Science & Philosophy, 3(2)/2015, 37--46 ISSN online: 2282-7765
print: 2282-7757

Semantic Web: Metadata, Linked Data, Open Data

Vanessa Russo[1]

[1] Department of Humanities, Arts, and Social Sciences
University "G. d'Annunzio"
of Chieti-Pescara, Via dei Vestini 31, 66100 Chieti, Italy
russov1983@gmail.com

Abstract

What's the Semantic Web? What's the use? The inventor of the Web Tim Berners-Lee describes it as a research methodology able to take advantage of the network to its maximum capacity. This metadata system represents the innovative element through web 2.0 to web 3.0. In this context will try to understand what are the theoretical and informatic requirements of the Semantic Web. Finally will explain Linked Data applications to develop new tools for active citizenship.

Keywords: Web Semantic; DBpedia; Open Data; Linked Data; Web 3.0

1 Background: What is Semantic web?

In 2002 Tim Berners-Lee, one of the inventors of the World Wide Web in his book The architecture of the Web describing an ideal future for the network:

«have a dream for the Web . . . and it has two parts. In the first part, the Web becomes a much more powerful means for collaboration between people. I have always imagined the information space as something to which everyone has immediate and intuitive access, and not just to browse, but to create. The initial World Wide Web program opened with an almost blank page, ready for the jottings of the user. Robert Cailliau and I had a great time with it, not because we were looking at a lot of stuff, but because we were writing and sharing our ideas. Furthermore, the dream of people-to-people communication through shared knowledge must be possible for groups of all sizes, interacting electronically with

as much ease as they do now in person. In the second part of the dream, collaborations extend to computers. Machines become capable of analyzing all the data on the Web—the content, links, and transactions between people and computers. A "Semantic Web," which should make this possible, weaving the web has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy, and our daily lives will be handled by machines talking to machines, leaving humans to provide the inspiration and intuition. The intelligent "agents" people have touted for ages will finally materialize. This machine-understandable Web will come about through the implementation of a series of technical advances and social agreements that are now beginning. Once the two-part dream is reached, the Web will be a place where the whim of a human being and the reasoning of a machine coexist in an ideal, powerful mixture. Realizing the dream will require a lot of nitty-gritty work. The Web is far from "done." It's in only a jumbled state of construction, and no matter how grand the dream, it has to be engineered piece by piece, with many of the pieces far from

glamorous.» (Berners Lee, 2001)

The utopian idea of the proposed network by Berners-Lee in recent years is taking shape in the revolution of Big Data and the Semantic Web.

The Semantic Web is a computer system that allows detecting the content of network data, analyzing and finalizing them in a specific use. It consists of a metadata system that is based on a protocol called Resource Description Framework through which you can make a unique data collection system using a URI (Universal Resource Identifier).

Through this new detection technology becomes possible to organize the data according to a "logic of predicates" and finding for each of the ad hoc resource: in this way a document may be analyzed and elaborated according to various needs.

The Semantic Web technology has given rise to the idea of the Internet as a huge database full of information traveling in real time and is called Big Data. (Hitzler, Janowicz, 2013)

The Big Data consist in the set of all the movements that users perform within the network and with certain characteristics attributable to five variables:

1) Speed: Data are traveling in real time;

2) Varieties can be text files, multimedia files, geocalizzazioni, connections between users, clicks, cell streams, sensors, commercial transactions of various kinds;

3) Volume: have a magnitude in terms of tens of zettabytes per day;

4) Variability: the interpretation of the same data varies depending on the context in which It's detected;

5) Virality: the spread of the data is likely to change in real-time events and public opinion in such contributions.

The accumulation sources of Big Data can be classified into three main categories: information to Machine People, People to People and Machine to Machine.

The information Machine People to represent the source of origin of most of the accumulated data and the object of interest.

It derive from the data "click" or the number and type of sites visited daily and purchases made in the online shops. These P2M Data have become essential tools for use by marketing agencies, to analyze the performance of a Web site, and therefore its placement within the search engines.

Furthermore, the analysis of the variety of sites visited by each user allows to design "ad hoc" advertising bid for each user. The data classified as a People to People concern the information content derived from human interactions within the social media, the blogs and the discussion forum. In fact, are under analysis all the data derived from post content on social networks, discussions generated on a given topic, and the types of selected Hashtags to categorize a certain theme.

The information Machine to Machine, finally, gather the data generated by the interaction between our electronic devices with other electronic devices. Specifically big data M2M include: The localisation; Economic transactions with credit cards; Data from the I.o.T.

Data Machine to Machine allow you to detect our attitude to the purchase of a given asset. Finally, through the localisation, it's possible to detect the popularity of a particular trade or tourist spot.

The big data are the boom of the new economy of the digital and have been defined the oil of the new millennium. However, given the speed with which the web has evolved and continues to evolve it would be short sighted not to expect from the future a further digital revolution.

The era of the post-information, sets out a new vision for communication. The development of digital media allows the design of an ad hoc information system for the user. Negroponte in this context carries out a prospective analysis of what later will be called the third generation of the Web.

«When you know my address, my marital status, my age, my income, the brand of my car, my purchases, what I like to drink, and how much I pay taxes you know me, an entity formed by demographic one person. [...] This information related only to me personally, however, identify new services that I may want, for example information on a small undisclosed location, a person not so famous, the today's weather forecast in Virginia [...] It was the post- information means knowledge that extends over time: machines that understand people with the same level (if not more) of insight that we can expect from humans, machines that take into account the small individual quirks (such as those for shirts with blue stripes) and made completely random, good or bad, that are part of our everyday life.» (Negroponte, 2004)

Negroponte wrote these considerations when the Web, in its current form, it was only fiction. Therefore it can be considered a precursor of the era 3.0. In addition, according to the author, "is currently the media business models are based almost exclusively on the concept of" push "information to the public and spectators. Tomorrow will be based rather more on the idea of "pull," that is, we will be you and I to get into the net to look for something, as we do today in a library or a shop where are rented videotapes.

2 RDF for representation of knowledge

With the Semantic Web term means the transformation of the web in an environment where you can publish not only documents but also information about themselves (metadata documents) in a format suitable for the question, the interpretation and, more generally, automatic processing.

Through the semantic web you can develop a software environment that allows each user to make the best possible use of the resources available on the Web.

In fact, through the extraction and metadata reading we can customize your search and identify new forms of connections through the establishment of semantic graphs.

In browsing the web, we follow the link, leading to what is formally known resource. In the current language an asset is also called "paper", to make it readable. So "the resource" is not a separate entity, but It’s accompanied by information describing it.

The information on the resource are generally said Metadata. Si can therefore say that the metadata is information, understandable by the machine, relating to a web resource or some other thing and, therefore, constitute a kind of information that can be used, to make a appropriate use of resources, making it faster and easier the operation of the Web.

For example, when it retrieves a document (or object) on the web, using the HTTP protocol, It’s possible that the server may send some information on the resource, such as his date of renovation, the maximum validity date information, its author, etc. So the Web, as a set of resources and resource information (metadata) is already a reality to which we are accustomed. (Signore, 2005).

At the methodological level for an effective use of metadata, it requires a encode for semantics, syntax and structure. The acronyms RDF (Resource Description Framework) is the basic tool to encode, exchange and re-use of structured metadata, and allows' interoperability between applications exchanging information on the Web. The sectors in which RDF can be used and bring advantages are various:

• description of a Web site, or a page, or a library digital;

• software implementation for the exchange of knowledge and better use of Web resources;

• classification of the content, to apply selection criteria;

• description of a set of pages, which represent a single document logical;

• establish the intellectual property policies for the individual pages;

• express a preference privacy of user identity and privacy policies of a website;

• with the digital signature mechanism, contributing to the creation of the Web of

Trust, for applications in e-commerce, cooperation, etc..

The RDF Data Model is based on three fundamental concepts:

• Resource: indicates whatever is being described by an RDF expression.

• Properties: it describes a specific aspect, an attribute or relation used to describe a resource;

• Statement: It describes the characteristics of a resource and relationships with other resources. Each assertion is formed, then, by: subject, predicate and object. The subject of a triple is the URI that identifies the described resource. The object can be a simple literal value (such as a string, number, or date), or the URI of another resource that is in some way related to the subject. The predicate, in the middle, shows the type of relationship exists between the subject and the object. (Signore 2002)

3 Linked Data

The protocols listed are regulated by the World Wide Web Consortium (W3C), an NGO whose mission is "Leading the Web to Its Full Potential...." The W3C has defined a series of projects and protocols that exploit the semantic web. Among them we can find the "linked data" system. By this term It's meant a mode of publication of structured data in order to interconnect them, and then make them usable through semantic query.

It's based on web technologies and open standards such as HTTP and URI, and extending its application to provide information which can be read and understood by computers. This makes it possible to connect and use data from different sources.

The term Linked Data, created by Tim Berners-Lee, It's often used in the context of the Semantic Web when you want to highlight the possibility of building a data network to connect information belonging to many different domains and allows the creation of new applications can access it, freely creating added value. The assumption at the basis of this idea is based on the principle that

the value and usefulness of the data grows as much as these are interconnected. (Salarelli 2014)

In summary therefore the Linked Data is the use of the Web to create specific links between data from different sources. The tenets of this approach include using the RDF as a model for the publication of structured data on the Web and the links that RDF provides to interconnect data. (Bizer, Heath, Berners-Lee, 2009).

4 Application: Open Data and Civic Hacking

The sources for the creation of semantic knowledge networks depend on the possibility of access to and from the queries to the specific database. For this reason, there is the problem of accessibility to the sources. In this sense the social utility of the linked data lies in the idea of considering the data as common goods that can provide benefit.

In this context born the project Linked Open Data. The aim of the project is to extend the W3C Web publishing several open RDF dataset as on the Web and by setting RDF links between data from different resources. In October 2007, the dataset contained more than two billion RDF triples, connected by more than two million RDF links. From May 2009 rose to 4.2 billion RDF triples, connected by about 142 million RDF links.

An example of Linked Open Data is DBpedia, a database that allows to convert all the information in Wikipedia linked data. DBpedia is one of the most important dataset of Linked Open Data, collects as RDF extracted information in articles from Wikipedia, and Geonames. (Auer, Bizer, Kobilarov, Lehmann Cyganiak e Ives, 2007)

The project takes advantage of this huge information resource by extracting from it semantically structured information. The transformation process basically consists of three elements:

- Extracting data from Wikipedia: it has been developed a system for the extraction of the Wikipedia data that should convert its content in a base of multi-domain knowledge.
- Creation of the knowledge base: the information obtained during the extraction have been developed, providing a semantic ontology and defining where the information are mapped.
- Creation of access points to the knowledge base: various means have been made available in order to access the DBpedia data such as RDF links pointing from DBpedia to other datasets.

The system Linked Open Data is the foundation of philosophy Open Data and forms of civic hackerism. (Solodovnik, 2015)

The civic hacker is one who finds the mechanism to easily make available to citizens to public data, offering a potential contribution to improving the quality of life of people.

The civic hacking is based on the model of the Innovation without permission, expression with which you want to define the innovation that does not knock at the door, who does not ask if you can, but that just happens. (Townsend, 2013)

In Italy, the civic monitoring system as a powerful example Spaghettiopendata. The civic hacker group was founded in September 2010 by Alberto Cottica Post: «Launching a proposal: If you are aware of public and open data, in all sizes (from the national to the neighbourhood: there must be some municipality or some "geek "out there!), And on anything, let me know. Anyhow: blog comments, social networks, phone, smoke signals, whatever you want. I commit myself to share them in some way (many have gathered here in quite spontaneously). Besides me, you can also report it to Federico Bo, Matteo Brunati, Laura Tagle; at the end we will put together all that we have collected, so as to facilitate access to data for citizens who want to play. If someone else wants to join in the hunt, welcome (Gaspar Torriero and Massimo Mantellini, for example, have been shown to be on the piece): we are trying to give life to a common resource, the more the merrier.

I ask you a favour: if you feel the topic important enough to share it on the network, use the hashtag #opendataitaly».

The activist group meets online inside a mailing lists Google and offline through the annual gatherings.

The Spaghettiopendata community in about five years of activity has implemented several civic Hacktivism projects. Among them It’s to underline the Monithon \ ConfiscatiBene.

Monithon is a portal of civic monitoring of EU-funded projects aimed at structuring of social cohesion and territorial development.

Through Monithon you can create "civic monitoring report."

For example, in February 2014 a group of Palermo hackers have carried out a report on the implementation of the Ring Rail Palermo comprising: pre-construction site monitoring, project description, report on the state of knowledge and on citizens' opinions, efficiency evaluation communicative the individuals responsible for the project, technical evaluation on work and concluding remarks.

ConfiscatiBene, however, is a project aimed at monitoring of property confiscated by the state from criminal organizations of the Mafia. The aim of bringing citizens and report on:

- How many goods are;
- where am I;
- How valuable;
- As they are re-used by the state;

The Monithon and ConfiscatiBene projects share the same assessment form and are both based on open sources available from OpenCoesione portal.

Openpolis instead it's an independent association aimed at promoting transparency and public administrations and public budgets.

OP, presently, is active for seven projects throughout the national territory:

- Open Parliament: within this portal you can monitor activity talking in relation to specific keywords;
- Productivity Parliamentary Index: data from Open Parliament flow into the IPP site in order to assess the activities of individual MPs in relation to the amount and effectiveness of the activity undertaken;
- Open Financial Reporting: the portal contains all of the common budgets of the last ten years of operation all in downloadable format and easily readable; Open Politicians: The platform allows you to monitor the activity of Italian politicians, local and national. In addition It's possible to know which Parliamentarians have offered their declaration of assets;
- Open Town Hall: The project consists in the creation of local authorities who, through the instrument of Open Data, are able to offer the people the opportunity to participate, comment and vote on the acts, follow political, topics and territory.
- You are here: within the site the citizen can make a test to see which parties responding to their needs by analyzing the electoral programs and public speeches;
- Open Blog: The platform is a container of initiatives and proposals of Civic Hacking.

Finally, among the virtuous examples of HC, it should be emphasized the OpenRicostruzione project, promoted by Wikitalia Openpolis and associations for the Emilia Romagna region.

The portal is designed to monitor the status of post-earthquake reconstruction in the Emilia Romagna region and the effectiveness of the donations received.

5 Conclusion

The study of web 3.0 requires: updated tools for working, a fast and easy navigation and a conscious use of the data. The Semantic Web overcome the idea of just browsing the web, in fact allows you to customize every detail to our possible need for knowledge, all this through a system of procedures and specific languages that can query databases. Therefore, the semantic web function is to access to structured information for to pull out the information that a query has requested.

Currently, the major limitation of the Linked Open Data is its difficulty of use, because for access at large amounts of data and metadata available need knowledge of specific languages and computer techniques. For this reason the future of open data management is in the development of intelligent semantic agents, a set of programs that can autonomously explore and interact with the computer systems.

Tim Berners-Lee in his paper Semantic Web, identifies this system as an element that can help in the evolution of human knowledge as a whole.

«The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web. This structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work and learn together.» (Tim Berners Lee, 2001)

References

Auer S., Bizer C., Kobilarov G., Lehmann J. Cyganiak R., Ives Z. (2007). "DBpedia: A Nucleus for a Web of Open Data" ISWC/ASWC 2007, 4825, 722-735.

Berners-Lee T. (2000). Weaving the web, New York: Ed. HarperCollins.

Berners-Lee T. (2001). "The Semantic Web", Scientific American, 284, 34-43.

Bizer C., Heath T., Berners-Lee T. (2009). Linked Data—The Story So Far, in International Journal on Semantic Web and Information Systems, 5, 1–22.

Hitzler P., Janowicz K. (2013), "Linked Data, Big Data, and the 4th Paradigm", Semantic Web, 1, 1-3.

Negroponte N. (2004). Essere Digitali, Milano: Ed. Sperling & Kupfer.

Salarelli A. (2014). "Sul perché, anche nel mondo dei Linked Data, non possiamo rinunciare al concetto di documento" AIB Studi, 54, 279-293.

Signore O. (2002). "RDF per la rappresentazione della conoscenza". www.w3c.it

Signore O. (2005). "Semantic Web: il futuro è già qui?". www.w3c.it

Solodovnik I. (2015). Repository Istituzionali Open Access e strategie Linked Open Data, Firenze: Ed. Firenze University Press.

Townsend A. T. (2013). Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia, London: W. W. Norton&Company.