# Linking up the web of data

The current Web is configured as a series of pages or ‘documents’. While these documents draw on rich sources of data, they disguise it beneath pages designed for humans to view. In this module, we explore what would happen if all the pages or documents were removed from the Web.

Imagine you only had the raw data, all open, all usable and all linked together in a network or 'web' of data.

This module also introduces the web of open linked data and looks at how the 5 Stars of linked open data provides a roadmap for achieving this vision.

**In this module we’ll explore the following:**

What is the web of data?

How web identifiers are used

What a web of open linked data looks like

# Surfing the web of data - in 2 minutes

ODI Trainer David Tarrant introduces the web of linked open data and examines how the 5 Stars of open linked data can help you start browsing the web of data.

# Understanding the web of data

The web of linked open data is changing the way we access and browse open data.

**What is the web of data?**

The current web is configured as a series of pages or ‘documents’. These documents are designed by humans to be visually attractive for other humans to read and to provide them with information.

There is a difference between data that is ‘on the web’ and the ‘web of data’.

Data ‘on the web’ is a file that can be downloaded from a webpage. The web of data is a set of deeply linked data resources that can be instantly read and understood by both humans and machines.

## Linking up the web of data

Imagine that all visually attractive documents were stripped away from the Web. How would you find your way around?

To navigate, you would need signposts to tell you how your current location relates to others around it.

The ‘web of data’ provides these signposts by requiring that all data be connected, via links, to other data.

The links allow you to navigate between data points and also allows machines to do so.

**What can a web of linked open data do?**

The web of data already exists in certain parts of the Web.

Large corporations like Microsoft have even researched tools that would allow us to navigate the Web in a whole new way by using data, rather than webpages, to move around.

[Take me to Microsoft Pivot talk](http://www.ted.com/talks/gary_flake_is_pivot_a_turning_point_for_web_exploration?language=en)

# Building a web of data

In order to build a web of open linked data, two things are required:

Data that can be referenced on the Web

People to provide Web links to this data from within their own data

These are the two conditions that meet the requirements of the 4th and 5th stars of linked open data.

# 4th star: open identifiers

The 4th star is all about being able to reference things on the Web.

## Open identifiers

An identifier is a short-form way of referencing a more complex thing.

Company numbers for businesses and Zip/Postal codes for places are common examples.

However, there are many other identifiers used both by the public and private sectors.

The challenge of many identifiers is that it is not obvious from the number itself what it refers to. Most people therefore rely on search engines to tell us more about what these identifiers are for. Web-based identifiers don’t have this problem as you can simply click the identifier to find out more.

For example, the company number for a business is 08030289. To find the company from this number, you would have to type it into a search engine to find out what the identifier was for before linking through to the resource. By contrast, you can follow a web-based identifier for the same book directly from any Web browser, for example: <https://beta.companieshouse.gov.uk/company/08030289>

## Using the 4th star

The 4th star requires the creation and use of Web-based (http://) identifiers for things, both physical and digital.

Identifiers are commonly sets of charecters, like post/zip codes (e.g. EC2A 4JE) or simply numbers (08030289).

Making these Web-based identifiers enables the use of the HTTP name resolution service to look up these things, and even retrieve an actual copy. eg

<http://id.southampton.ac.uk/building/59>

<http://data.ordnancesurvey.co.uk/id/postcodeunit/EC2A4JE>

## 4th star open data

The 4th star also requires that when a user clicks on a link and cannot be provided with the exact item (for example a book, object, person, or building) they should instead be provided with contextual information about that item.

This is often done by redirecting the client (human or machine) to that information. The information might be in the form of a webpage, or raw data.

**Try it out...**

Some Web-based identifiers tell a reader what the identifier is just from the text in the URL. Before you click the URL, why not try this for yourself? Below, try to guess what identifier it links to:

<http://data.ordnancesurvey.co.uk/id/postcodeunit/EC2A4JE>

The link sent a request to the server for a postcode. To fulfil the request, the server returned a webpage about the postcode as it couldn’t send you the postcode itself. Your browser has undergone a process known as content negotiation with the server, where you get a webpage because you're viewing through a browser. If you accessed this with a machine you would get it in another format, depending on what you said you preferred.

Why not see what a machine would see with content negotiation on the Ordnance survey link above by adding extensions like “.json” and “.xml” to the end of the URL?

# 5th star: linked data

The 5th star is about linking individual data points together.

## The link

The 5th star introduces the concept of linking data points together.

These links enable humans and machines to navigate between data sources on the Web.

Linking data points directly to one another allows you to direct the person or machine to a definitive record on that subject, rather than each person maintaining their own record of the data. Linked data functions like an official website by providing everyone with an authoritative resource on the subject. Linked data also saves everyone from duplicating data by sharing a single resource among an infinite number of datasets, using URIs.

## Using the 5th star

Example - A database of buildings owned by a company will often have Postcodes/Zip codes against each building in the data records:

**Building number**: 32

**Postcode**: SO17 1BJ

**Ward**: Postwood

**District**: City of Southampton

With linked data we have no need to replicate data – like **ward** and **district** – that are manged by other authorioties. Therefore, the record becomes:

**Building number**: 32

**Postcode**: <http://data.ordnancesurvey.co.uk/id/postcodeunit/SO171BJ>

## Try out using the 5th star

You may have noticed some links in the data that we explored earlier in this module.

Try loading the link from earlier in the module:

<http://data.ordnancesurvey.co.uk/id/postcodeunit/SO171BJ>Now try loading the same link into a data browser:

[Take me to Graphite linked data browser](http://graphite.ecs.soton.ac.uk/browser/)

The data browser uses automatic content negotiation for data, not webpage content, and displays the raw data in a human-readable form.

Any red text in the browser is a link in the data, which should go to more data.

Use the following URL to see what you can discover within the web of linked open data. <http://www.bbc.co.uk/programmes/b006q2x0>

**Are you ready to surf the web of data?**

There are four helpful things to remember when considering surfing the web of linked open data, can you recall them?

**What is the web of data?**

**The web of data is...**

a web of data portals where you can download data a set of interlinked data resources in the Web a new social network

**That’s right!**

The web of data is a set of interlinked data resources that can be instantly read and understood by both humans and machines.

**Are you sure?**

The web of data is a set of interlinked data resources that can be instantly read and understood by both humans and machines.

**What does the following identify?**

<http://id.southampton.ac.uk/room/32-3077>

a webpage

a building a room

**That’s right!**

Remember that the web-based identifier points to the thing itself and not to the webpage on which it is sitting.

**Are you sure?**

Remember that the web-based identifier points to the thing itself and not to the webpage on which it is sitting.

**Which of these are identifiers for the artist Taylor Swift?**

<http://www.bbc.co.uk/music/artists/20244d07-534f-4eff-b4d4-930878889970><https://musicbrainz.org/artist/20244d07-534f-4eff-b4d4-930878889970><http://dbpedia.org/resource/Taylor_Swift>

**That’s right!**

They all are! The first two use the same sub-identifier but are different authorities. The third one is the Wikipedia identifier (dbpedia) for the artist. They all link to each other and share each other's data, can you find the links?

**Are you sure?**

They all are! The first two use the same sub-identifier but are different authorities. The third one is the Wikipedia identifier (dbpedia) for the artist. They all link to each other and share each other's data, can you find the links?

# Surfing the web of linked open data

The web of data is a set of deeply linked data resources that can be instantly read and understood by both humans and machines.

In 1989, Sir Tim Berners-Lee invented the web of documents and it took us over 20 years to understand how to use it. It is now a fundamental part of our society.

In 2001, Tim envisioned the web of linked open data and we are still realising how this is going to fundamentally change the web all over again.

Main menu

# Module 13 – video

I’m Dave Tarrant, I’m the Trainer and Data Scientist here at the open data institute. In order to build the web of open linked data, we need two things. The first one is data that is referenceable on the web and the second thing is people to link to this data from within their own data. Thankfully, the five-star guideline gives us the fourth and fifth stars which can help us with this. The fourth star is all about identifiers. Now we use identifiers all the time, to identify places, people or maybe even a spare part for something. The problem is with these identifiers is that we can’t click on them, they’re just random numbers. So what if you turned these identifiers into full web addresses that you then could click on? We would be able to get back a human-readable web page about that thing. A machine would be able to get back data and a 3D printer might even be able to get back enough data to create the thing itself. So this identifier becomes the identifier for the physical thing not just something we can use to look up a webpage. And that’s what the fourth star is all about. The fifth star is all about adding context, and simply linking from this data or from this thing to other things that are contextually relevant so we can find out more information, and more knowledge. We live in an age where it’s more than just humans who are connected to the web. Everything from household appliances, to trains to even some clothing, is now connected. And this is of critical importance to help us derive this new value to bring new value to our society, our economy, and to help us save the environment. The web of open linked data is critical to helping us achieving this aim and unlock full potential. The web of open linked data will be the biggest evolution of the web, since the web itself.