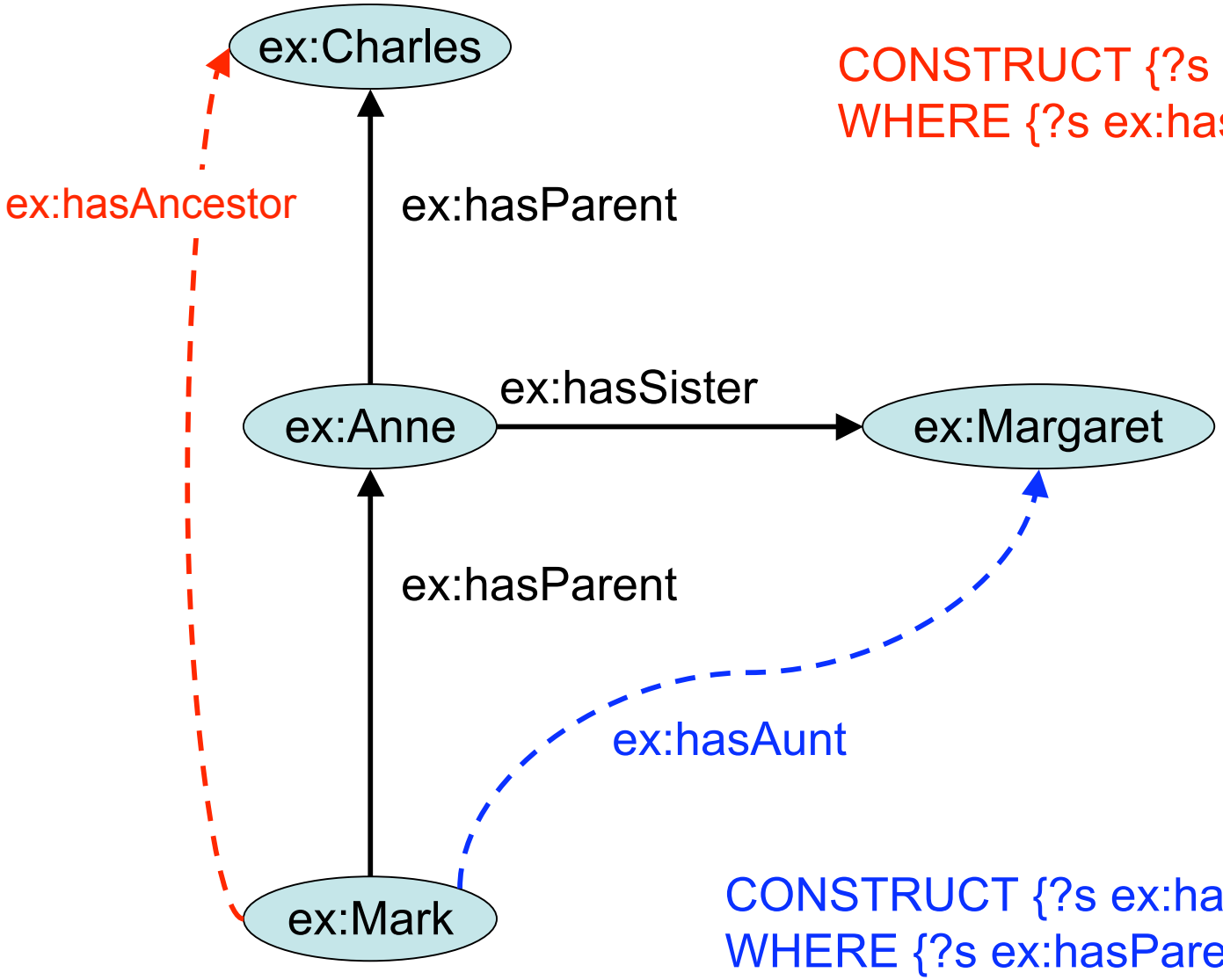
  www.w3.org/TR/sparql11-query (SPARQL 1.1 working draft)

- 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

SPARQL Query Language - very simple example



CONSTRUCT {?s ex:hasAncestor ?o}
WHERE {?s ex:hasParent+ ?o .}

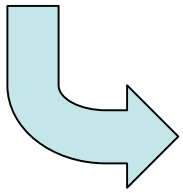
CONSTRUCT {?s ex:hasAunt ?o}
WHERE {?s ex:hasParent/ex:hasSister ?o .}

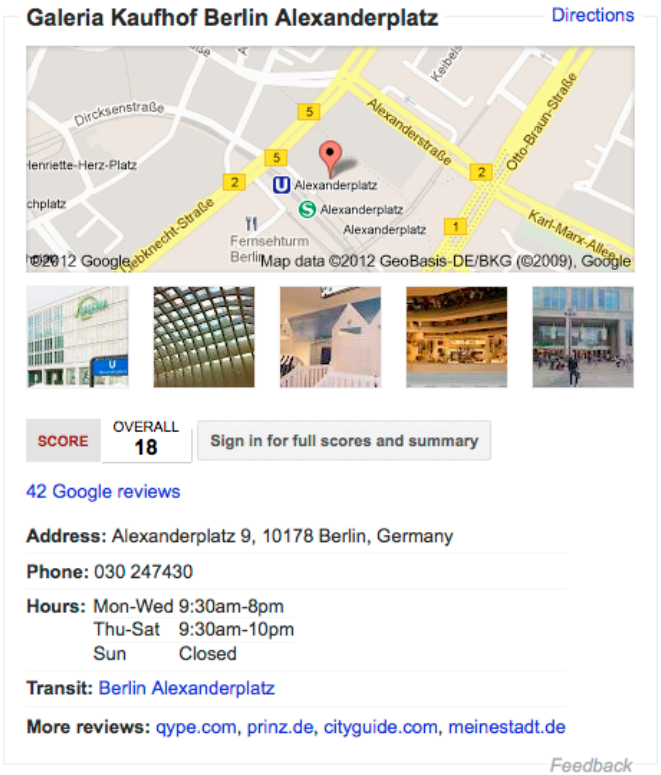
Why is this important now?

- 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 ...





Why is this important now?

- Many linked open data sources exist
  - see the Linked Open Data cloud at http://lod-cloud.net/
  - Government Data (financial, demographic, geographic / mapping)
    e.g. http://www.data.gov/ http://www.data.gov.uk/ http://publicdata.eu
  - Geographic Data, e.g. http://www.geonames.org
  - Info boxes from Wikipedia = dbPedia (http://dbpedia.org)

You can start using this now!

- Internal use:
  - Privately mash-up your own enterprise data with public open linked data (for easier visualization, discovering new insights)

- Externally:
  - To make it easier for search agents to find YOUR products / services (especially 'niche' or specialist: Kosher, Vegan, Gluten-Free)
  - To make it easier for consumers to find YOU as a local supplier/retailer (to help combat the increasing loss of trade to online-only retailers)

Examples of using linked data

www.publicdata.eu/app

Why is this important now?

- Web search engine companies are actively encouraging website owners to use semantic markup.
- Schema.org - joint initiative from Google, Yahoo, Bing, ...
  - basic details about products, services, companies, contact details
- GoodRelations ontology and productontology.org
  "The Web Vocabulary for Electronic Commerce"
  - Richer vocabulary especially for describing product master data
  - developed by Prof. Martin Hepp and the E-Business and Web Science research group at the University of the Bundeswehr in Munich
  - GoodRelations now integrated with schema.org and recognized by major search engines, used to provide enhanced search results

  http://purl.org/goodRelations
- "Core Business Vocabulary" and "Core Location Vocabulary"
  - 2 of 3 eGovernment Core Vocabularies from the ISA programme of the European Commission

  http://joinup.ec.europa.eu/site/core_business

Why should this be important to manufacturers and retailers?

- Manufacturers and retailers have worked together within the GS1 community to develop standards and services for the B2B sharing of product master data about organizations and their products (GDSN, Align Trade Item Business Message Standard)
- GS1 has recently launched B2C initiatives:
  - Extended Packaging
  - Trusted Source of Data
- These appear to currently focus on scanning a barcode to find additional trusted information about a product
- What about ***online*** product searches, ***before*** we have the physical product in our hands? ... ***before*** we have even selected the product?
- Online tools can be developed to make it easy to export a subset of the rich B2B master data in linked data format, to help brand owners and retailers to improve their search engine rankings *(particularly attractive for SMEs with limited in-house IT capabilities / expertise)*

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

GDSN Master Data (Products)

- 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*.

References and further reading

W3C Semantic Web Activity
(presentations, links to 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/