

Altair Acquires Cambridge Semantics, Powering Next-Generation Enterprise Data Fabrics and Generative AI. [Read More \(https://altair.com/newsroom/news-releases/altair-acquires-cambridge-semantics-powering-next-generation-enterprise-data-fabrics-and-generative-ai/\)](https://altair.com/newsroom/news-releases/altair-acquires-cambridge-semantics-powering-next-generation-enterprise-data-fabrics-and-generative-ai/).

Introduction to the Semantic Web

≡ MENU

TOWARDS THE SEMANTIC WEB ([HTTPS://CAMBRIDGESEMANTICS.COM/BLOG/SEMANTIC-UNIVERSITY/INTRO-SEMANTIC-WEB/TOWARDS-THE-SEMANTIC-WEB/](https://cambridgesemantics.com/blog/semantic-university/intro-semantic-web/towards-the-semantic-web/))

INTRODUCTION TO LINKED DATA ([HTTPS://CAMBRIDGESEMANTICS.COM/BLOG/SEMANTIC-UNIVERSITY/INTRO-SEMANTIC-WEB/INTRO-LINKED-DATA/](https://cambridgesemantics.com/blog/semantic-university/intro-semantic-web/intro-linked-data/))

WHAT IS LINKED DATA ([HTTPS://CAMBRIDGESEMANTICS.COM/BLOG/SEMANTIC-UNIVERSITY/INTRO-SEMANTIC-WEB/WHAT-IS-LINKED-DATA/](https://cambridgesemantics.com/blog/semantic-university/intro-semantic-web/what-is-linked-data/))

SEMANTIC WEB LANDSCAPE ([HTTPS://CAMBRIDGESEMANTICS.COM/BLOG/SEMANTIC-UNIVERSITY/INTRO-SEMANTIC-WEB/SEMANTIC-WEB-LANDSCAPE/](https://cambridgesemantics.com/blog/semantic-university/intro-semantic-web/semantic-web-landscape/))

THE MANY NAMES OF THE SEMANTIC WEB ([HTTPS://CAMBRIDGESEMANTICS.COM/BLOG/SEMANTIC-UNIVERSITY/INTRO-SEMANTIC-WEB/MANY-NAMES-SEMANTIC-WEB/](https://cambridgesemantics.com/blog/semantic-university/intro-semantic-web/many-names-semantic-web/))

SEMANTIC WEB MISCONCEPTIONS ([HTTPS://CAMBRIDGESEMANTICS.COM/BLOG/SEMANTIC-UNIVERSITY/INTRO-SEMANTIC-WEB/SEMANTIC-WEB-MISCONCEPTIONS/](https://cambridgesemantics.com/blog/semantic-university/intro-semantic-web/semantic-web-misconceptions/))

Introduction

“The Semantic Web is a webby way to link data” is one of the most brilliant definitions of the Semantic Web. Dave Beckett genially captured the very final essence, in one single sentence, of what we call Linked Data. If we look closer, we can find in it the main idea behind the Linked Data paradigm: using the Web model to publish and connect raw data.

In today’s lesson you’ll learn about one of the most important movements in the Semantic Web community—Linked Data—which strives to expose and connect all of the world’s data in a readily queryable and consumable form.

For a lighter, less technical introduction to Linked Data, see our brief video introduction: What is Linked Data?

Today's Lesson

The model behind the Web could be roughly summarized as a way to publish documents represented in a standard way (HTML), containing links to other documents and accessible through the Internet using standard protocols (TCP/IP and HTTP). The result could be seen as a worldwide, distributed file system of interconnected documents that humans can read, exchange and discuss.

You can imagine it as a set of books with references to other books or articles: browsing the Web is not really different from going through a book and following its citations to other publications, as described in Introduction to the Semantic Web.

Let's focus on a lower degree of granularity and look at data instead of just documents. As most data exists in databases of one form or another, we'll start there.

Figure 1 shows two relational tables of two different databases. The first contains data about movies, with titles, actor and director names. The second is a collection of theaters in a town where such movies are playing. It's immediately clear that there is an implicit link between the rows of those two tables.

Now let's say that you wanted to publish this information so that it could be easily consumed by other Web sites or computer programs. For example, so that a program can automatically query your Web site and any other site that has movie scheduling information in order to show a complete view in one place.

This is where the traditional Web here pays its most serious technological debt. There is no standard way to make such tables available on the Web and have them be accessible from an application. You could export the values in CSV and putting them on a HTTP Server (everything reads CSV, right?), but this has its own problems.

Even though CSV files are raw data that are not human readable and are meant for machines, you haven't succeeded. The problem is that my website's CSV format might differ from yours. Even a simple difference such as changing the order of the columns can make it impossible for an automated computer program to consume the data, since they would have to have separate rules for every kind of CSV anywhere that talks about movies or theaters! Let alone the infinite other kinds of data you might want to publish and consume.

This is where Linked Data is revolutionary.

The goal of Linked Data is to publish structured data in such a way that it can be easily consumed and combined with other Linked Data.

A future lesson will show you how to actually publish your own Linked Data.

The Four Rules of Linked Data

So in a way, Linked Data is the Semantic Web realized via four best practice principles.

1. Use URIs as names for things. An example of a URI is any URL. For example: <https://cambridgesemantics.com/blog/semantic-university/> is the URI for Semantic University.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards such as RDF* and SPARQL.
4. Include links to other URIs so that they can discover more things.

The Four Rules Applied

Let's now imagine that you would like to publish the two tables mentioned above applying the Linked Data principles.

1. Instead of using application-specific identifiers—database keys, UUIDs, incremental numbers, etc.—you map them to a set of URIs. Each identifier must map to one single URI. For example, each row of those two tables is now uniquely identifiable using its URI.
2. Make your URIs dereferenceable. This means, roughly, to make them accessible via HTTP as we do for every human-readable Web page. This is a key aspect of Linked Data: every single row of our tables is now fetch able and uniquely identifiable anywhere on the Web.
3. Have our web server reply with some structured data when invoked. This is the Semantic Web “juicy” part. Model your data with RDF. Here is where you need to perform a paradigm shift from a relational data model to a graph one.
4. Once all the rows of our tables have been uniquely identified, made dereferenceable through HTTP, and described with RDF, the last step is providing links between different rows across different tables. The main aim here is to make explicit those links that were implicit before shifting to the Linkeddata approach. In our example, movies would be linked to the theaters in which they are playing.

Once our tables have been so published, the Linked Data rules do their magic: people across the Web can start referencing and consuming the data in our rows easily. If we go further and link from our movies to external popular data sets such Wikipedia and IMDB then we make it even easier for people and computers to consume our data and combine it with other data.

In some sense, the four Linked Data rules, which strongly rely on the Semantic Web stack, could be seen as a new layer in the OSI model, on top of the networking layer, but below the application layers.

The Big Picture: Building a Web of Data

Linked Data is not only about exposing data using Web technologies. Nor is it simply an elegant way to solve interoperability issues. Linked data is fundamentally about building a Web of Data.

Imagine hundreds of different data sets published on the Web according to the Linked Data principles: thousands and thousands different identifiers you can rely to grab data about books, movies, actors, cities, or anything your can imagine. In few words, such datasets form a giant Web-scale database you could potentially embed in your applications and reference whenever you needed.

The four principles really shine when links are provided between different data sets. To return to our book reference analogy, instead of having links citations simply between books or Web pages, this allows links between anything to be followed for more information. If a single author has published in two different journals, for example, and both journals expose their catalogs as Linked Data, and the author's bio is on DBpedia, then your application can easily mash it all together with a simple query, automatically. As if all the data was in one database.

Where Open Data meets the Semantic Web: Linked Open Data

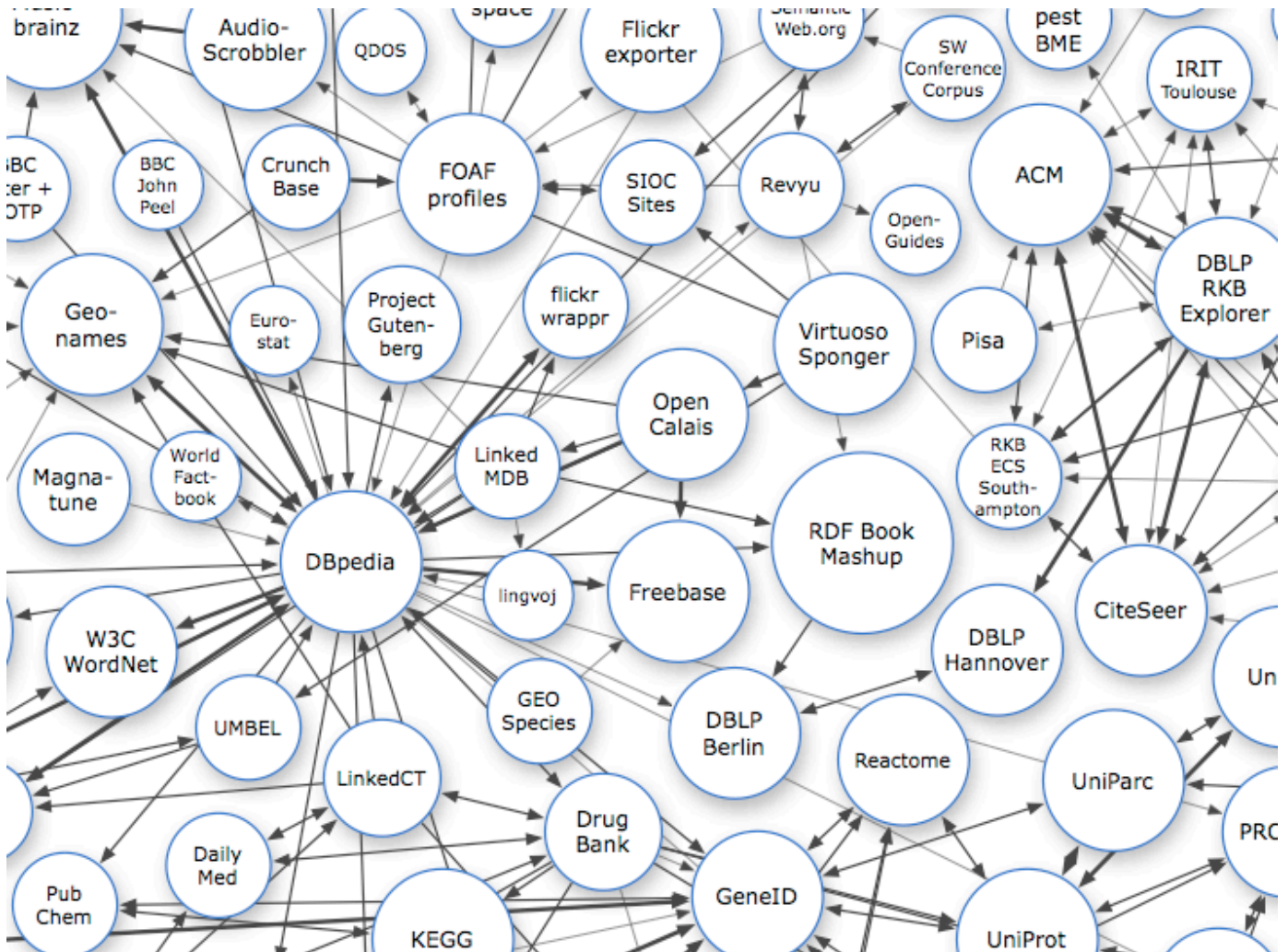
Over the last few years, the Linked Data paradigm has found a huge application when combined with the publication of data with liberal licenses. The Open Data Movement, which aims to release huge data sets often from local government authorities, embraced the Linked Data technologies and best practices to publish a plethora of different interlinked data sets.

Roughly speaking, a bunch of data published with an open license is intended to be freely available to everyone to use and republished without restriction from copyright, patents or other restrictions.

Where Open Data meets Linked Data, we have Linked Open Data.

The Linked Open Data movement has experienced exponential growth in term of published data sets. Within four years, the number of published data sets has grown from 12 to 295.

The most popular and emblematic diagram about Linked Open Data is the “cloud diagram” made by Richard Cyganiak and Anja Jentzsch. They are maintaining and releasing a diagram showing all the published data sets, with their sizes and inter-links. The diagram is automatically sketched using the CKAN API to get JSON for each of the data sets and then manually clustered and colored by their characteristics.



Dataset Quality in the Linked Data Web

As the open data movement gains momentum worldwide, the community faces the need for a set of metrics to estimate the quality of different datasets as reliable dataset quality is a crucial aspect for a long-term success and sustainable data re-usage.

A pragmatic and informal set of metrics have been introduced by Tim Berners-Lee, inventor of the Web: the 5 Stars of Linked Open Data. Basically, it's a set of incremental characteristics that published data sets should have. The more of these features a data set has, the higher the quality of its data.

One star – make your stuff available on the Web (whatever format)

Two stars – make it available as structured data (e.g. Excel instead of image scan of a table)

Three stars – non-proprietary format (e.g. CSV instead of Excel)

Four stars – use URLs to identify things, so that people can point at your stuff

Five stars – link your data to other people's data to provide context

Data published in accordance to the Linked Data principles is clearly 5 Star quality. 5 Star datasets are intrinsically more valuable because the data within has been made accessible through the Web (so you can get at it), has been exposed with open standards (so its easily consumable), and has been linked to

other data (so it's verifiable and more easily integrated). The combination makes application development absolutely more cost-efficient and less time-consuming compared to using less quality datasets.

Conclusion: From the Web of Data to the Web of Things

Looking at the growing rate of Linked Data we can bet on its having a bright future. Linked Data is the new de-facto standard for data publication and interoperability on the Web and is moving into enterprises as well. Big players such as Google, Facebook and Microsoft, already adopted some of the principles behind it. The Open Graph Protocol and Schema.org are somehow proving that linking data on the Web is the new strategy to grab structured data into their own platforms.

Relying on semantic markup technologies, such as Microformats, Microdata, and RDFa—which all allow structured data to be embedded in traditional HTML pages—millions of pages containing structured data are popping out on the Web. And all this data can be linked and consumed according the Linked Data principles.

The far future is even more promising. Think for a moment about the growth of mobile devices and the rising pervasiveness of connected devices. The next Web will be about connecting every device or tangible thing in our world.

In this scenario, often referred as the “The Internet of Things”, the Linked Data paradigm is well positioned to play a strategic role as a pillar of the whole new Internet infrastructure. URIs could be used to uniquely identify real World objects, RDF could be used as a standard way to syndicate data about them and, most importantly, the interlinking could be used to represent objects interactions and relationships with other surrounding objects.

< PREVIOUS LESSON

([HTTPS://CAMBRIDGESEMANTICS.COM/BLOG/SEMANTIC-UNIVERSITY/INTRO-SEMANTIC-WEB/TOWARDS-THE-SEMANTIC-WEB/](https://cambridgesemantics.com/blog/semantic-university/intro-semantic-web/towards-the-semantic-web/))

NEXT LESSON >

([HTTPS://CAMBRIDGESEMANTICS.COM/BLOG/SEMANTIC-UNIVERSITY/INTRO-SEMANTIC-WEB/WHAT-IS-LINKED-DATA/](https://cambridgesemantics.com/blog/semantic-university/intro-semantic-web/what-is-linked-data/))



1 Beacon St | 15th Floor | Boston, MA 02108

Copyright © 2024 Cambridge Semantics. All Rights Reserved. |

Privacy Policy (<https://www.cambridgesemantics.com/privacy-policy/>)