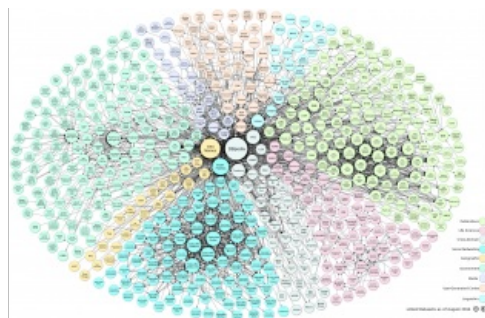


TECHNOLOGICAL BACKGROUND

Smart Cities and Linked Data



The European Union defines a 'smart city' as "a city well performing in 6 key fields of urban development, built on the 'smart' combination of endowments and activities of self-decisive, independent and aware citizens". Amongst the key fields are smart mobility and a smart environment. Smart cities rely heavily on reliable, accurate and available data. Linked open data (LOD) is a way to make that data available and

therefore to 'enable' the smart city.

(By Huibert-Jan Lekkerkerk, contributing editor, GIM International, The Netherlands)

Linked data (LD) refers to data that is made available in a structured way through the internet. Linked open data (LOD) takes the linked data concept one step further and exposes the linked data to everyone on the internet. LOD is based on the 5-Star Open Data scheme, which was developed by Tim Berners Lee, one of the founders of the internet in the early 1980s.

Creating LOD

The first step (equivalent to one star) is to make information available on the web (in any format) under an open licence so that it can be reused. The second step or star is to make it available in a structured, computer-readable format such as the ESRI shape format or an MS-Excel file. Even better (three stars) would be to do this in an open, non-proprietary format such as GML. Most datasets that are available today adhere to one, two or three stars, including the large variety of web services making data available through, for example, WFS.

But it is only with four and five stars that one enters the realm of LOD. So what makes this different? In LOD all 'things' are uniquely identified. The final step is to link all these objects together in such a way that the data can be navigated and that additional context is provided. This can be done within a single dataset by creating a link between a house and a parcel, but can also be done between datasets. By naming the links between the objects, context can be provided about the relationship as well as effectively creating what is called the 'semantic web' or 'Web 3.0'. For example, when the municipality keeps track of new houses and the cadastre of the parcels on which they are built, there could be a link from the house to the cadastral parcel. Anybody interested in the house could not only see the information about the house but could also navigate towards the parcel data. And a wider context is also possible. If, for example, the house has been featured in a movie, a link to the house (and as a result also to the cadastral parcel) could be established from the movie database.

LOD Technology

LOD requires the implementation of a number of standards that are not often used outside the LOD community, with the exception of object identification. Object identification is done by so-called universal resource identifiers (URIs). An example of a URI is the way universal resource locators (URLs, e.g. <http://gim-international.com>) are used to find web pages. In LOD, however, the URI identifies a specific geographical object (road, building, cadastral parcel) rather than a specific web page. The URI needs to be stable over time; changes in the data should not change the URI (otherwise, the network of links would crumble). To keep the URI stable, governments may adopt a URI strategy for their linked data which describes how URIs should be created and navigated.

Two further World Wide Web Consortium (W3C) standards are required to make LOD work, namely the Resource Description Framework (RDF) and the query language SPARQL. RDF defines the XML structure to use when describing objects and linking them. For the links, 'triples' are used that describe how object types (and therefore objects) are related by identifying a 'from' object and a 'to' object as well as the relationship between them. In the example above, the triple between the house and the cadastral parcel could be that the object 'House A' is related to the object 'Parcel B' as 'Located on': 'House A' - 'Located on' - 'Parcel B'. Between the movie and the house, the triple could become 'Movie C' - 'Features' - 'House A'. This would not only allow navigation between the objects but would also provide the context of the relationship; instead of people having to guess how the two objects are related, RDF identifies what type of relationship the two objects have. A special type of link in RDF is that of defining 'sub classes', which allow a hierarchy of objects to be created. For example, 'House' is a 'sub class' of 'Building', which also defines House A as a building.

SPARQL was designed to navigate through RDF datasets and to query them for specific information. In that respect SPARQL is to RDF what SQL is to a relational database. Using SPARQL, it would be possible to perform the query 'On what Parcel is the building featured in Movie C located?'. SPARQL can translate the query, navigate the triples and arrive at Parcel B. By default, SPARQL does not support geospatial queries. To remedy this, the Open Geospatial Consortium (OGC) has extend SPARQL and RDF with a geospatial vocabulary; this extension is called GeoSPARQL. These extensions allow not only linking between identifiers, but also storing (and querying) coordinate information, thus giving even more linking options.

Creating Smart Cities

Over the last few years, LOD has slowly become an accepted way of exposing data to the internet. One could also say that LOD, together with the Internet of Things (IoT) is one of the key requirements for smart cities. If governments would open up their datasets, and in particular their sensor networks, over the internet using LOD, then this could 'enable' smart cities. The 'things', such as sensors, expose their data in a structured way and (potentially) linked to other datasets. This in turn may lead to applications that are not yet foreseen as data is not yet available in this manner. For example, a traffic intensity sensor not only exposes the traffic intensity itself to the internet but also the information about the road on which it is located. Using that road location, the information can be combined with other data such as road maintenance information, an air pollution sensor in the vicinity and/or meteorological information.

Threats and Weaknesses

Although the potential benefits could be huge, there are still some hurdles to be overcome. The first is that (even) more datasets need to be opened up based on at least a three-star approach. What may also be a potential threat is that RDF and GeoSPARQL are quite remote for today's app builders who rely on relatively simple and semantically poor API and JSON interfaces. Making the switch to full five-star LOD may be beneficial to data in general but detrimental to the actual use of that data by a wide audience. Whilst RDF and SPARQL were published over five years ago, mainstream IT has been slow to support it so far. Furthermore, the publication of LOD by organisations is not yet a common part of data management processes (although, considering that nor is publishing open data, perhaps this should not come as a surprise). A government-inspired community is experimenting with LOD in the Netherlands but, even with such a large force behind it, developments are relatively slow.

Conclusion

Linked open data (LOD) could be a key enabler for the smart city and has the potential to generate applications that are as yet unknown. However, the technical and organisational hurdles that need to be overcome should not be underestimated.

Huibert-Jan Lekkerkerk

Huibert-Jan Lekkerkerk is a contributing editor and author of various publications on GNSS and hydrography as well as a principal lecturer on hydrography at Skilltrade. He is also a technical manager for the Dutch government where he works on nationwide standardisation and information management issues.

info@hydrografie.info

Further reading

- smart-cities.eu
- w3.org/TR/rdf-primer
- w3.org/TR/sparql11-query
- opengeospatial.org/standards/geosparql