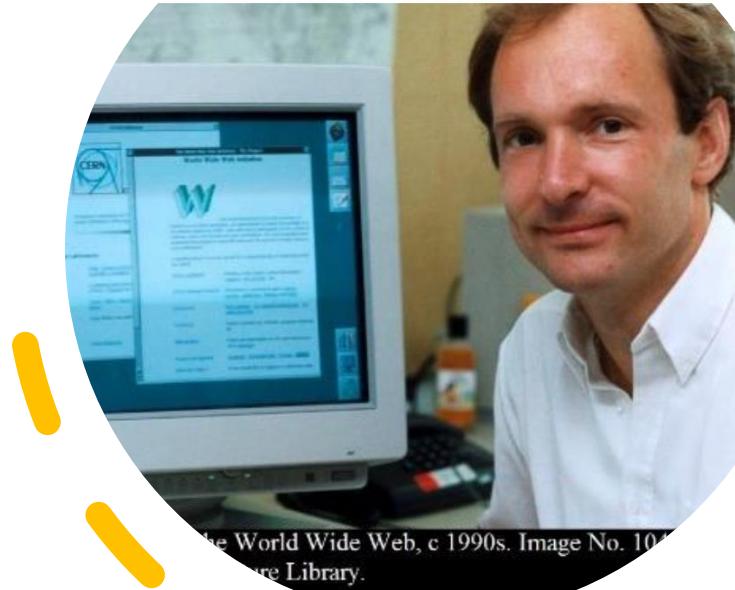


Semantic Web & Knowledge Graphs

Nature of the web

- Invented by T. Berners-Lee in 1989 at CERN to provides rapid electronic access to reports (documents) created by the high-energy physics lab
- From its origin, the world wide web is a **web of documents** written in human readable format



The World Wide Web, c 1990s. Image No. 10
Source: CERN Photo Library.

The screenshot shows the W3C website homepage from 2004. The header features the W3C logo and the tagline "Leading the Web to Its Full Potential...". The main navigation menu includes "Local Reports", "Site Index", "New Visitors", "About W3C", "Join W3C", and "Contact W3C". Below the menu, a banner for the "2004-09-02: The 26th Internationalization & Unicode Conference" is displayed. The page is filled with various links to W3C's work on topics like HTML, CSS, XML, and RDF. A blue arrow points from the text "" to a specific link on the left sidebar under the heading "Internationalization & Unicode Conference".

Nature of the web

Two important services of web:

- Providing a convenient means to retrieve and view information
- Providing a language "HTML" describing to computers how to display documents written in this language.
- While keeping this same nature, It evolved to
 - Social networks (blogging, sharing, chatting...)
 - Retail (buying, selling, advertising...)
 - Administration (registering, obtaining admin documents, paying taxes...)
 - Education (learning, cooperating...)
 - Entertainment (viewing video, listening to music...)
 - Media (reading newspapers, finding news...)

Limitations of the web of documents

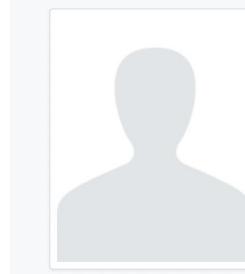
- It is meant for human to share documents
- A human reader can browse through the contents and make sense of them
- A machine is unable to make sense of the web documents because they lack context
- Human can figure out the context
- Machine cannot
- So even if the machine had a software that is able to perfectly recognise text, it cannot make use of the information contained in web documents

Solutions to handle limitations:

1. To take the web as it is currently implemented, and to **use Artificial Intelligence techniques** to analyze the content of web pages in order to provide an interpretation of its meaning
Problem: This approach would require validation. Furthermore, the rate at which the web is growing would render it practically impossible to achieve.
2. To represent the web pages in a **common representation form** in which computers can represent and interpret the data easily **and to use intelligent techniques** to take advantage of these presentations. This is the approach of the **Semantic Web**.

Example

- Typing “Radu Mihailescu” in google search provides implicit information
- Human reader can “understand” that it is a human; whose name is Radu Mihailescu; he works at Heriot-Watt university; he has a list of research domains...
- A machine lacking the context cannot do the same
- How can we add such context to make the information **explicit and readable by a machine?**
- Use additional **metadata**, that describe each piece of information



View Scopus Profile

Radu-Casian Mihailescu
Associate Professor, School of Mathematical & Computer Sciences

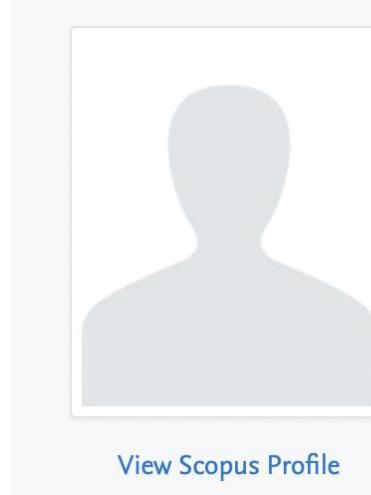
Email
R.Mihailescu@hw.ac.uk

EH14 4AS
United Kingdom

```
<human/>  
<male/>  
<name>Radu Mihailescu</name>
```

Possible metadata

```
<human/>
<male/>
<name>Radu Mihailescu</name>
<academic/>
<Work at>
  <Organisation>
    <type>university</type>
    <name>Heriot-Watt</name>
  </Organisation>
<Researcher>
<Research domains>
  <domain>Machine learning</domain>
  <domain>Deep Learning</domain>
  <domain>Artificial Intelligence</domain>
  <domain>Optimisation</domain>
</research domains>
<research profile>
  <link>https://scholar.google.com/Radu Mihailescu/</link>
<research profile>
```



[View Scopus Profile](#)

Radu-Casian Mihailescu

Associate Professor, [School of Mathematical & Computer Sciences](#)

Email

R.Mihailescu@hw.ac.uk

EH14 4AS

United Kingdom

Machine would connect to other data

Heriot-Watt University
Website Directions Save
University in Dubai

Heriot-Watt University Dubai Campus is a satellite campus of Heriot-Watt University based in Dubai, United Arab Emirates. Established in 2005, it was the first campus of an overseas university to open in Dubai International Academic City. [Wikipedia](#)

Address: Al Sufouh - Knowledge Park - Dubai
Hours: Closed - Opens 9AM Wed+
Phone: 04 571 7000
Number of students: 3,700
Founded: 2005
Principal: Professor Richard Williams
Chancellor: Robert M Buchan

Works at

Such web would be a web of data!

Radu-Casian Mihailescu

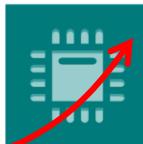
Associate Professor, [School of Mathematical & Computer Sciences](#)

Email
R.Mihailescu@hw.ac.uk

EH14 4AS
United Kingdom

	TITLE	POSTED BY	YEAR
<input type="checkbox"/>	Dynamic coalition adaptation for efficient agent-based virtual power plants RC Mihailescu, M Vasirani, S Ossowski Multiagent System Technologies: 9th German Conference, MATES 2011, Berlin ...	49	2011
<input type="checkbox"/>	Towards machine learning explainability in text classification for fake news detection L Kurasinski, RC Mihailescu 2020 19th IEEE international conference on machine learning and applications ...	28	2020
<input type="checkbox"/>	Collaborative sensing with interactive learning using dynamic intelligent virtual sensors A Tegen, P Davidsson, RC Mihailescu, JA Persson Sensors 19 (3), 477	27	2019
<input type="checkbox"/>	Towards collaborative sensing using dynamic intelligent virtual sensors RC Mihailescu, J Persson, P Davidsson, U Eklund Intelligent Distributed Computing X: Proceedings of the 10th International ...	17	2017

Has research profile



sensors

an open access journal by MDPI

Published by

Semantic web

- The **Semantic Web** is a ‘web of data’ that facilitates machines to understand the semantics, or meaning, of information on the WWW. It extends the network of hyperlinked human-readable web pages by inserting machine-readable metadata about pages and how they are related to each other, enabling automated agents to access the Web more intelligently and perform tasks on behalf of users

T. Berners Lee

Berners-Lee is the director of the [World Wide Web Consortium](#) (W3C), which oversees the continued development of the Web.

At W3C, the activities of **Semantic Web** are taken by the **Web of Data** activities. The goal of the is to make Internet data machine-readable. This is done by Encoding **semantics** in the data using W3C technologies such as **RDF, OWL...**

TED Talk – Tim Berners Lee

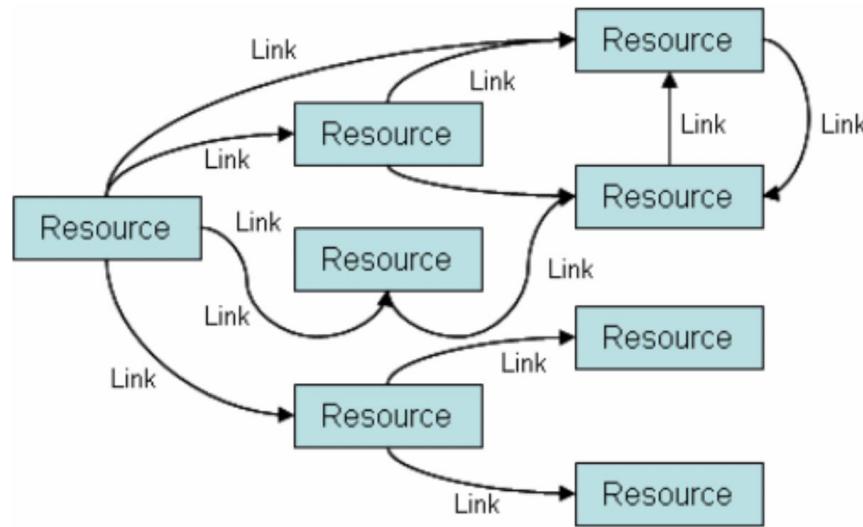


Semantic web = web of data

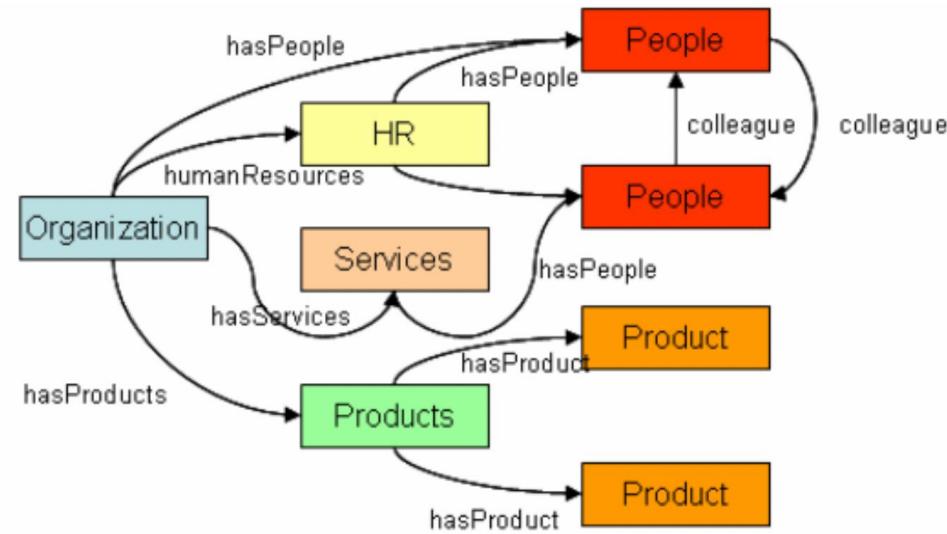
- Extension of WWW with additional **metadata** associated with all pieces of information (data, links...)
- This additional information gives the **meaning** of the data and **links** it to other data
- The additional information should be **machine readable**
- It must be **standardised**
 - The terms, the format...

The web of data would be a **knowledge graph**

Using graph representation



Syntactic Web



Semantic Web

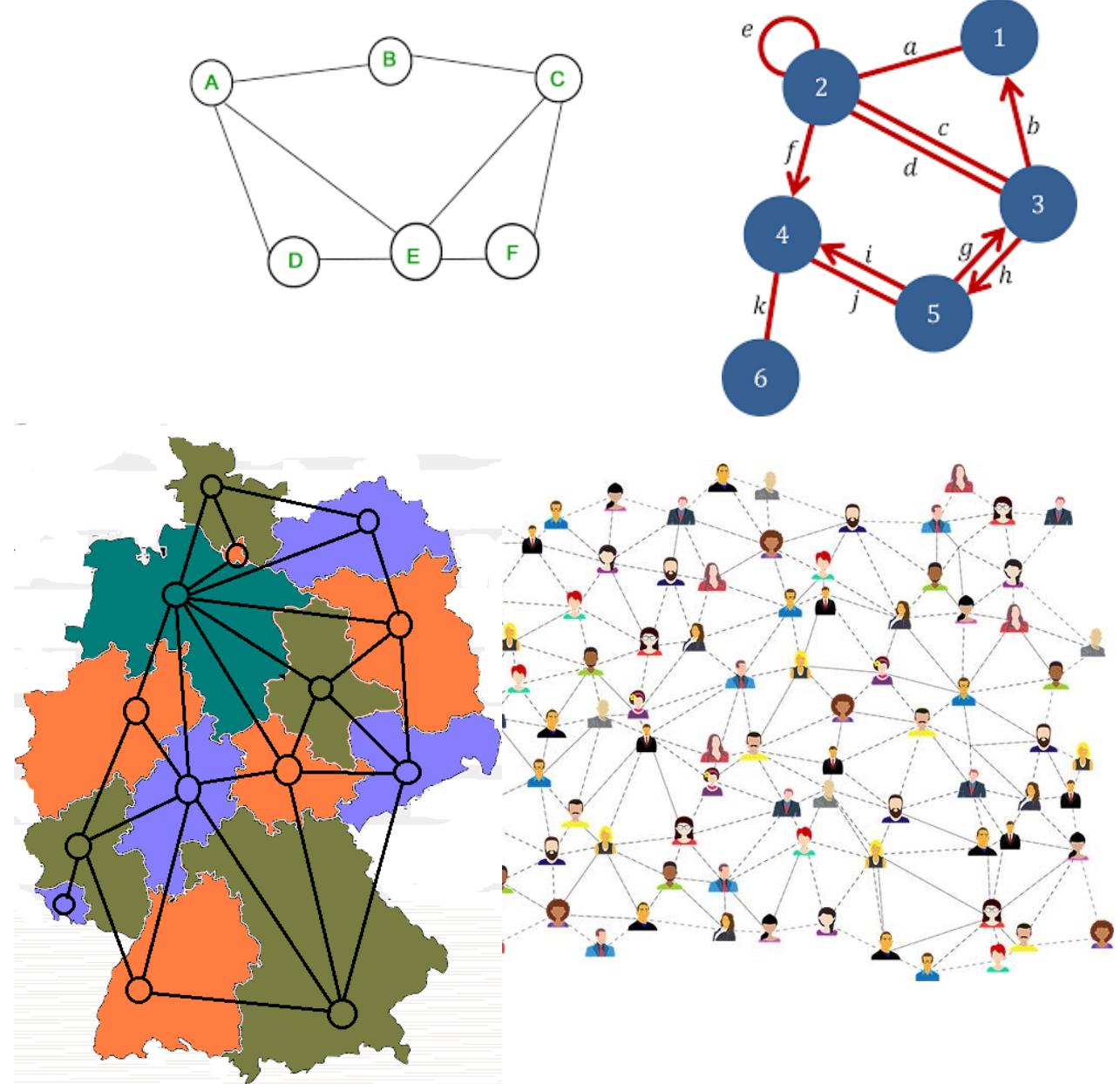
- The "syntactic" web is about documents while the semantic web is about “things” concepts we are interested in (people, places, events etc.), and the relationships between these concepts.

Knowledge graph

- A knowledge graph is a knowledge base in the form of a graph
- A knowledge base is
 - A technology to store complex structured or unstructured information used by a computer system (Wikipedia)
 - A collection of knowledge expressed using some formal knowledge representation language (free online dictionary of computing)
 - A store of information or data that is available to draw on; the underlying set of facts, assumptions, and rules which a computer system has available to solve a problem (Google Dictionary)

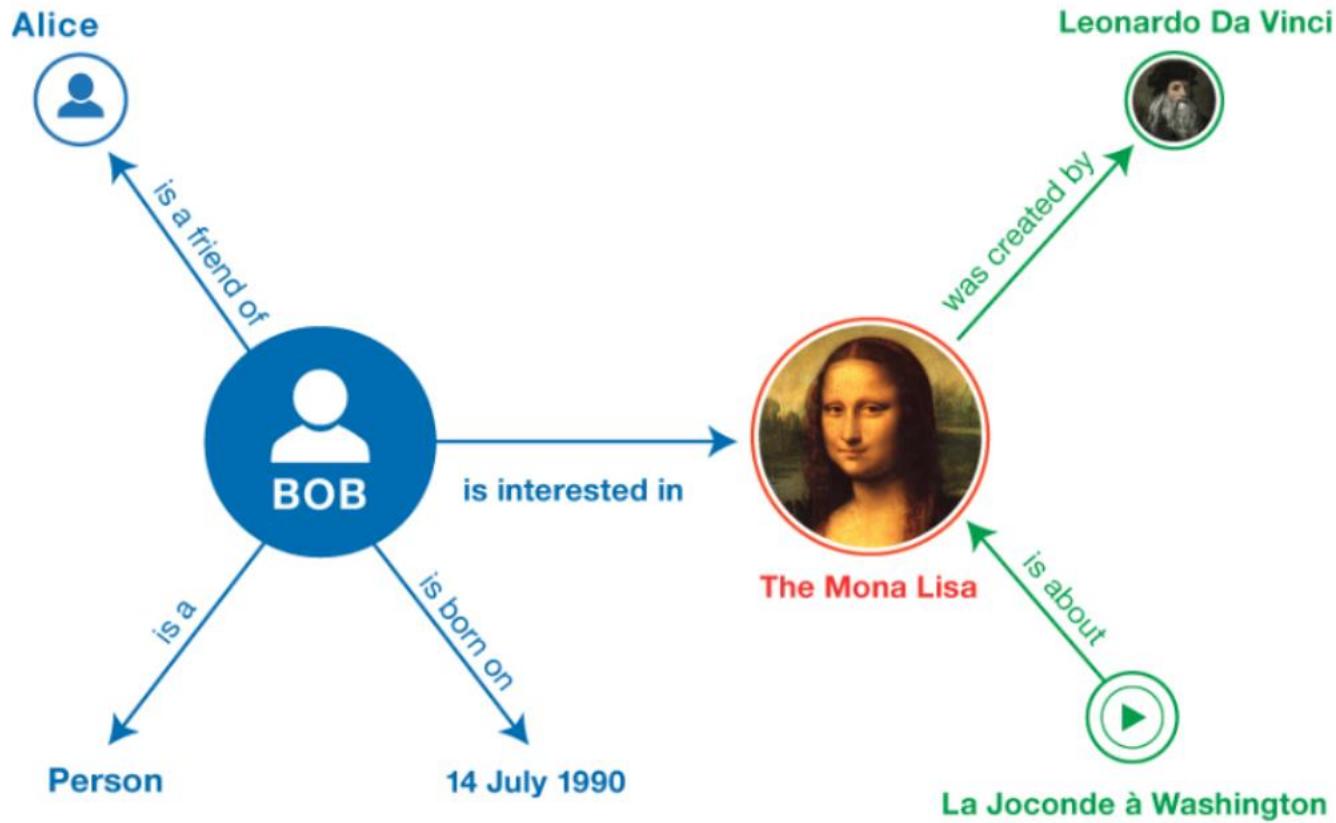
What is a graph?

- A set of vertices (or nodes) and edges connecting nodes.
- It can be directed or undirected
- Nodes can include any information
- Edges can also have labels and properties (weights...)
- Used to represent data in various domains

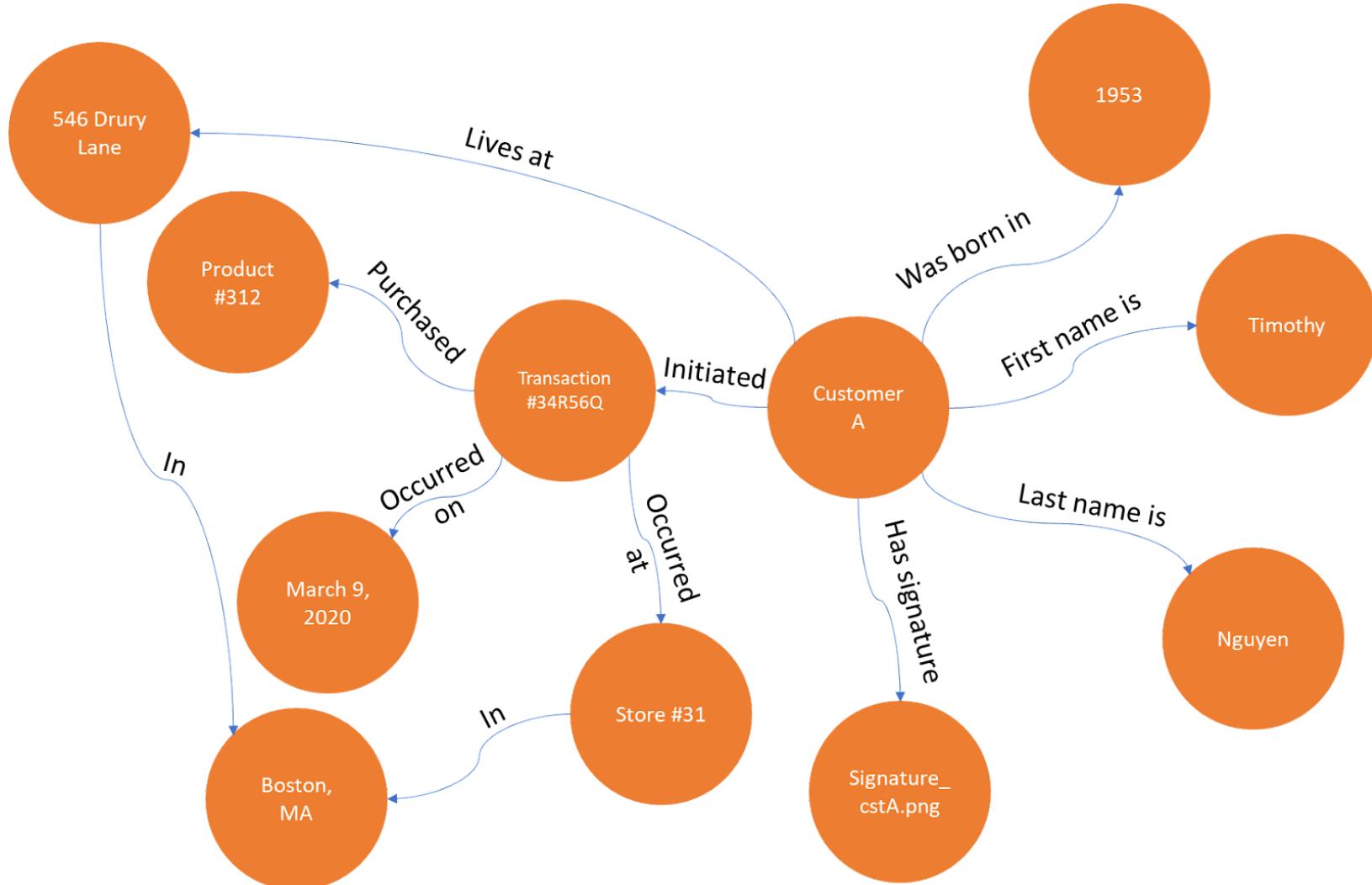


Example knowledge graph

Directed, labelled graph

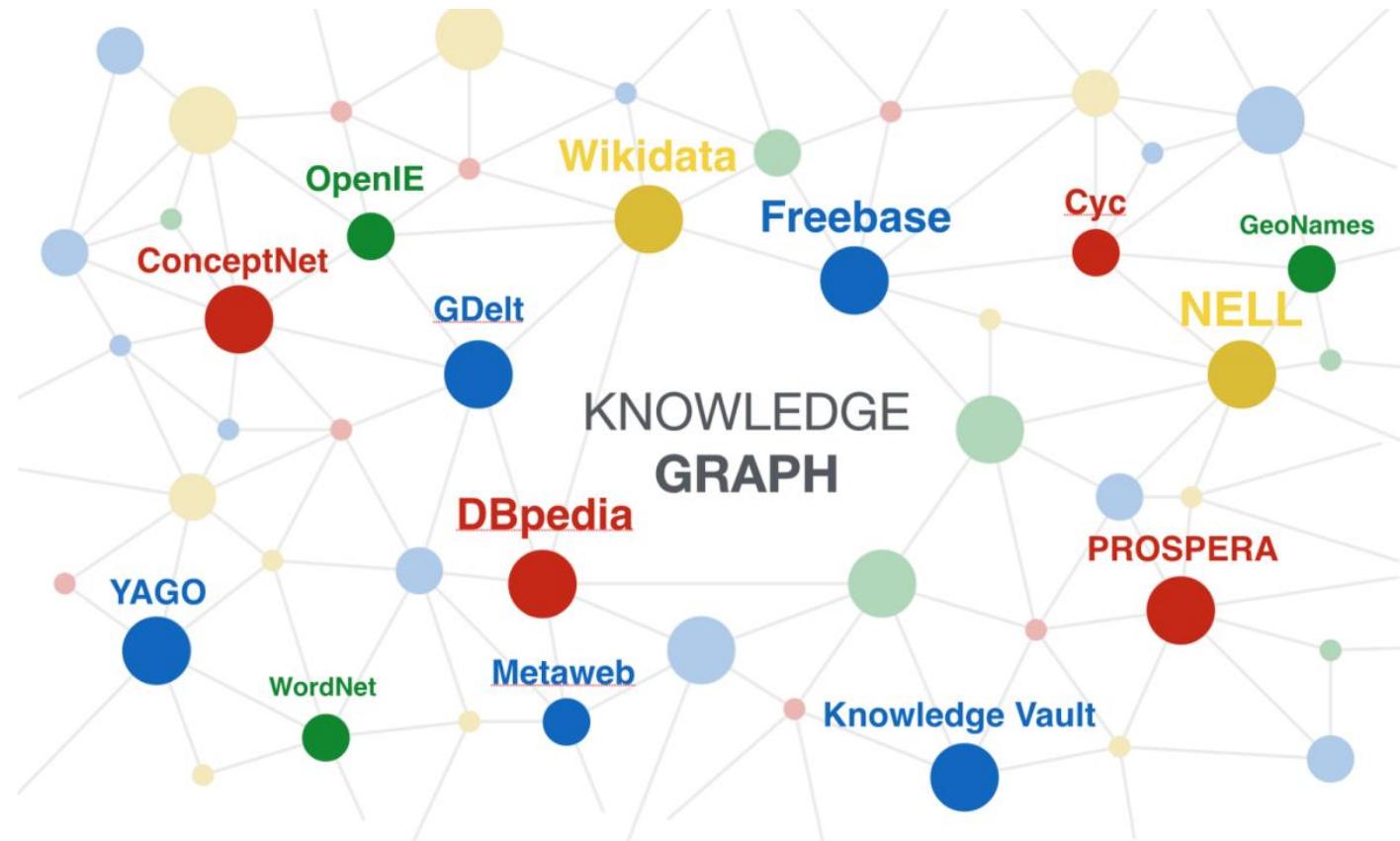


A customer in knowledge graph



Who uses knowledge graphs

“Major companies such as Google, Yahoo!, Microsoft, and Facebook have created their own ‘knowledge graphs’ that power semantic searches and enable smarter processing and delivery of data. The use of these knowledge graphs is now the norm rather than the exception” (ISWC 2014)



Applications of knowledge graphs

Application	Example KG
Semantic search	Microsoft Satori, Google Knowledge Vault
Competitive analytics	Quid, RecordedFuture
Reputation analytics	Bottlenose, Cleareye
Integrate productivity apps	Microsoft Graph
AI assistants	Amazon Information Graph
Intelligence in defence and security	Palantir, Primer

Companies create various types of knowledge graphs

- Customer graphs
- Employee graphs
- CRM graphs
- Contents graphs
- Machines graphs

Example 1: Google knowledge graph in search

The screenshot shows a Google search result for 'Florence Price'. A red box highlights the 'Florence Price' Knowledge Graph panel at the top right of the page. This panel includes a large portrait of Florence Price, a grid of smaller images, a title 'Florence Price American composer', and tabs for 'OVERVIEW', 'VIDEOS', 'SONGS', 'ALBUMS', 'LISTEN', and 'PEOPLE ALSO KNEW'. Below the panel are sections for 'Available on' (Spotify, YouTube, Pandora, and more), 'About' (biography, birth date April 9, 1887, death date June 3, 1953, education at The University of Chicago and New England Conservatory of Music, children Edith Cassandra Price, Thomas Jr., and Florence Price Robinson, albums Symphonies and Violin Concertos), and 'Top results' (link to Wikipedia). Red arrows point from the 'About' section and the 'Songs' section below to their respective components in the Knowledge Graph panel. The main search results list includes 'Dances in the Canebrakes', 'Fantasie Negre', and 'Song To The Dark Virgin'.

- Information panel generated from Google Knowledge Graph
- Provides overview of concept searched for
- Can answer the user's information need without accessing and reading web pages

Example 2: music on the BBC site

BBC – Music – Eric Clapton
http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294 BBC Music

Netvibes Feedly Social Private Mailing lists SW Python RDFa it! Bookmarklets Add Zemanta bit.ly To Mendeley TinyURL To Faviki Dokuwiki

MUSIC BETA GENRES ARTISTS REVIEWS NEWS BLOG QUICK FIND Enter an artist name ...

BBC Music > Artists > Eric Clapton

Eric Clapton
Born 30 March 1945.



David Redfern/Redferns

Biography

Eric Patrick Clapton, CBE (born 30 March 1945) is an English blues-rock guitarist, singer, songwriter and composer. Clapton has been inducted into the Rock and Roll Hall of Fame as a solo performer, as a member of rock bands; the Yardbirds and Cream. Clapton is the only person ever to be inducted three times. Often viewed by critics and fans alike as one of the most important and influential guitarists of all time, Clapton was ranked fourth in Rolling Stone magazine's list of the "100 Greatest Guitarists of All Time" and #53 on their list of the Immortals: 100 Greatest Artists of All Time.

MOST PLAYED ON BBC RADIO 2

Latest Tracks Played On The BBC

- Promises BBC Radio 2 | Ken Bruce 22/02/2010
- Bad Love BBC Radio 2 | Alex Lester 22/02/2010
- Lay Down Sally BBC Radio 2 | Chris Evans Breakfast 18/02/2010
- I Ain't Gonna Stand For It BBC Radio 2 | Alex Lester 15/02/2010
- Wonderful Tonight BBC Radio 2 | Ken Bruce 10/02/2010

Audio Previews From Latest Album Review



- Me And Mr Johnson
- 8 Milkcow's Calf Blues
- 10 Come on in My Kitchen

Music on the BBC site (contd)

BBC – Music – Eric Clapton

http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

BBC Music

Netvibes ▾ Feedly Social ▾ Private ▾ Mailing lists ▾ SW ▾ Python ▾ RDFa it! ▾ Bookmarks ▾ Add Zemanta bit.ly To Mendeley TinyURL To Faviki Dokuwiki

this focus, he is credited as an innovator in a wide variety of genres. These include blues-rock (with John Mayall & the Bluesbreakers and The Yardbirds) and psychedelic rock (with Cream). Clapton's chart success was not limited to the blues, with chart-toppers in Delta Blues (Me and Mr. Johnson), pop ("Change the World") and reggae (Bob Marley's "I Shot the Sheriff") (He is often credited for bringing reggae and Bob Marley to the mainstream.) Two of his most successful recordings were the hit love song "Layla", which he played with the band Derek and the Dominos, and Robert Johnson's "Crossroads", which has been his staple song since his days with Cream.

Read more at Wikipedia...

WIKIPEDIA This entry is from Wikipedia, the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the GNU Free Documentation License. If you find the biography content factually incorrect, defamatory or highly offensive you can edit this article at Wikipedia. Find out more about our use of this data.

Links & Information

LINKS

- Official homepage at ericclapton.com
- Fanpage at whereseric.com
- Wikipedia article on [Eric Clapton](#)
- MySpace at myspace.com/ericclapton
- Last.fm page on [Eric Clapton](#)
- MusicBrainz entry on [Eric Clapton](#)

MEMBER OF Derek and the Dominos, [Blind Faith](#) (1968–1969), [Cream](#) (1966–1968), John Mayall & The Bluesbreakers (1965–1966), The Yardbirds (1963–1965)

COLLABORATED ON J.J. Cale & Eric Clapton, Eric Clapton & The Immediate All Stars, Eric Clapton & The Impressions, Eric Clapton & Jimmy Page, Eric Clapton & David Sanborn, Eric Clapton & Stan Webb's Chicken Shack, Eric Clapton & The Powerhouse, Eric Clapton & Stevie Ray Vaughan, Eric Clapton & Marc Shaiman, The Dirty Mac, Bob Dylan, Roger McGuinn, Tom Petty, Neil Young, Eric Clapton & George Harrison, Jimmie Vaughan, Eric Clapton, Bonnie Raitt, Robert Cray, B.B. King, Buddy Guy, Dr. John & Art Neville, Elton John & Eric Clapton, Michael Kamen, Eric Clapton and David Sanborn, B.B. King & Eric Clapton, Mark Knopfler & Eric Clapton, Paul McCartney & Eric Clapton, Sting with Eric Clapton, Steve Winwood & Eric Clapton

Links & information come from [MusicBrainz](#). You can add or edit information about Eric Clapton at musicbrainz.org. Find out more about our use of this data. The BBC is not responsible for the content of external sites

Latest News Stories

Played By

Since December 2008

Alex Lester BBC Radio 2
Steve Wright in the Afternoon BBC Radio 2
Sarah Kennedy BBC Radio 2
Wake Up to Wogan BBC Radio 2
Ken Bruce BBC Radio 2
Steve Wright's Sunday Love Songs BBC Radio 2
Jeremy Vine BBC Radio 2

Information displayed about artists played on BBC programmes is incomplete. Find out more about this artist play count information.

Example 3: community newspaper from government data

Data.gov.uk Newspaper | Newspaper Club

RSS Dogpile

Netvibes Feedly Social Private Mailing lists SW Python RDFa it! Bookmarks Add Zemanta bit.ly To Mendeley TinyURL To Favik

Newspaper Club

Data.gov.uk Newspaper

Friday, October 16th 2009

Over the last three days we've been working on a side project. A design exercise if you like.

We've been thinking about the beta [Data.gov.uk](#) repository, and wanted to explore putting some of the information contained within into people's hands in a form that is accessible, timely, and relevant.

And perhaps unsurprisingly, we thought a good way to do that was with a newspaper. So here it is, the Postcode Paper:



It's a prototype of a service for people moving into a new area. In our exercise we imagined you might receive it after paying your council tax for the first time. It gathers information about your area, such as local services, environmental information and crime statistics.

This is a post by Tom from the **Newspaper Club** Blog. File under [case studies](#).

We're building a service to help people make their own newspapers. This is the blog where we're alarmingly honest about where it's all going wrong. And occasionally smug about where it's going right.

WE'RE IN ALPHA

You can stick your name on the beta invite list [here](#).

Search

[Subscribe via RSS](#)

Archives

[January 2010](#)
[December 2009](#)
[November 2009](#)
[October 2009](#)
[September 2009](#)
[August 2009](#)
[July 2009](#)
[June 2009](#)

Filed under

[4ip \(3\)](#)
[art \(8\)](#)
[case studies \(9\)](#)
[engineering \(5\)](#)
[fulfillment etc \(2\)](#)
[investors \(2\)](#)
[media \(1\)](#)
[printers \(6\)](#)
[Uncategorized \(31\)](#)



Figures from knowledge graphs

Google Knowledge graph

- 500 billion facts about 5 billion entities
- Access
 - Limited through API
- Proprietary access
- Managed internally

Wikidata

- 71 million facts
- Access
 - API
 - SPARQL queries (SPARQL endpoint)
 - Download
- Free and open access (creative commons zero)
- Community curated

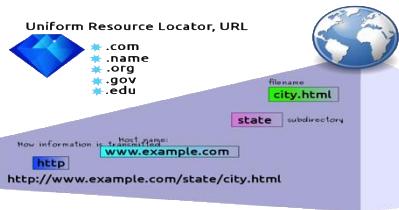
Benefits of knowledge graph representation

- Connect data from different sources
- Publish structured data to allow it to be interlinked and become useful through semantic queries
- Achieve the objective of Linked data (web of data)

Linked data principles

- Use IRI to uniquely identify things (entities)
- Use HTTP IRI to resolve ID
- Provide data through content negotiation
 - HTML for human
 - RDF for machines
- Link to other entities (data)

Identifiers: IRIs



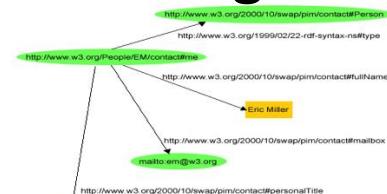
https://en.wikipedia.org/wiki/URL#/media/File:Uniform_Resource_Locator.svg

Protocol: HTTP



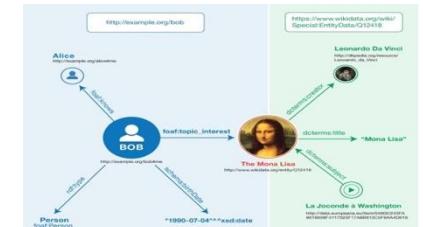
<https://en.wikipedia.org/wiki/File:Internet1.svg>

Encoding: RDF



https://en.wikipedia.org/wiki/Resource_Description_Framework#/media/File:Rdf_graph_for_Eric_Miller.png

Links: Data reuse



<http://www.w3.org/TR/rdf11-primer/>

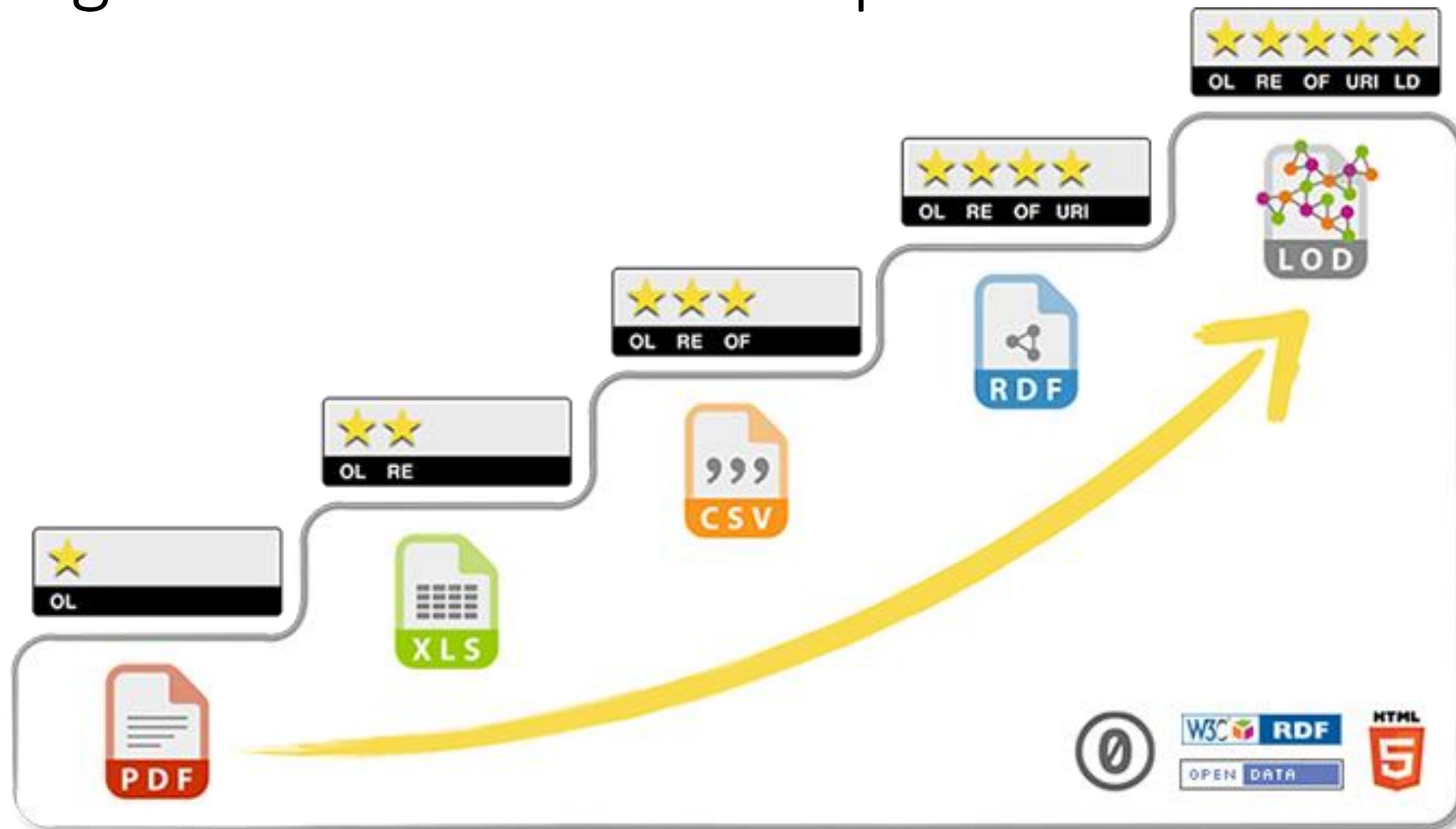
Five-star linked open data

- **Linked Open Data** is data that can be freely used and distributed, and that follows a set of design principles for sharing machine-readable interlinked data on the Web
- T. Berners Lee suggested a 5-star deployment scheme
 - Data gets more stars when proprietary formats are removed, and links are added

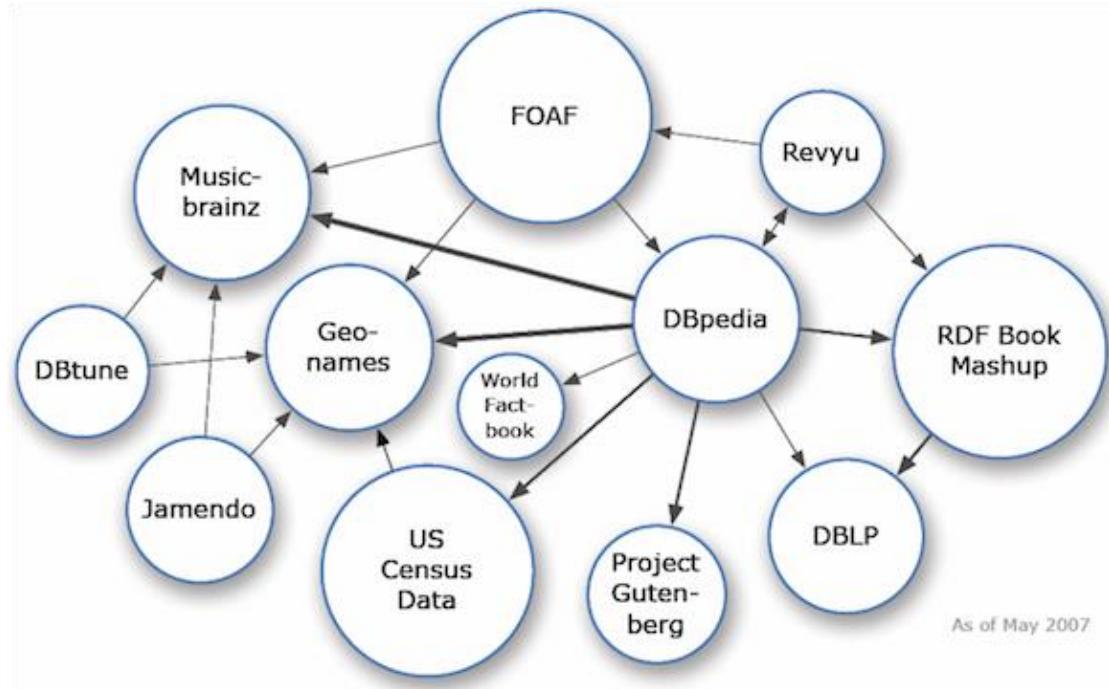
The five stars

stars	availability	format	license	Link
One star	data available online	Whatever format	Open license (OL)	Not linked
Two stars	Machine readable (RE) structured data	Processed with proprietary software	Open license (OL)	Not linked
Three stars	Data is available in open formats (OF)	No requirement for proprietary software	Open license (OL)	Not linked
Four stars	Data in open standards (IRI, RDF)	Query through SPARQL possible	Open license (OL)	
Five stars	W3C standards	Semantic access to data possible	Open license (OL)	Linked

Stages to achieve linked open data

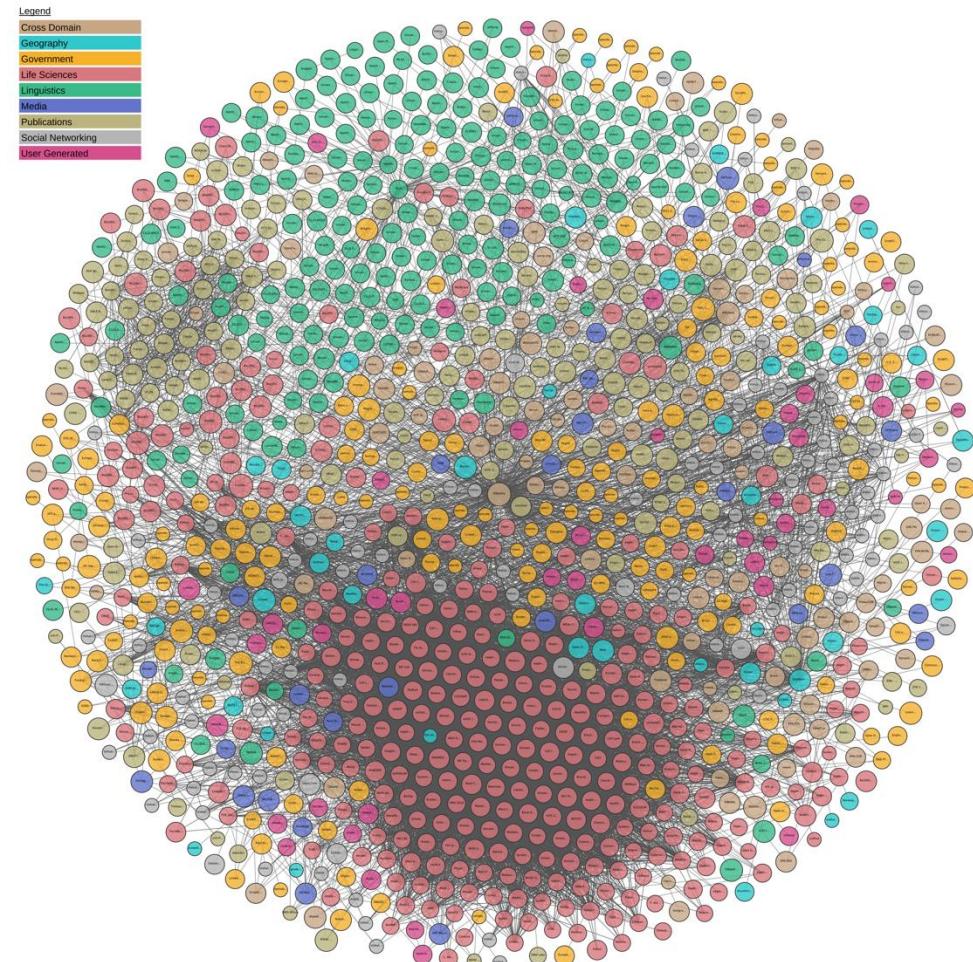


Evolution of LOD (2007-2019)

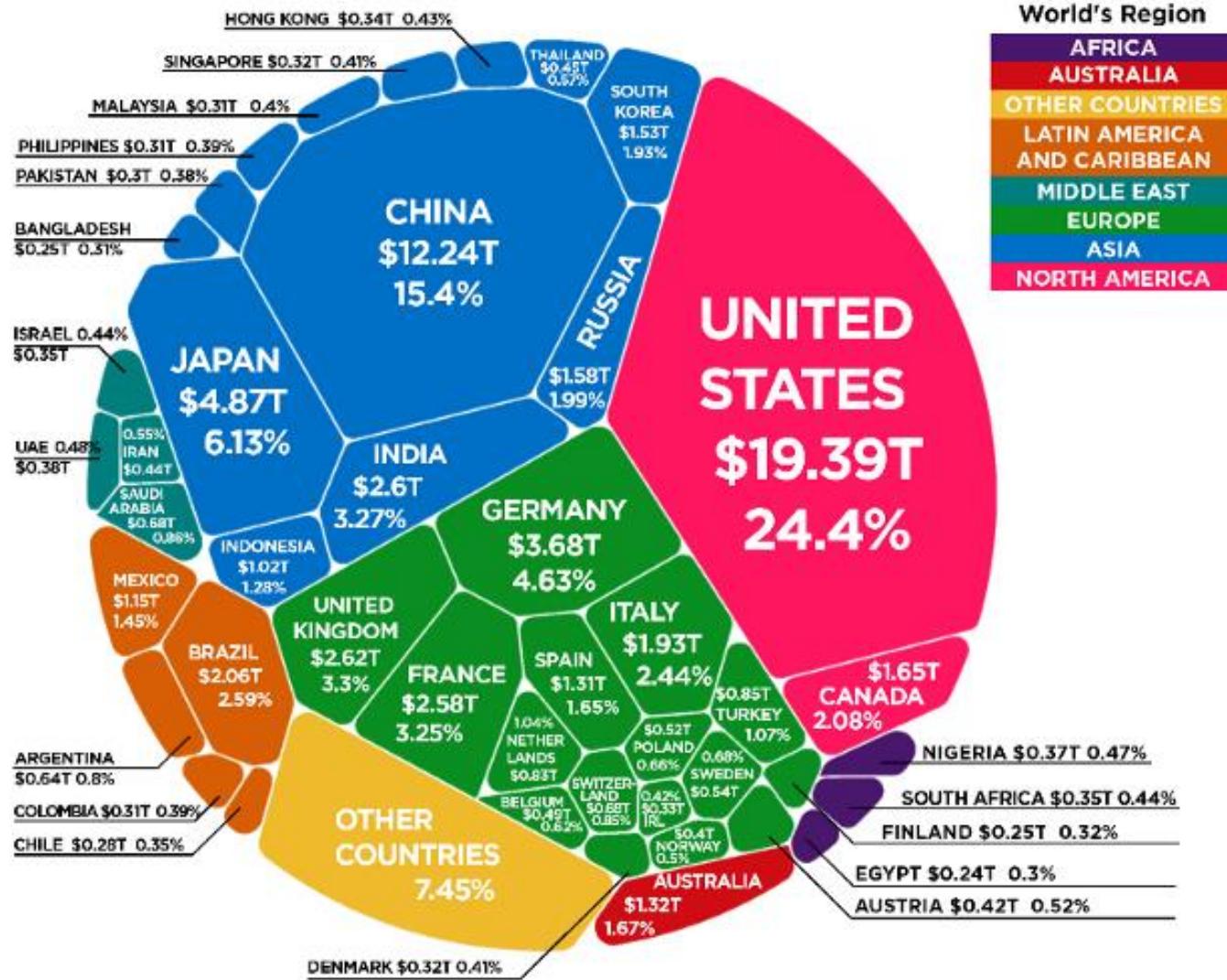


May 2007: 500 million RDF triples, 120 000 RDF links

- September 2011: 31.6 billion RDF triples, 503 million links
- April 2015: the number of sources is multiplied by 4 compared to 2011



Situation of LOD in 2019



Different types (stages) of linked data

Open data
OD

Free licence
Not linked

Linked data
LD

No Free licence
linked

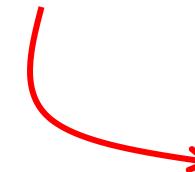
Linked Open data
LOD

Free licence
linked

Enterprise web of data (aka: knowledge graph)

Linked Closed data
LCD

No Free licence
Linked –business models



Knowledge graph construction

1. Identify the entities: named entity recognition
 - Objects or individuals
 - Disambiguate strings
 - This ensures that there is no data inconsistencies in the knowledge graph
2. Identify relationships
 - A link between two or more entities
 - This shows how the knowledge links together

Terminology Recap

Concept	Description
Semantic web	A set of technologies to make data retrievable and processable by machines (RDF, RDFS, OWL, IRI, SPARQL, SKOS, JSON-LD...)
Web of data	A web that implements semantic web
Open data	Data accessible online through free licence
Linked data	Data connected to other data
Linked open data	Open data linked with other linked data
Linked closed data	Linked data accessible under specific licences
Knowledge graph	Knowledge base based on graph-data-model (old concept - 1972)
(Semantic) knowledge graph	Knowledge graph that stored triples (subject, predicate, object)
Ontology	A semantic model of a domain in terms of concepts and relationships
Freebase	Knowledge graph base by metaweb, acquired by google (focused on public datasets)
BDpedia	An early knowledge graph base (focused on data from Wikipedia)

Exercise

- Consider the following statements
 - John is a lecturer
 - John teaches Big Data Management
 - Big Data Management is a course at Heriot-Watt University
 - Big data management is taught in Edinburg and Dubai
- Tasks
 - Identify entities (e.g., Dubai) and relationships (is a)
 - Which entities would be nodes (with IRIs) and which ones would be data values (no IRI)
 - What common sense reasoning can be applied to the statements? Give examples
 - Draw a graph with entities and relationships

Solution

Entities	Relationships
John	Is a
Lecturer	Teaches
Big Data Management	At
Course	Taught in
Heriot-Watt University	
Edinburgh	
Dubai	

