

# Structured Knowledge on the Web 7.0

Steffen Staab

Universität Koblenz-Landau, DE &  
University of Southampton, UK  
staab@uni-koblenz.de, s.r.staab@soton.ac.uk

Jens Lehmann

University of Bonn, Germany &  
Fraunhofer IAIS, Germany  
jens.lehmann@cs.uni-bonn.de

Ruben Verborgh

Ghent University — imec  
Belgium  
Ruben.Verborgh@UGent.be

## ABSTRACT

Structured Knowledge on the Web had an intriguing history before it has become successful. We briefly revisit this history, before we go into the longer discussion about how structured knowledge on the Web should be devised such that it benefits even more applications. Core to this discussion will be issues like trust, information infrastructure usability and resilience, promising realms of structured knowledge and principles and practices of data sharing.

## CCS CONCEPTS

• **Information systems** → **Web data description languages**; *Data management systems*; Social networks;

## KEYWORDS

RDF, Semantic Web, Knowledge Graph, Schema.org, Linked Data

## ACM Reference Format:

Steffen Staab, Jens Lehmann, and Ruben Verborgh. 2018. Structured Knowledge on the Web 7.0. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3184558.3190666>

## 1 A BRIEF HISTORY

Structured knowledge on the Web has a rich history, most, but not all of which is deeply interwoven with the Semantic Web, its standards and its research community. Unsurprisingly, this history included failures, such as the very beginning of structured knowledge on the Web as foreseen with the HTML META tag, which was predominantly used for spam and hence widely ignored by search engines and other applications.

A second generation of more powerful knowledge structures were crafted with the development of SHOE [5] and Ontobroker [3]. They may count as precursors to the third and fourth generation of knowledge structures on the Web that were standardized with the likes of the RDF graph data model, RDF-A, OWL, and the — much later — but impactful JSON-LD.

For many years, these languages and the practices around them were considered as futile by industry and indeed by a significant part of the research community. Who would ever annotate his/her content? Why would companies deliver well-structured data for

free, would this not adversely impact their profits? Would the Semantic Web not immediately fall prey to spam such as occurred to the HTML META tag before?

Ten years ago two ideas were conceived that made a big difference to how Web developers would view structured knowledge on the Web. The DBpedia [2] and Freebase [4] *knowledge graphs* were taken as blueprints for managing (Web) content by a multitude of companies. Almost at the same time Yahoo!SearchMonkey found a way to (mostly) de-incentivize the spamming observed with HTML META and encourage people to provide appropriately structured knowledge, a practice that has converged into Schema.org with trillions of corresponding triples now found by search engines on the Web.

## 2 THE FUTURE OF STRUCTURED KNOWLEDGE ON THE WEB

We consider the ongoing development of structured knowledge on the Web, the challenges one encounters when publishing knowledge and data and the applications that might be built. To this end we discuss questions like the following ones:

*Which data publishing paradigm?* Linked data has been promoted as a data publishing paradigm that reflects well the autonomy of content providers over entities they maintain. Drawbacks include the necessity to digest knowledge in tiny fractions that are hard to scale. SPARQL endpoints offer full access to data, but put heavy loads on servers and cost models are still discussed. Approaches such as triple pattern fragments [7] try to strike a balance between the piecemeal approach of Linked Data and the heavy machinery of SPARQL endpoints.

*How much consistency?* Structured knowledge on the Web must be open by definition. Yet how much consistency can be enforced in order to achieve guarantees when programming with such data? Where are the most pressing pain points: Having more inference or improving the data handling for the developer? How will ontology languages such as the OWL family and constraints, such as SHACL, be used together or apart?

*Which structured knowledge should be represented on the Web?* Until now schematic and factual knowledge dominates on the Web, while approaches like ConceptNet [6] or argumentative knowledge [1] cannot fully be captured in existing languages. The Semantic Web pursued a strategy of avoiding over-abundance of knowledge representation paradigms that would not converge to common denominators useful for adoption and reuse. For many domains, e.g. Web Services, Semantic Web defined abstract syntax rather than full-fledged representation languages. Should this strategy be

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3190666>

(dis-)continued? Should or will this also be the case for multi-modal knowledge, such as visual, perceptual or auditory knowledge?

*How is knowledge distributed across the Web?* Knowledge distribution on the Web often follows a long tail, with a small number of large datasources, and a very large number of smaller datasources. We need to think about where data should be published, replicated or aggregated, and how we will discover and query it. Distribution also comes with challenges regarding provenance and trust.

*In which knowledge we trust?* Initially, structured knowledge was (and is continued to be) spammed, today we must fight fake news and other kind of misinformation. How do we tackle these challenges? How do we build up infrastructures for signing knowledge on the Web? Which lessons can we take from other areas of the Web of how to counter attacks and incur a more trusted knowledge infrastructure? We might even talk about knowledge on the blockchain, why not?

*New applications:* How do we extend the reach of structured knowledge into new applications? How is the publishing process furthered? Why do we not see more knowledge in applications like  $\text{\LaTeX}$ , Word<sup>TM</sup>, or PDF? Scientific data publishing targets the four principles (FAIR): (i), findable, (ii), accessible, (iii), interoperable and, (iv), re-usable, which seems to be greatly furthered by structured data on the Web — why don't we see more take-up then and how to we need to meet our customers/colleagues from data publishing?

### 3 CONCLUSION

“The Semantic Web cannot work.” was a widely held belief, which has been proven wrong. Just like the Web of 2018 does not look like anyone might have imagined, when the first Web conference convened in 1994, the structured knowledge we find on the Web now is there for other purposes and enriched by other languages than one might have imagined in the late 90ies for the Semantic Web. Novel opportunities are immediate — as well as new challenges.

### REFERENCES

- [1] Katie Atkinson, Pietro Baroni, Massimiliano Giacomin, Anthony Hunter, Henry Prakken, Chris Reed, Guillermo Ricardo Simari, Matthias Thimm, and Serena Villata. 2017. Towards Artificial Argumentation. *AI Magazine* 38, 3 (2017), 25–36. <https://www.aaai.org/ojs/index.php/aimagazine/article/view/2704>
- [2] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007*. 722–735. [https://doi.org/10.1007/978-3-540-76298-0\\_52](https://doi.org/10.1007/978-3-540-76298-0_52)
- [3] V. Richard Benjamins, Dieter Fensel, Stefan Decker, and Asunción Gómez-Pérez. 1999. (KA)<sup>2</sup>: building ontologies for the Internet: a mid-term report. *Int. J. Hum.-Comput. Stud.* 51, 3 (1999), 687–712. <https://doi.org/10.1006/ijhc.1999.0275>
- [4] Kurt D. Bollacker, Robert P. Cook, and Patrick Tufts. 2007. Freebase: A Shared Database of Structured General Human Knowledge. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada*. 1962–1963. <http://www.aaai.org/Library/AAAI/2007/aaai07-355.php>
- [5] Sean Luke and James A. Hendler. 1997. Web Agents That Work. *IEEE MultiMedia* 4, 3 (1997), 76–80. <https://doi.org/10.1109/93.621584>
- [6] Robert Speer, Joshua Chin, and Catherine Havasi. 2017. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*. 4444–4451. <http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14972>
- [7] Ruben Verborgh, Miel Vander Sande, Olaf Hartig, Joachim Van Herwegen, Laurens De Vocht, Ben De Meester, Gerald Haesendonck, and Pieter Colpaert. 2016. Triple Pattern Fragments: A low-cost knowledge graph interface for the Web. *J. Web Sem.* 37-38 (2016), 184–206. <https://doi.org/10.1016/j.websem.2016.03.003>