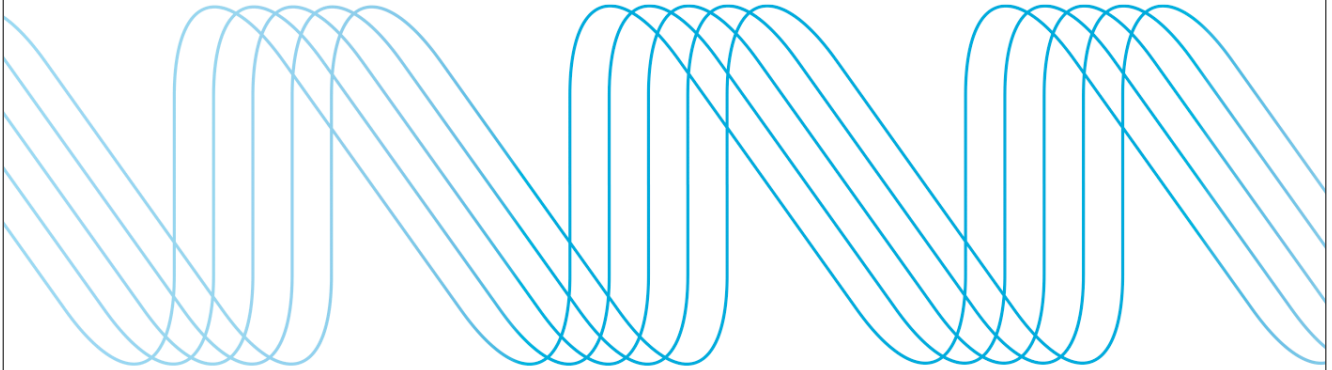

Maximising Online Resource Effectiveness



Session 7
Joining the semantic web

Subtopics

- What do we mean by semantic web
- The significance of metadata
- Resource Description Framework (RDF)
- Embedded metadata and RDFa
- Vocabularies and ontologies
- Linked data
- Mashups and mashouts
- Open data and open government

Semantic web

W3 future directions keynote at first WWW conference

Tim Berners-Lee, Geneva, May 1994

To a computer, the Web is a flat, boring world, devoid of meaning. This is a pity, as in fact documents on the Web describe real objects and imaginary concepts, and give particular relationships between them. For example, a document might describe a person. The title document to a house describes a house and also the ownership relation with a person. Adding semantics to the Web involves two things: allowing documents which have information in machine-readable forms, and allowing links to be created with relationship values. Only when we have this extra level of semantics will we be able to use computer power to help us exploit the information to a greater extent than our own reading.



Metadata

- Data can not be associated if not well defined
- In a web context, metadata can describe *things* mentioned in a web page and not just the page itself
- Can provide a definitive description of something and not just supplementary information (as generally perceived)
- Is most useful when it is unambiguous and easily understood by *everyone*
- Could be extremely useful when understood by computers

The importance of metadata

- Metadata
- Increases specificity
- Improves common, wide, unambiguous understanding
- Content becomes computable



Compiling metadata



Working individually, prepare some metadata about familiar entities by completing the following table



Compiling metadata

Subject	Property	Value
me		
me		
me		
my organisation		
my organisation		
my organisation		
my home town		
my home town		
my home town		

Evolution of metadata

- The World Wide Web—HTML, HTTP, URLs
- Primitive metadata, some semantics about a document
- Use of meta tags in head of HTML document
- Need for standards and common vocabularies
- Early vocabularies
- Resource Description Framework
- RDF files and separate metadata records
- Issues with RDF
- Microformats and similar pragmatic solutions
- Embedded metadata
- RDFa



Resource Description Framework

- A model for describing the universe unambiguously
- RDF describes *things*, defined by uniform resource identifiers, URIs, by assigning properties (predicate) and corresponding values (object) in statements known as *triples* consisting of *[subject] [predicate] [object]*
- The predicate URI usually references a term from a standard set of metadata terms (vocabulary), resulting in unambiguous meaning



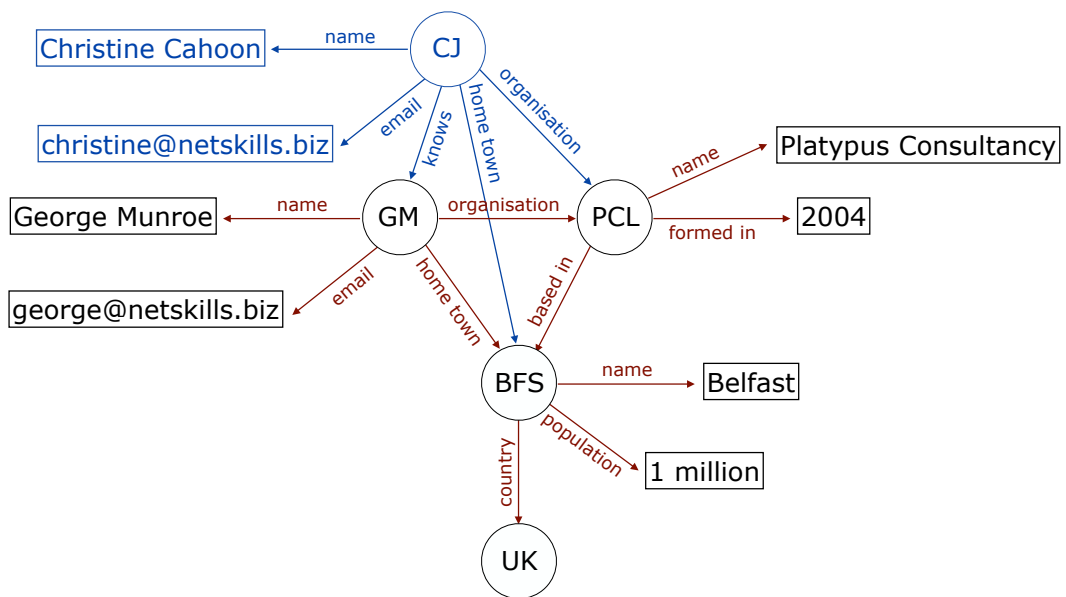
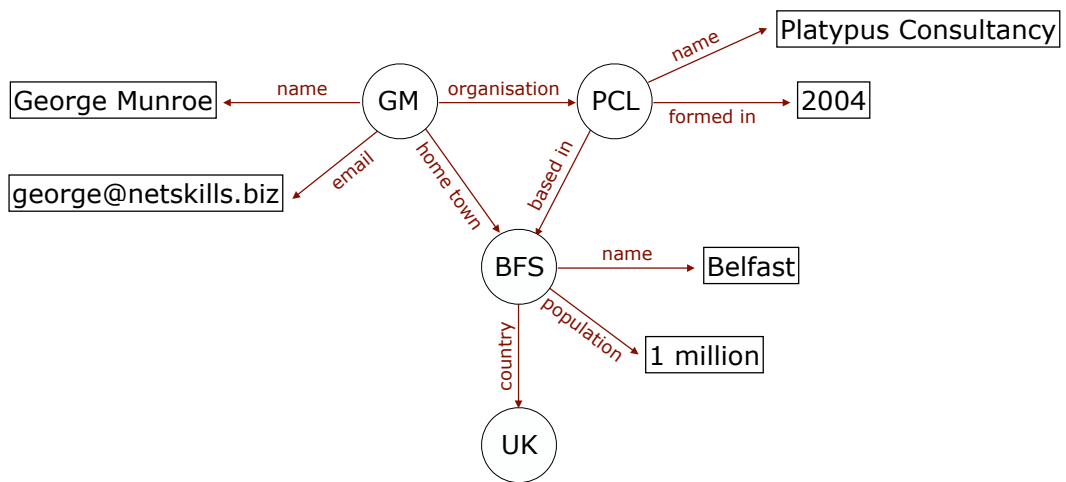
<http://www.w3.org/TR/rdf-primer/>

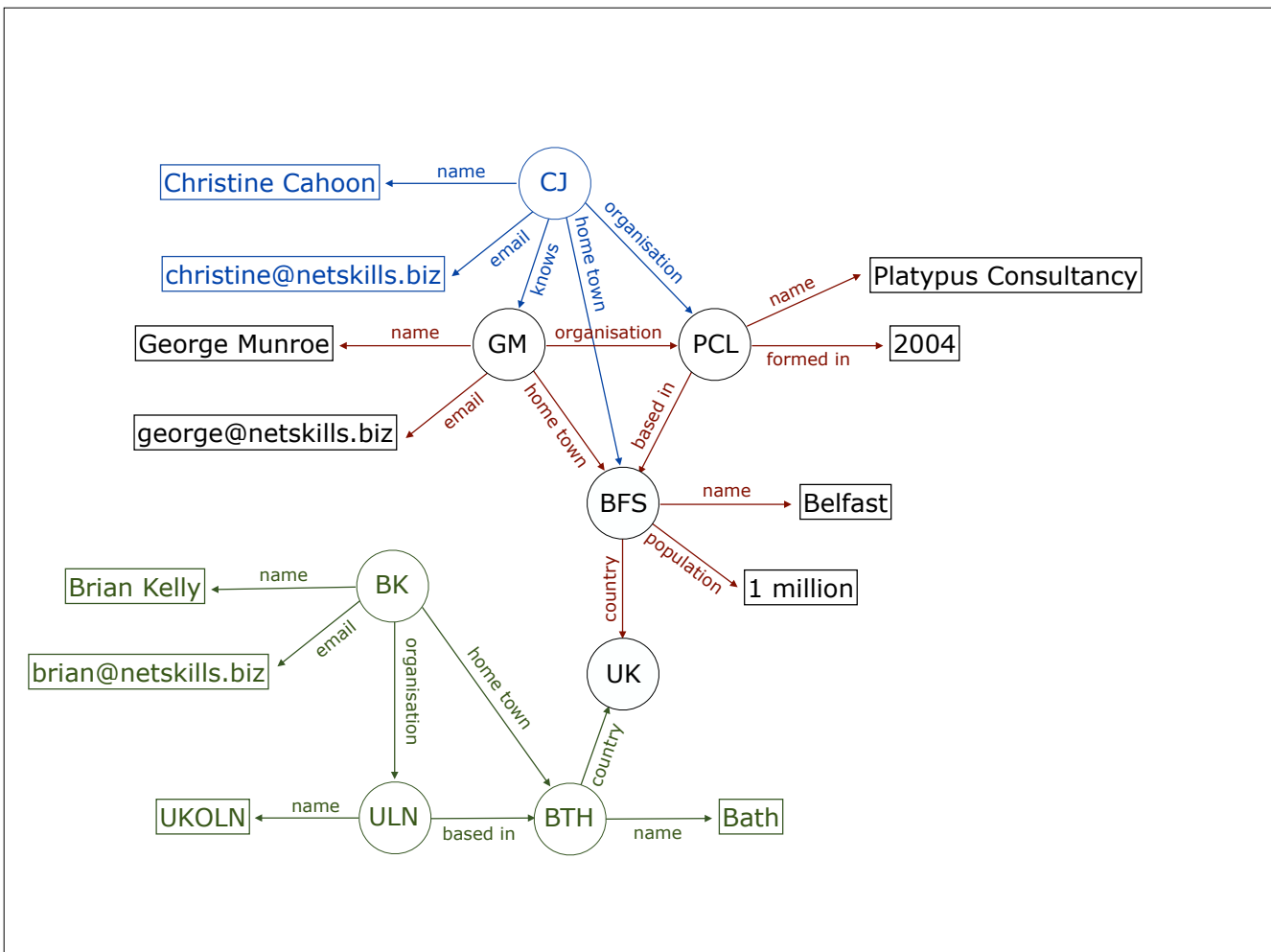
Simple metadata scheme

- Declare values of properties for a subject
- Works for anything
- Simplicity

Subject	Property	Value
me	name	"George Munroe"
me	organisation	my organisation
me	email address	"george@netskills.biz"
my organisation	name	"Platypus Consultancy"
my organisation	year formed	"2004"
my organisation	town/city	my home town
my home town	name	"Belfast"
my home town	country	UK
my home town	population	"1 million"

Subject	Property	Value
GM	name	"George Munroe"
GM	organisation	PCL
GM	email address	"george@netskills.biz"
PCL	name	"Platypus Consultancy"
PCL	year formed	"2004"
PCL	town/city	BFS
BFS	name	"Belfast"
BFS	country	UK
BFS	population	"1 million"

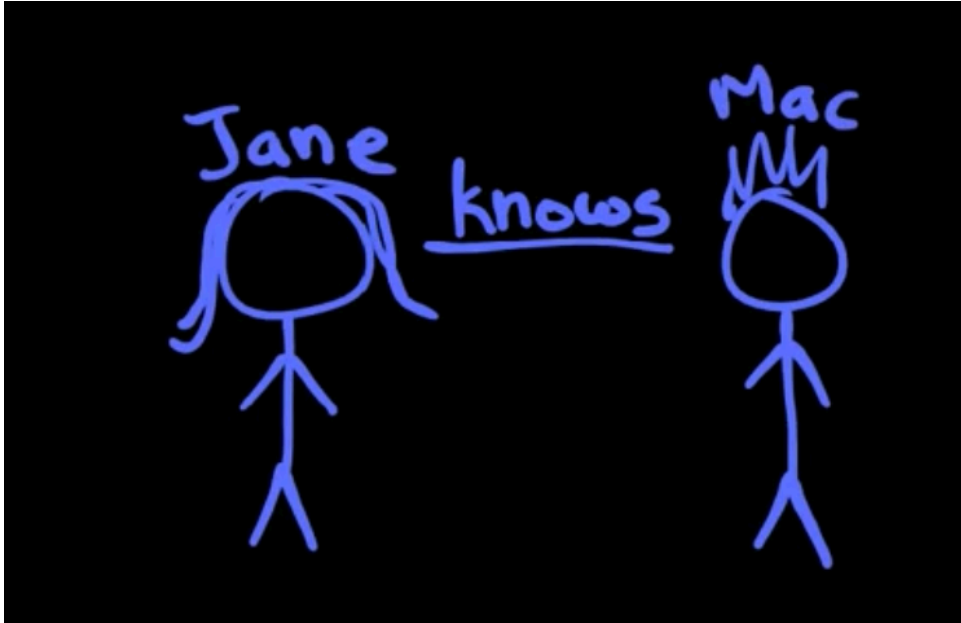




RDFa

- Previously RDF statements were usually provided in separate .rdf files and were not widely used because of the extra effort required to produce and link the files
- RDFa allows RDF statements to be included in ordinary HTML files using formally defined attributes within blocks, with metadata vocabularies referenced in the <head> block of the document
- The vocabularies are specified using XML namespaces, so mandating use with *XHTML* document types only

An RDFa basics tutorial by Manu Sporny



http://www.youtube.com/watch?v=ldl0m-5zLz4&feature=player_embedded

Using RDFa overview

- **Get HTML document headings right:**

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
"http://www.w3.org/Markup/DTD/xhtml-rdfa-1.dtd">
<html>
```

- **Use HTML tag attributes to identify triples:**

about—a URI or CURIE identifying the subject (usually in a <div>)

typeof—specifies the type (class) of the subject (usually in <div>)

property—specifies the predicate (usually in)

rel and *rev*—relationship or reverse-relationship predicate (in <a>, etc.)

href, *src* and *resource*—specifying object (usually in <a>, etc.)

content—overrides or supplements HTML content specifying the object

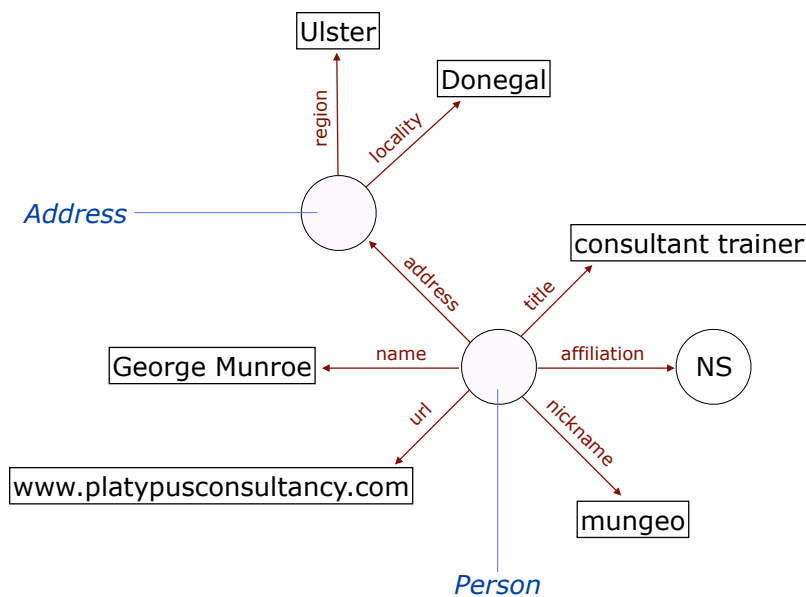
datatype—specifies a datatype for object (content)

<http://www.w3.org/TR/rdfa-syntax/#rdfa-attributes>

A simple RDFa example

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
"http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
xmlns:v="http://rdf.data-vocabulary.org/#">
<head profile="http://www.w3.org/1999/xhtml/vocab">
<title>Simple RDFa example</title>
</head>

<body>
<div xmlns:v="http://rdf.data-vocabulary.org/#" typeof="v:Person">
My name is <span property="v:name">George Munroe</span>,
also known online as <span property="v:nickname">mungeo</span>.
I am involved in several ventures but my home web site is at:
<a href="http://www.platypusconsultancy.com"
rel="v:url">www.platypusconsultancy.com</a>.
I live in
<span rel="v:address">
<span typeof="v:Address">
<span property="v:locality">Donegal</span>,
<span property="v:region">Ulster</span>
</span>
</span>
and work as a <span property="v:title">consultant trainer</span>
at <span property="v:affiliation">Netskills</span>.
</div>
</body>
</html>
```



RDFa support

“RDFa is an extension to HTML5 that helps you markup things like People, Places, Events, Recipes and Reviews. Search Engines and Web Services use this markup to generate better search listings and give you better visibility on the Web, so that people can find your website more easily”.

Linked Data
Linked Data is a relatively new way of thinking about how we publish data. At present, there is a great deal of data on the Web, but it is either locked down in a private database, or published in a way that is not very useful to computers. The data we have today is disconnected and lives in data silos. These data silos are not linked to one another, which goes against the core design of the Web. It is important that we fix both the data silo-ing and data linking problem. If we do so, computers will be able to use the data more easily and thus make our lives better as a result. **RDFa is a way of publishing Linked Data in HTML5 documents – it solves both the data silo-ing and data linking problem.**

[Learn more about Linked Data »](#)

Immediate Benefits
There are immediate benefits for integrating RDFa into your website today. By using RDFa to express [Google Rich Snippets](#) or [schemas.org](#), your web pages will be displayed in an enhanced format on all major search engines. You can also use Facebook's [Open Graph Protocol](#), which is based on RDFa, to express concepts that are contained in your web pages, like people, places, events, movies and recipes. **The search and social companies are serious about indexing RDFa content, which means that you may want to get serious about adding it into your pages before your competitors do.**

Using RDFa in HTML5
RDFa not only works in HTML5, but is a general technology for Linked Data markup in a variety of HTML-like languages. This means that it works well in XML documents, SVG documents, as well as HTML4, XHTML1 and XHTML5 documents. In fact, if your markup language is based on XML, then you can already use RDFa in your documents today. There is a version of RDFa for beginners called [RDFa Lite](#). The [full version of RDFa](#) is more expressive and allows Web authors to solve just about any structured data markup problem that they might throw at it. **Whatever popular Web document format you use, chances are that RDFa already works in it.**

[Learn more about RDFa Lite »](#)

Learn
There are a number of good reads available on this website covering everything from basic RDFa usage, to markup recipes, to advanced technical concepts.

[Learn more about RDFa »](#)

Tools
There are a variety of tools available for RDFa, ranging from online structured data preview tools, to validators, to processing tools and software libraries.

[Learn more about RDFa software »](#)

Developers
If you are a software developer, there are a number of tools and libraries that you can use in your software projects to extract and process RDFa markup in Web documents.

[Learn more about RDFa development »](#)

<http://rdfa.info/>

Google tools

- Tutorial on producing RDFa
- Description of Google provided vocabularies
- Testing tools for rich snippets information built using RDFa



<http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=146898>

Using RDFa and schema.org

Google indexing RDFa 1.0 + schema.org markup

Manu Sporny, The Beautiful, Tormented Machine, 14 February 2012

Evidence that Google recognises RDFa in web pages using the *schema.org* vocabulary that top search engines have agreed for common *things* in web pages. Metadata is used to enhance the search result presentation, providing a more useful immediate summary for the search engine user and increasing attraction to the address containing the embedded metadata.

<http://manu.sporny.org/2012/google-indexing-schema-rdfa/>

Vocabularies

- A particular set of metadata terms or properties, usually referring to a specific *thing*—i.e. a class or entity
- XML namespaces identify where a vocabulary may be found, usually with an HTTP dereference-able address (i.e. it can be accessed on the web)
- As well as the actual properties, the vocabulary contains descriptions of what those properties are

Ontologies

- In a semantic web everyone (and every computer) must have a common understanding of what particular things and related properties actually mean
- A real understanding involves being aware of the relationships between classes of things, as well as what properties can correspond to a particular class
- These relationships can be defined using additional vocabularies using a web ontology language—OWL—essentially a vocabulary for relationships
- A very complex relationship map or graph can be built quite simply from triples with the object of one triple as the subject of another triple

<http://www.w3.org/standards/semanticweb/ontology>

Commonly used metadata vocabularies

- Google (person, organisation, review, event, recipe)
<http://rdf.data-vocabulary.org/>
- FOAF (Friend Of A Friend)
<http://xmlns.com/foaf/0.1/>
- GoodRelations (ecommerce)
<http://www.heppnetz.de/ontologies/goodrelations/v1>
- Dublin Core (generic document)
<http://dublincore.org/specifications/>
- Creative Commons (licensing)
<http://creativecommons.org/ns>
- SKOS (Simple Knowledge Organisation System)
<http://www.w3.org/2004/02/skos/>

SPARQL

- SPARQL Protocol and RDF Query Language
- Used to access diverse data sources that contain RDF—all are treated as a single *database*
- Enables exploring, mining, combining RDF triples using a simple query language



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

Using SPARQL

- Requires a SPARQL endpoint as query engine to examine data and present results
- Typical SPARQL query using dbpedia data set

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?name
FROM <http://dbpedia.org/>
WHERE {
    ?person foaf:name ?name .
    GRAPH ?g1 { ?person rdf:type foaf:Person }
    GRAPH ?g2 { ?person rdf:type foaf:Person }
    FILTER (?g1 != ?g2) . }
```

- Some pre-set services available, for example:
<http://data.gov.uk/sparql>

<http://www.cambridgesemantics.com/2008/09/sparql-by-example/>

RDFa Developer

Firefox add-on

This firefox add-on lets you visualise all the RDFa triples in a web page. A list of errors and warnings found while parsing the document is shown. it is also possible to execute SPARQL queries on the RDFa content.



<https://addons.mozilla.org/en-US/firefox/addon/195085/>

RDFa distiller

- W3C service to identify and list RDF in a web page

- Extract RDF from HTML + RDFa

Using a web address, local file or direct text inputs this provides a clean view of the implied data hierarchy enabling a simple check on the markup and intended meaning.

Get RDF directly from <http://example.com/sample.html> by appending the following to the service address:

`/extract?uri=http://example.com/sample.html`



<http://www.w3.org/2007/08/pyRdfa/>

RDF validator and grapher

- W3C service for validating content containing RDF
- Lists the RDF triples contained in supplied input
- Produces optional graphical visualisation of the data model



<http://www.w3.org/RDF/Validator/>

Google's knowledge graph

Introducing the Knowledge Graph— *things, not strings*

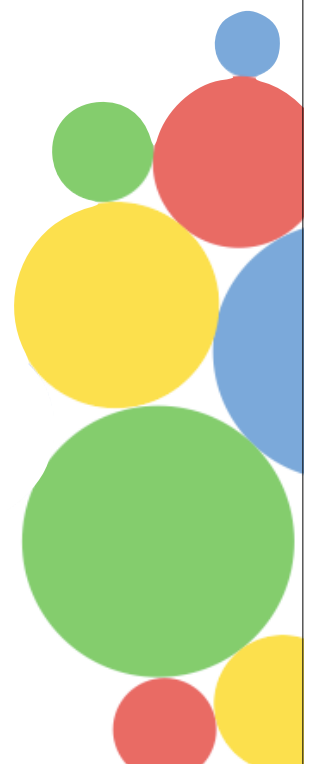
Amit Singhal, Google official blog, May 16, 2012

The Knowledge Graph enables you to search for things, people or places that Google knows about—landmarks, celebrities, cities, sports teams, buildings, geographical features, movies, celestial objects, works of art and more—and instantly get information that's relevant to your query. This is a critical first step towards building the next generation of search, which taps into the collective intelligence of the web and understands the world a bit more like people do.

Results:

- Find the right thing
- Get the best summary
- Go deeper and broader

<http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html>
<http://searchengineland.com/can-retailers-benefit-from-the-knowledge-graph-schema-you-bet-122676>



Facebook

The open graph protocol

Facebook Developers Conference, 21 April 2010

The Open Graph protocol enables you to integrate your web pages into the social graph. It is currently designed for web pages representing profiles of real-world things—things like movies, sports teams, celebrities, and restaurants. Once your pages become objects in the graph, users can establish connections to your pages as they do with Facebook Pages. Based on the structured data you provide via the Open Graph protocol, your pages show up richly across Facebook: in user profiles, within search results and in News Feed.



<http://developers.facebook.com/docs/opengraph>

Central Office of Information

RDFa and linked data in UK government web-sites

Mark Birbeck, Nodalities Magazine, 29 July 2009

The UK government's Central Office of Information had a straightforward problem to solve: how could they create a centralised web-site of information that the public could search and access, when the source of that information could be any government department database or any public sector web-site?

By using RDFa to address the challenge of making distributed data available in one place, the COI avoided having to make changes to each department's systems. But once each department is publishing RDFa, it becomes possible for third parties to consume that information however they see fit. Such a flexible architecture is crucial in the age of open government, and is a cornerstone of linked open data.



<http://blogs.talis.com/nodalities/2009/07/rdfa-and-linked-data-in-uk-government-web-sites.php>

The Stationery Office

TSO announces major new platform to accelerate open data drive

TSO partners with Garlik on hosted "trillion triple" RDF platform, 18 January 2010

TSO (The Stationery Office), the public sector division of Williams Lea, has today announced a partnership with Garlik, the leading semantic technology innovator, to launch what is believed to be the world's most scalable, securely hosted RDF platform for use by UK Central and Local Government departments. As the largest publisher in the UK of public sector documents (over 8,000 titles a year), TSO has taken this proactive step to provide its core public sector customers with the ability to participate with confidence in the Government's open data initiative.



<http://www.tso.co.uk/press/latestnews/archive/2010/triplestore/> <http://openup.tso.co.uk/des>

Microformats

- Microformats consist of informal metadata terms which are not referenced in the document but assumed to be known
- These terms have been established by rapid user adoption, ease of use and a desire to create richer semantics with embedded metadata
- They are used with *class* attributes in <div> and blocks or with *rel* attributes in anchor <a> tags



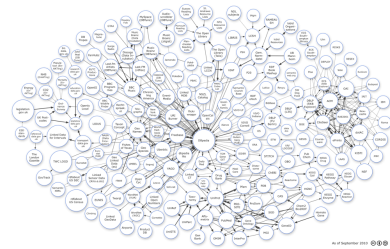
http://microformats.org/wiki/Main_Page

Linked data

A web of data

To make a web of data a reality, it is important to have the huge amount of data on the web available in a standard format, reachable and manageable by semantic web tools. Furthermore, not only does the semantic web need access to data, but relationships among data should be made available, too, to create a web of data and not just a massive collection of datasets. Interrelated datasets on the web are referred to as Linked Data.

Data can be simply linked when they are associated with a common unique identifier or more complexly linked when they are defined as a particular class and that class is known to be related to another particular class via a defined relation.



<http://www.w3.org/standards/semanticweb/data>

The linking open data project

Linking Open Data project

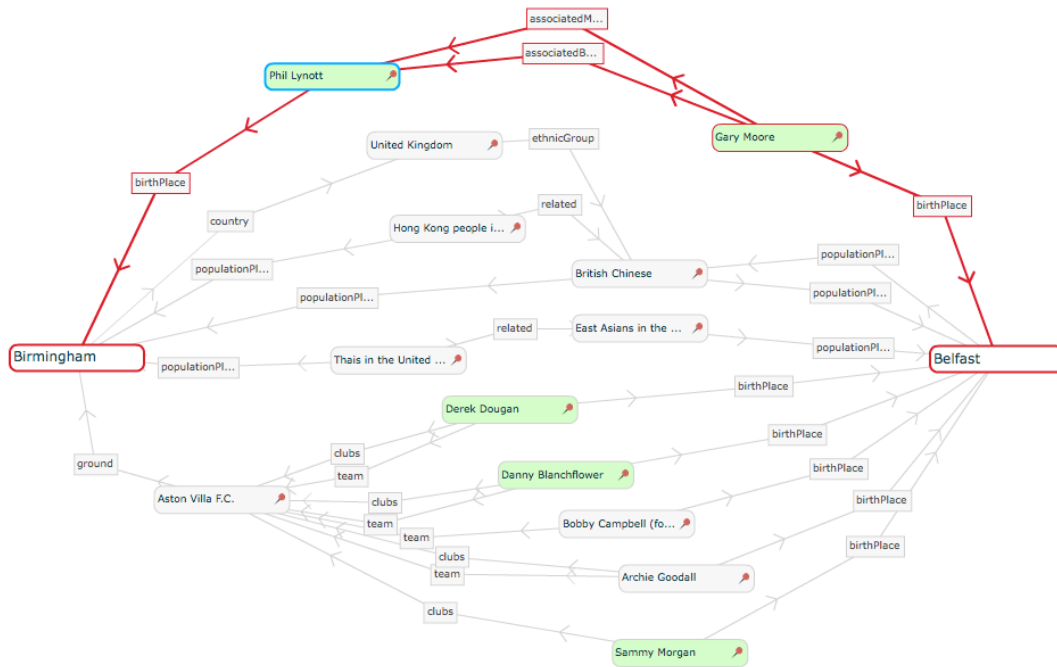
The goal is to extend the web by publishing various open data sets as RDF on the web and by setting RDF links between data items in different data sources. These RDF links then enable navigation from a data item within one data source to related data items within other sources using a semantic web browser.

RDF links can also be followed by the crawlers of semantic web search engines, which may provide sophisticated search and query capabilities over crawled data. As query results are structured data and not just links to HTML pages, they can be used within other applications.



<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

New discoveries!



<http://www.visualdataweb.org/relfinder/relfinder.php>

The *open data* movement



http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide.html

Open government

The currents of our time

Carl Malamud, Gov 2.0 Summit, 7-8 September 2010

If our government is to do the jobs with which we have entrusted it—if government is to ensure that the air we breathe and the water we drink are safe, or that every child is to be given a chance to flourish—if we are to accomplish these goals, the machinery of our government must be made to work properly...

Our federal government spends \$81.9 billion a year on Information Technology. Much of that is wasted effort. We build systems so badly, it is crippling the infrastructure of government.



<http://public.resource.org/currents/>

The great big database in the sky

Giant Global Graph

timbl's blog, 21 November 2007 (excerpts only)

Now, people are making another mental move. There is realisation now, "It's not the documents, it is the things they are about which are important". Obvious, really.

We can use the word Graph, now, to distinguish from Web.

I called this graph the Semantic Web, but maybe it should have been Giant Global Graph! Any worse than WWW? ;-). Now the "Semantic Web" term has been established for a long time, I'm not proposing to change it. But let's think about the graph which it is.

So, if only we could express these relationships, such as my social graph, in a way that is above the level of documents, then we would get re-use.

That's just what the graph does for us. We have the technology—it is Semantic Web technology, starting with RDF, OWL and SPARQL. Not magic bullets, but the tools which allow us to break free of the document layer.



<http://dig.csail.mit.edu/breadcrumbs/node/215>

Call a spade a spade



photo: ©GM 2010