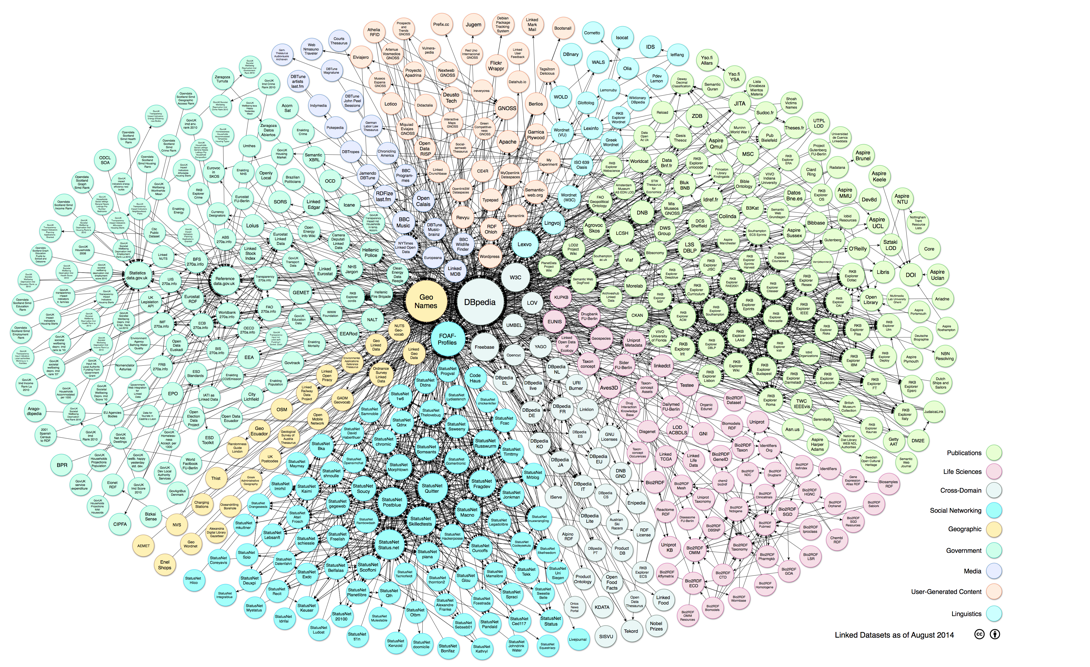
## Linked Data

Suzanne Little, [suzanne.little@dcu.ie](mailto:suzanne.little@dcu.ie)

Another way that data can be made open and accessible is to use Linked Data. Linked data arose from the efforts to create machine understandable linked information using the infrastructure and standards from the world wide web (See Figure 1 below).

**

*Figure 1: The Linked Open Data Cloud represents some of the publically available linked datasets (explore the interactive and most recent version* [*here*](https://lod-cloud.net/)*.*

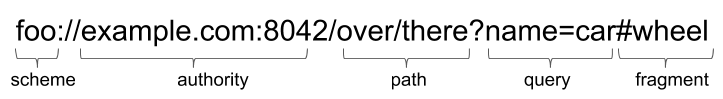
Watch this video - <https://youtu.be/uju4wT9uBIA> (3m42s) - for an explanation of linked data and its uses.

The key components of linked data that you should understand are:

**Uniform resource identifier** ([URI](https://www.w3.org/wiki/URI))

* Identifier string for everything (webpages, you, me, books, data, cars, …)
* See also [ORCID](https://orcid.org/), [DOI](https://www.doi.org/), [IBAN](https://www.isbn-international.org/content/what-isbn) for other unique reference schemes

See Figure 2 below for the components that make up a common URI.

**

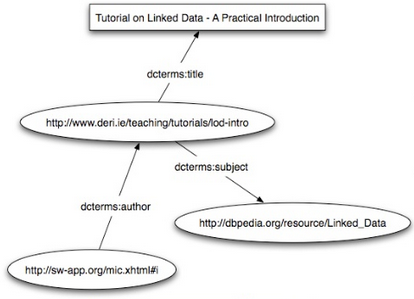
*Figure 2: Components that make up a common URI string*

The URI includes the scheme (eg: http/https), the authority (domain or host of the data), which may optionally have a port number, the path that often identifies the dataset or application, optionally a query that starts with ‘?’ mark and may have one or more attribute=value pairs. You sometimes also see content after the ‘#’ mark that identifies the part of the document identified by the URI.

**Resource description framework** ([RDF](https://www.w3.org/RDF/))

* Triples: Subject → Predicate → Object
* Can be recorded as text documents in XML format or [turtle](https://www.w3.org/TR/turtle/) (Terse RDF triple)
* This forms the *links* that make linked data

Figure 3 provides examples of linked data triples below.

**

*Figure 3: Example of linked data triples [*[*Source*](https://www.slideshare.net/mediasemanticweb/linked-data-michael-hausenblas-2009-03-05)*].*

**SPARQL Protocol and RDF Query Language** (SPARQL)

* Pronounced “sparkle”
* Like SQL but for RDF data
* Specifies triple patterns to match against in the WHERE clause
* PREFIX sets the namespace (semantic domain) that defines the meaning or topic of the subject/predicate/object.

Below is an example of a SPARQL query.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>;

PREFIX foaf: <http://xmlns.com/foaf/0.1/>;

PREFIX : <http://dbpedia.org/resource/>;

SELECT ?name ?birth ?description ?person WHERE {

?person a dbo:MusicalArtist .

?person dbo:birthPlace :Dublin .

?person dbo:birthDate ?birth .

?person foaf:name ?name .

?person rdfs:comment ?description .

FILTER (LANG(?description) = 'en') .

} ORDER BY ?name

*Figure 4: Example SPARQL query that will return the name, birth date, English description and link to the resource page of Musicians (“MusicalArtists”) born in Dublin, ordered by their name.*

### Exercise: using SPARQL on DBPedia

You can see an example of Linked Data and SPARQL in use at [DBPedia](https://wiki.dbpedia.org/about), a community-created database version of Wikipedia. Try the example query above using the online [SPARQL endpoint](http://dbpedia.org/snorql). Note that the list may not be exhaustive as it depends on what has been entered into DBPedia from Wikipedia but it’s often useful if you need to get a set of supplementary statements for a dataset.

Can you get a list of films starring your favourite actor? SPARQL is a bit counter-intuitive at times. You’ll want to include a ‘PREFIX dbo: <<http://dbpedia.org/ontology/>>’ and a triple statement ‘?movie rdf:type dbo:Film .’. The relationship (predicate) for the actor should be ‘dbo:starring’ and most actors will work if you use ‘:Firstname\_Lastname’ as the value (Object).

Try it for yourself first but if you’re stuck then you can see an [example query](https://www.computing.dcu.ie/~slittle/futurelearn/ca682i/linked_data_eg.html) for Tom Hanks.

Linked Data, RDF and SPARQL are not especially common in business applications. However, they are useful to know about for those occasions when you need to extract data from an open linked dataset and can be fun to play around with. These simple queries also don’t demonstrate the scope of linked data or SPARQL that can potentially link across multiple different, independently curated data stores. For this course, it is important to know the main concepts that make up Linked Data but I won’t be testing you on your ability to write SPARQL.

### Further reading on linked data:

Hausenblas, M., 2009. *Linked Data Tutorial*. [online] Slideshare.net. Available [here](http://www.slideshare.net/mediasemanticweb/linked-data-michael-hausenblas-2009-03-05). [Accessed 8 June 2020].

Poblet, M., Casanovas, P. and Rodríguez-Doncel, V., 2019. Introduction to Linked Data. In *Linked Democracy* (pp. 1-25). Springer, Cham. Available [here](https://link.springer.com/chapter/10.1007/978-3-030-13363-4_1)