

Weaving a Web of Linked Resources

Editorial

Fabien Gandon^{a,*}, Marta Sabou^b and Harald Sack^c

^a *Wimmics, Université Côte d'Azur, Inria, CNRS, I3S, Sophia Antipolis, France*

E-mail: fabien.gandon@inria.fr

^b *CDL-Flex, Vienna University of Technology, Austria*

E-mail: marta.sabou@ifs.tuwien.ac.at

^c *FIZ Karlsruhe, Leibniz Institute for Information Infrastructure, KIT Karlsruhe, Germany*

E-mail: Harald.Sack@fiz-Karlsruhe.de

Abstract. This editorial introduces the special issue based on the best papers from ESWC 2015. And since ESWC'15 marked 15 years of Semantic Web research, we extended this editorial to a position paper that reflects the path that we, as a community, traveled so far with the goal of transforming the Web of Pages to a Web of Resources. We discuss some of the key challenges, research topics and trends addressed by the Semantic Web community in its journey. We conclude that the symbiotic relation of our community with the Web requires a truly multidisciplinary research approach to support the Web's diversity.

Keywords: Semantic Web, trends, challenges

1. From a Web of Pages to a Web of Resources

As we write this article Tim Berners-Lee just received the Turing Award “for inventing the World Wide Web, the first Web browser, and the fundamental protocols and algorithms allowing the Web to scale” [1]. But Tim not only invented the Web: he kept defending it year after year and continuously works to lead it to its full potential. In particular, a bit less than ten years after he submitted his proposal for an information management system in March 1989 [2], Tim wrote in September 1998 the Semantic Web Road map [3] to give a high-level plan of the architecture of the Semantic Web and as a continuation to his wish in 1994 to provide on the Web “more machine-oriented semantic information, allowing more sophisticated processing” [4]. This vision has been made visible to a broad audience in 2001 with an article in the *Scientific American* [5]. And again a few years later, Tim was instrumental in pushing what can be seen as a first wave of deployment of the Semantic Web with the Linked Data principles and the Linked Open Data

5-Star rules [6] leading to the publication and growth of linked open datasets towards the Web of Data, as depicted in Fig. 1 and Fig. 2

Twenty five years after its invention, the Web has exceeded its initial status of a distributed document-centric space. Following its numerous evolutions, it has become a virtual place where people and software can co-operate within hybrid communities [9]. It supports a mixed society where humans and Web robots interact in particular via shared metadata. Web sites such as Wikipedia, DBpedia, and Wikidata are the most prominent examples of this space, where software (Web robots) and humans interact in hybrid communities on a worldwide scale.

One of the major evolutions of the Web was the advent of the Web of Data based on Linked Data principles. It transformed the document centered Web into a distributed database. Linked Data provides an open interface based on widely used W3C standards, such as URIs (Uniform Resource Identifiers) as universal identifiers, HTTP (Hypertext Transfer Protocol) for web-based access, and RDF (Resource description Framework) as common standard for data encoding. These pillars enable efficient identification, access, and inter-

*Corresponding author. E-mail: fabien.gandon@inria.fr.

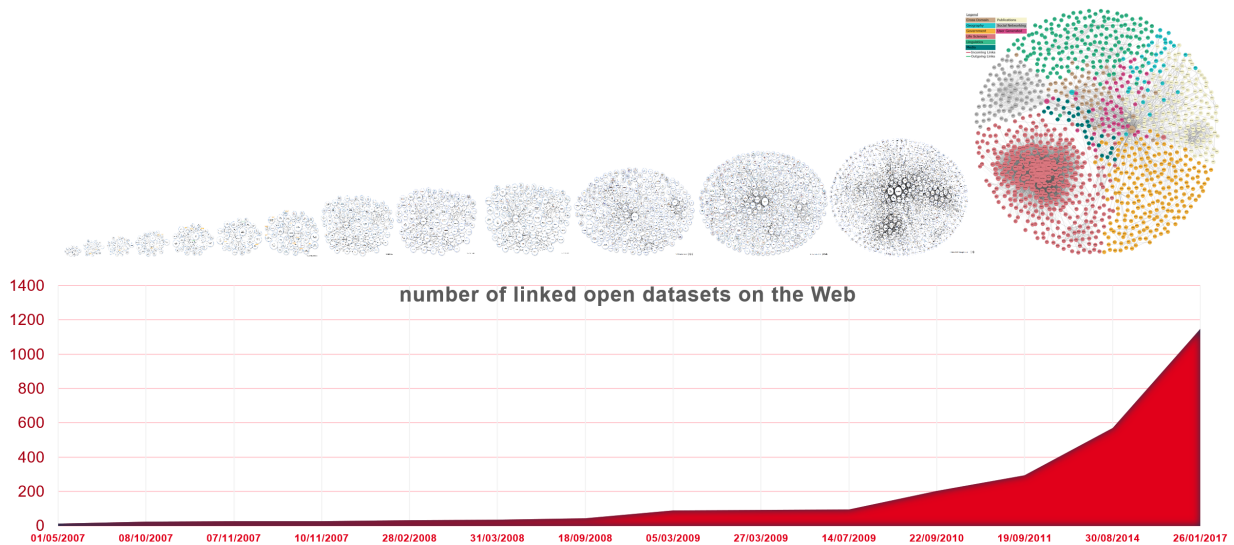


Fig. 1. Number of linked open datasets on the Web plotted from 2007 to 2017 with data from [7] and [8]

linking of data on the Web. Since its start in 2007, the Web of Data has grown up to almost 10,000 interlinked datasets covering a broad range of topics, as e.g., bibliographical data, geographical data, lexicographical data, biomedical data, or social networking data with DBpedia, the Linked Data version of the popular online encyclopedia Wikipedia, as its referential central hub (cf. Fig. 1, Fig. 2) [10].

2. This is for everything

During his participation in the London 2012 Summer Olympics opening ceremony, Tim Berners-Lee tweeted “This is for everyone” as a statement about the Web. With its fundamental idea of using URIs to literally identify and henceforth describe “every thing” around us, the Web is now used in every human activity. As a result many challenges have emerged from different visions of the Semantic Web and from its different contexts of use and the constraints they bring with them.

2.1. From V to S in data management

The so called “Vs” of big data (velocity, variety, volume, veracity) translate to many “S” of Semantic Web: scalability, storage, search, semantics, security, streaming, etc. In terms of data management, scalability of storage and efficiency of querying are active research and development domains to improve

RDF store performances. In terms of data access, many approaches now exist on a continuum from HTTP gets, to, Linked Data Fragments, Linked Data Platform REST approach, and SPARQL services, protocol and language. Moreover, to provide reliable, persistent, and trustworthy Linked Data services, topics such as access control, version management, long term preservation are of utmost importance. But again efficiency, federation and hybridization remain hot research questions.

2.2. Formal knowledge and artificial intelligence

On top of these core topics of data management and access, the Semantic Web community has always been interested in providing intelligent processing of the linked data of the Web, starting with reasoning. This remains a challenging and important topic with hard problems in scaling, approximating and distributing reasoning. In addition to classical logical derivation, many other artificially intelligent behaviours are studied in the community including machine learning and data mining, induction of knowledge, deontic reasoning on data licences, etc. In particular, hybrid approaches investigating whether and how techniques of description logic and reasoning can be combined with statistical machine learning are in the focus of current research with the promise to further improve artificial intelligence beyond the current state-of-the-art.

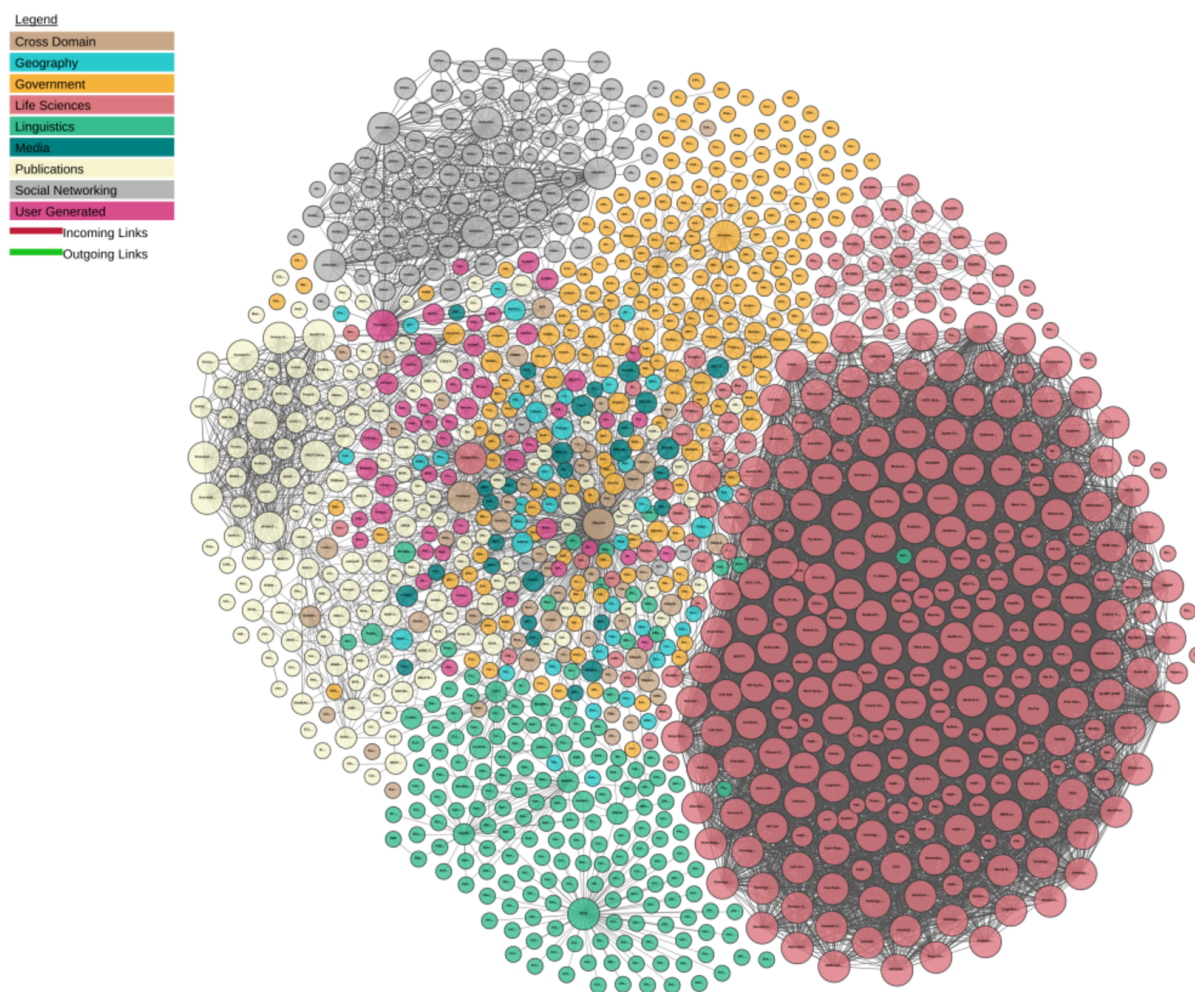


Fig. 2. The Linked Open Data Cloud diagram shows major datasets published in Linked Data format as of 2017-02-20 [8]

2.3. Heterogeneity of graph types and life cycles

The initial graph of linked pages of the Web has been extended by a growing number of other graphs including: sociograms capturing social network structures, workflows specifying decision paths to be followed, browsing logs capturing trails of navigation, automata of service compositions specifying distributed processing, linked open data from distant datasets, etc. Moreover, these graphs are distributed over many different sources with very different characteristics. Some sub-graphs are public (e.g. DBpedia), while others are private (e.g. semantic intrawebs). Some sub-graphs are small and local (e.g., a user's profile on a device), and some are huge and hosted on clusters (e.g., Wikipedia). Some are largely stable (e.g., a thesaurus for Latin),

some change several times per second (e.g., sensor data in a smart city), etc. And each type of graph of the Web is not an isolated island. Graphs interact with each other: the networks of communities influence the message flows, their subjects and types, the semantic links between terms interact with the links between sites and vice versa, the small changing graphs of sensors are joined to the large stable geographical graphs that position them, etc. Not only do we need methods to represent and analyse each kind of graph, we also require the means to combine them and to perform multi-criteria analyses on their combinations.

2.4. A truly open-world assumption

One of the major changes when trying to port database and knowledge base approaches to the Web is the open world assumption (OWA) principle that underlies Semantic Web technologies. Many existing results from more established research domains, such as databases, knowledge based systems or model-based engineering, have to be revisited to account for this open-world assumption. But in a broader sense, the open world of linked data on the Web also adds additional challenges to address uncertainty, data quality, data and processing traceability in order to be able to provide reasons for the users to trust the systems and their results.

2.5. Human-machine partnership

This expanding Web of data, together with the schemas, ontologies and vocabularies used to structure and link it, form a formal Semantic Web with which we have to design new interaction means to support the next generation of Web applications. The Semantic Web has a role to play in addressing challenges at the intersection of knowledge-based interactions and Web-augmented interactions [11]: there is not only a need for Human Computer Interaction (HCI) to provide methods to design human-data interaction applied to linked data on the Web, but also inversely a need for the Semantic Web community to investigate how linked data and the intelligent inferences they support can improve human-machine interactions. On the Web, large-scale interactions also create many challenges, and in particular the ongoing need to reconcile the formal semantics of computer science (logics, ontologies, typing systems, etc.) on which the Web architecture is built, with the soft semantics of people (posts, tags, status, and so on) through which a lot of the Web content is created. And as the Web becomes a ubiquitous infrastructure reflecting all the objects of our world, we witness ever-increasing frictions between formal semantics and social semantics. A promising research avenue to span the gap between formal and social semantics is the use of Human Computation and Crowdsourcing techniques to involve large, distributed groups of users in the various stages of the knowledge engineering life-cycle, from ontology modeling and verification to annotation, data curation and entity linking [12].

2.6. Conquering new application domains

The benefits of Semantic Web technologies had been firstly apparent for information-intensive domains such as medicine, library science or cultural heritage where taxonomy structures were already well accepted. With the advent of Linked Data technologies, our community increasingly addressed large-scale data integration problems with enterprises knowledge graphs. Recent years have seen the uptake of Semantic Web technologies in application domains that span the borders of the digital and physical world. As a result we explored the use of our technologies as integral part of cyber-physical systems such as adaptive traffic control systems in smart cities [13] or flexible assembly lines in smart factories [14].

3. Leading the Semantic Web to its full potential

3.1. Dynamics of research topics in Semantic Web

The previous topics are only a taste of the many important topics now addressed at international venues such as ISWC, ESWC and WWW conference series. As shown in Fig. 3 these topics were born, grew and went mainstream very dynamically over the past 14 years as we moved from solved to new challenges. Starting out with a small number of key topics focused on representation languages (rdf, web ontology language), query answering, ontology engineering and learning primarily, Semantic Web research has been extended with further topics including: semantic web services (with a peak of activity between 2002-2009), linked data (which represents an major research share since 2007) as well as data management topics such as ontology matching and SPARQL.

This special issue is based on the top articles selected from the 12th ESWC in 2015. Besides having a main focus on advances in Semantic Web research and technologies, we, the Chairs of ESWC 2015, decided to broaden the scope to span other relevant research areas as well. In particular, as a response to the emerging human-machine partnership challenge, the core tracks of the research conference were complemented with new tracks focusing on linking machine and human computation at Web scale (Cognition and Semantic Web, Human Computation and Crowdsourcing). The current special issues contains two extended papers from the Reasoning track and one extended paper from the new track on Cognition:

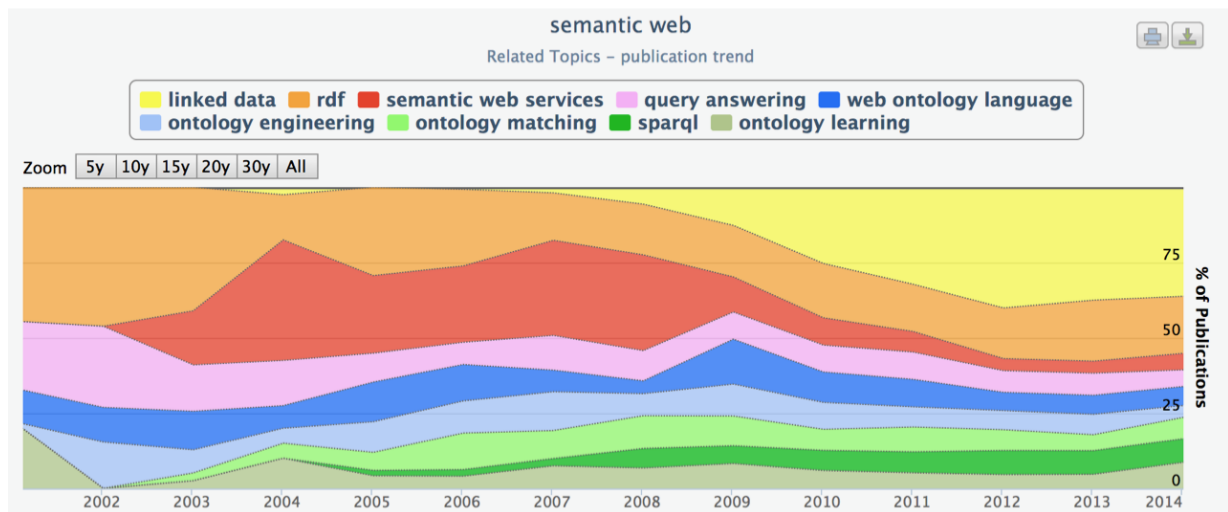


Fig. 3. Rexplore showing the evolution of major topics and keywords in the Semantic Web community over the last 14 years [15]

- *SPARQL with Property Paths on the Web*. This paper extends work in the area of querying Linked Data on the Web, thus addressing challenges related to data management and graph type heterogeneity. To query Linked Data on the Web, a range of approaches have been proposed spanning from centralized to fully distributed querying. SPARQL property paths are an extension of SPARQL that allow graph navigation and therefore better capture the distributed and graph-like feature of Linked Data datasets. Yet, the semantics of SPARQL restricts the applicability of this new construct to a single RDF graph. Hartig and Pirro address this gap and propose the formal foundations for evaluating property paths on the Web, in a distributed fashion. They propose two different semantics, reachability and context-based, and experimentally evaluate these. Interestingly, their results show that some queries cannot be evaluated over the Web in practice and motivate introducing the notion of Web-safe queries.
- *Ontology Understanding without Tears: The Summarization Approach*. To deal with the increasing size and complexity of RDF knowledge basis, methods are necessary for representing these graphs in a more concise manner in order to aid their quick understanding and further foster the human-machine partnership. A family of approaches in this direction are summarization approaches which identify a subgraph in the knowledge base which includes the most representative concepts of the schema. This paper presents two

algorithms for summarization which take into account both the schema and the instances in an RDF knowledge base in order to provide a concise summary that captures crucial information within that knowledge base. These algorithms are part of the RDF Digest platform that automatically produces and visualizes high quality summaries of RDF/S Knowledge Bases (KBs).

- *FrameBase: Enabling Integration of Heterogeneous Knowledge* addresses data management issues while revisiting our current assumptions about formal knowledge representation practices adopted in the Semantic Web. The representation of knowledge that involves more than two entities (i.e., n-ary relations) can be achieved using a number of different approaches when relying on the subject-predicate-object representations as defined by the RDF model. This increases semantic heterogeneity and hampers integration across knowledge sources. As a potential solution, this paper proposes the FrameBase knowledge base schema. Derived from the combination of the FrameNet and WordNet linguistic resources, FrameBase aims to offer a more concise and expressive approach for representing n-ary relations compared to triple-based approaches. The authors, subsequently demonstrate the support of FrameBase for integrating heterogeneous knowledge. This is achieved through integration rules that transform data from external knowledge bases into FrameBase instances.

3.2. Multidisciplinary approaches to support diversity

“We need diversity of thought in the world to face the new challenges.”

- Tim Berners-Lee

The participatory nature of the Web makes it emerge as an openly co-constructed global and inherently heterogeneous artifact. The “world-wide way” of deploying the Web everywhere and for everything implies that, as the Web is spreading into the world, the world is spreading into the Web. The resulting world “wild” Web that is being created and is evolving every day is affected by, and at the same time reflects, the complexity of our world. From a Semantic Web point of view, as soon as we want to model, analyse and combine these many facets of one Web, we face the general challenge of its diversity and span. This complexity implies that a huge challenge for Web development in general, and for the Semantic Web community in particular, is the resulting need for large-scale multidisciplinary cooperation: the three ‘W’s of the World Wide Web call for the three ‘M’s of a Massively Multidisciplinary Methodology [16], and the Semantic Web is no exception to this.

References

- [1] ACM, 2016 A.M. Turing Award Citation For inventing the World Wide Web, the first web browser, and the fundamental protocols and algorithms allowing the Web to scale., 2017. http://amturing.acm.org/award_winners/berners-lee_8087960.cfm.
- [2] T.J. Berners-Lee, Information management: A proposal, Technical Report, 1989. <http://cds.cern.ch/record/1405411/files/ARCH-WWW-4-010.pdf>.
- [3] T. Berners-Lee, Semantic web road map, 1998. <https://www.w3.org/DesignIssues/Semantic.html>.
- [4] R.C. Tim Berners-Lee et al., The World-Wide Web, *Commun. ACM* **37** (1994), 76–82. doi:10.1145.179606.179671. <https://doi.org/10.1145/179606.179671>.
- [5] J.H. Tim Berners-Lee and O. Lassila, The semantic web, *Scientific american* **284**(5) (2001), 28–37.
- [6] T. Berners-Lee, Linked Data, 2006. <https://www.w3.org/DesignIssues/LinkedData.html>.
- [7] Linked Open Data, 2017. <http://linkeddata.org/>.
- [8] The Linked Open Data cloud diagram, 2017. <http://lod-cloud.net/>.
- [9] F. Gandon et al., Challenges in Bridging Social Semantics and Formal Semantics on the Web, in: *International Conference on Enterprise Information Systems*, Springer, 2013, pp. 3–15.
- [10] Linked Data Stats, 2017. <http://stats.lod2.eu/>.
- [11] F. Gandon and A. Giboin, Paving the WAI: Defining Web-Augmented Interactions, in: *International Conference Web Science*, ACM, 2017.
- [12] C. Sarasua, E. Simperl, N. Noy, A. Bernstein and J.M. Leimeister, Crowdsourcing and the Semantic Web: a Research Manifesto, *Human Computation* **2**(1) (2015), 3–17.
- [13] M. d’Aquin, J. Davies and E. Motta, Smart Cities’ Data: Challenges and Opportunities for Semantic Technologies, *IEEE Internet Computing* **19**(6) (2015), 66–70.
- [14] S. Biffl and M. Sabou, *Semantic Web Technologies for Intelligent Engineering Applications*, Springer, 2016.
- [15] F. Osborne, E. Motta and P. Mulholland, Exploring scholarly data with rexplore, in: *International semantic web conference*, Springer, 2013, pp. 460–477.
- [16] F. Gandon, The Three ‘W’ of the World Wide Web Call for the Three ‘M’ of a Massively Multidisciplinary Methodology, in: *International Conference on Web Information Systems and Technologies*, Springer, 2014, pp. 3–15.