

Semantic Web: A 20-year Perspective



Pascal Hitzler

Data Semantics Laboratory (DaSe Lab)
Kansas State University

<http://www.daselab.org>

A Review of the Semantic Web Field

By Pascal Hitzler

Communications of the ACM, February 2021, Vol. 64 No. 2, Pages 76-83

10.1145/3397512

[Comments](#)

VIEW AS:



SHARE:



Credit: Andrii Borys Associates.

Let us begin this review by defining the subject matter. The term *Semantic Web* as used in this article is a field of research rather than a concrete artifact—in a similar way, say, *Artificial Intelligence* denotes a field of research rather than a concrete artifact. A concrete artifact, which may deserve to be called "The Semantic Web" may or may not come into existence someday, and indeed some members of the research field may argue that part of it has already been built. Sometimes the term *Semantic Web technologies* is used to describe the set of methods and tools arising out of the field in an attempt to avoid terminological confusion. We will come back to all this in the article in some way; however, the focus here is to review the research field.

This review will be rather subjective, as the field is very diverse



2001-2006: Ontologies

Thesis

2006-2012: Linked Data

Antithesis

Since 2012: Knowledge Graphs & Wikidata

Synthesis

Thesis

Ontologies

2001-2006

The Hype: **Ontologies will enable Artificial General Intelligence
(it wasn't quite put that way, but see next slide)**

The Crash: **They didn't even come close.
They didn't even really get re-used.**

Thesis: Ontologies



Tim Berners-Lee, James Hendler, Ora Lassila: The Semantic Web. Scientific American, May 2001.

Agents that seamlessly and ad-hoc interact with structured (semantic) data on the Web, solving all kinds of tasks.

Earlier work:

SNOMED CT, 1960s

“Ontology” definition by Gruber, 1993

RDF first draft, 1997

Gene Ontology, 1998

DARPA DAML program, 2000-2006

EU On-To-Knowledge, 2000-2002

DAML+OTK merge into the OWL standard, 2004

Thesis: the good stuff

- **Some ontologies matured during this time and became very important, e.g.**
 - **the Gene Ontology**
 - **SNOMED**
- **Standards matured**
 - **OWL and RDF, 2004, V2 in 2012**
 - **SPARQL 2008, V2 in 2013**
- **A very active and diverse community grew to propel semantic technologies forward, including strong industry interest.**



Thesis: Ontologies



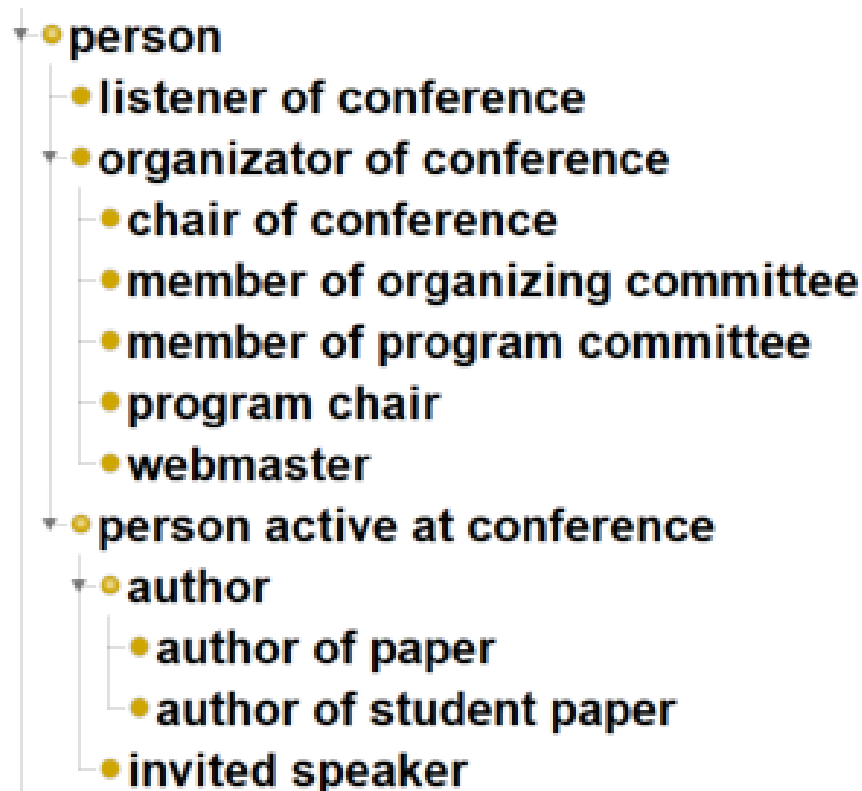
- **“A formal, explicit specification of a shared conceptualization”**
- **Main vehicle for data integration, sharing, discovery, re-use.**
- **A key idea is that ontologies are highly re-useable and connectable/mappable.**
- **However ontology development was often very ad-hoc, and resulting ontologies were**
 - **hard to understand**
 - **hard to maintain**
 - **hard to re-use**
 - **expensive to produce**
 - **of often very doubtful “quality”**

Ontological Problems

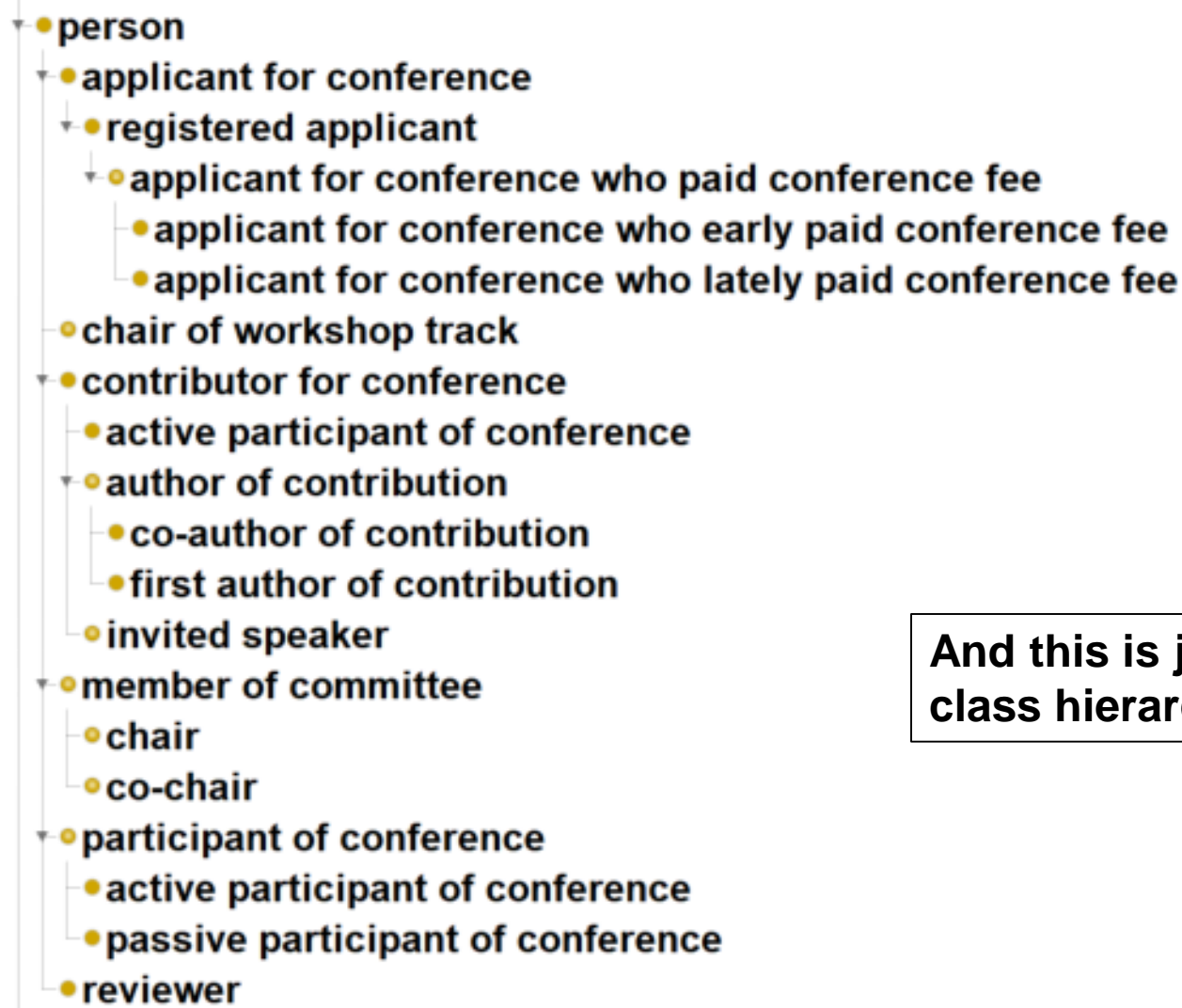


- Large, complex, with little or no internal structure
- Insufficient documentation.
- Non-obvious design choices, unexplained.
- Laden with ambiguity.

➔ Very difficult to understand what's going on inside.



Ontological Problems



And this is just looking at the class hierarchy!

Ontological Problems



- **Designed for single use case.**
 - **Granularity of representation highly varying.**
 - **Large and monolithic, hard to assess what any change will entail.**
- ➔ **Very difficult to adapt to a new setting.**

Arguably:

- **If often seemed easier to make a new ontology from scratch than to try understand an existing one and adapt it to your requirements.**

Antithesis

Linked Data

2006-2012



A few years later

- **Most ontologies were just a mess.**
- **Hardly any were reused or readily re-useable.**

Loudly voiced tenor:

- **“Ontologies don’t work”**
- **“Just scrap them”**

The Hype: **Linked Data will solve all your data management problems**

The Crash: **It didn't even come close.
Linked Data didn't even really get re-used.**

Antithesis: Linked (Open) Data

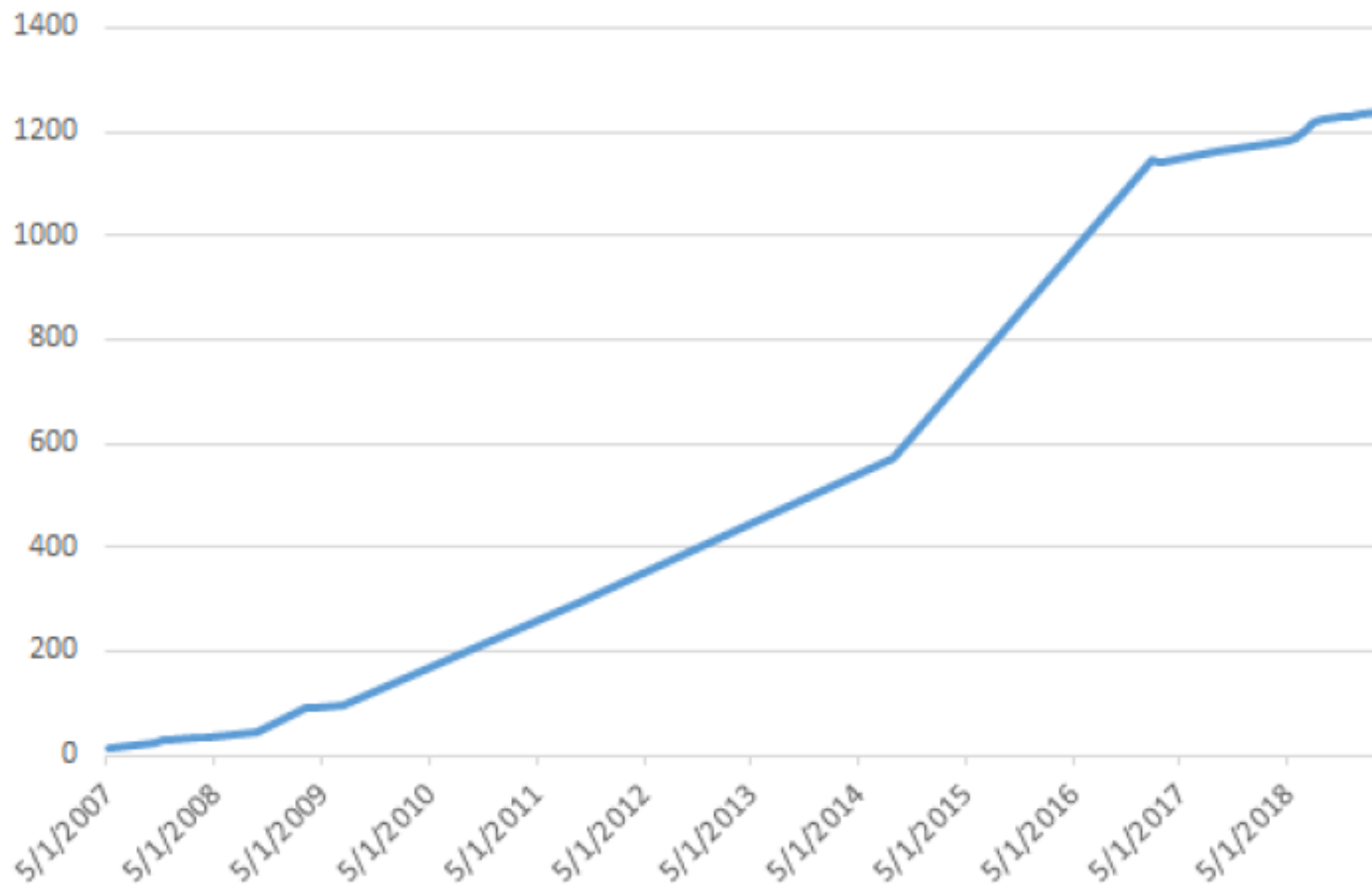


- **Just put your data on the web in RDF format.**
- **Or at least make it accessible through a SPARQL endpoint.**
- **And make sure it has plenty of links to other Linked Datasets.**

- **This will create a network of smart data that will solve all the data management problems that ontologies weren't able to solve.**

In particular: Very little concern regarding the organization or structure (schema) of the graphs, which could have been informed by ontologies!

Linked Data



Number of datasets in the linked data (interconnected) “cloud”.

Linked Data



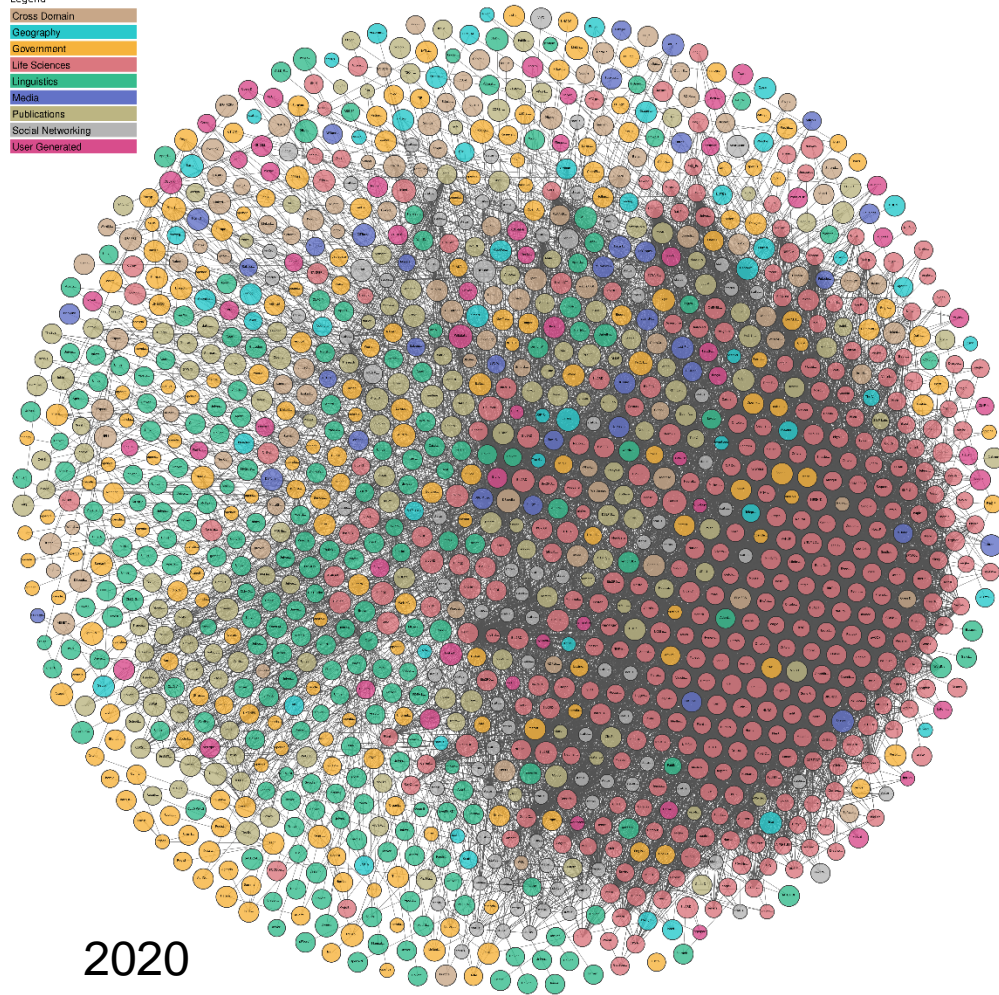
2015 (LOD Lab):

over 37,000,000,000 triples from over 650,000 data documents

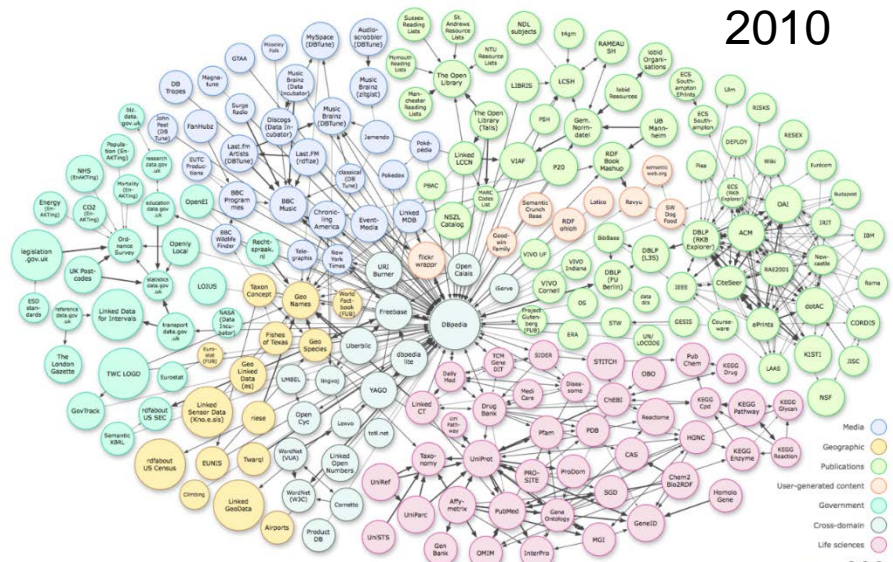


2007

- Legend
- Cross Domain
 - Geography
 - Government
 - Life Sciences
 - Linguistics
 - Media
 - Publications
 - Social Networking
 - User Generated



2020

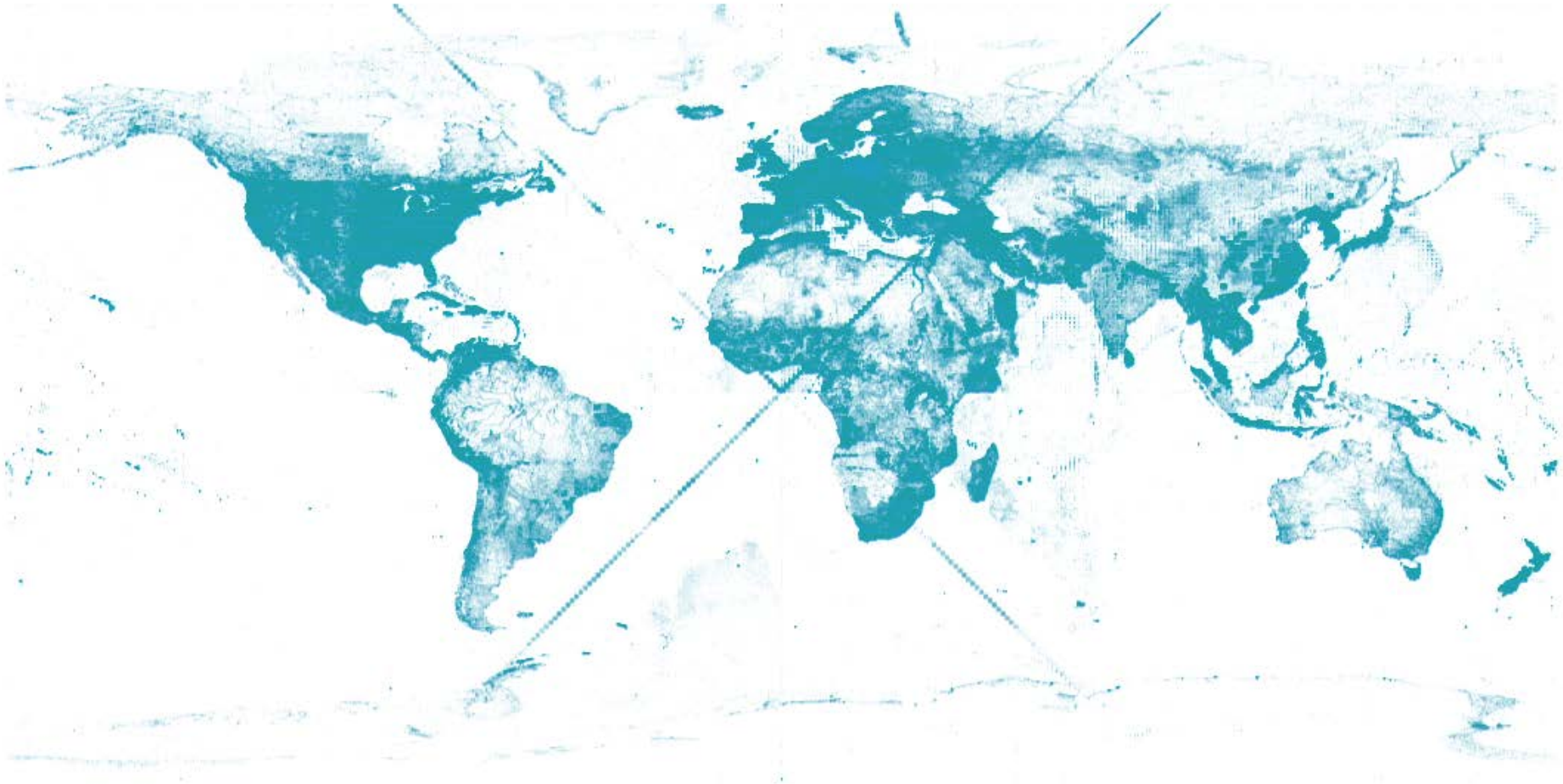


2010

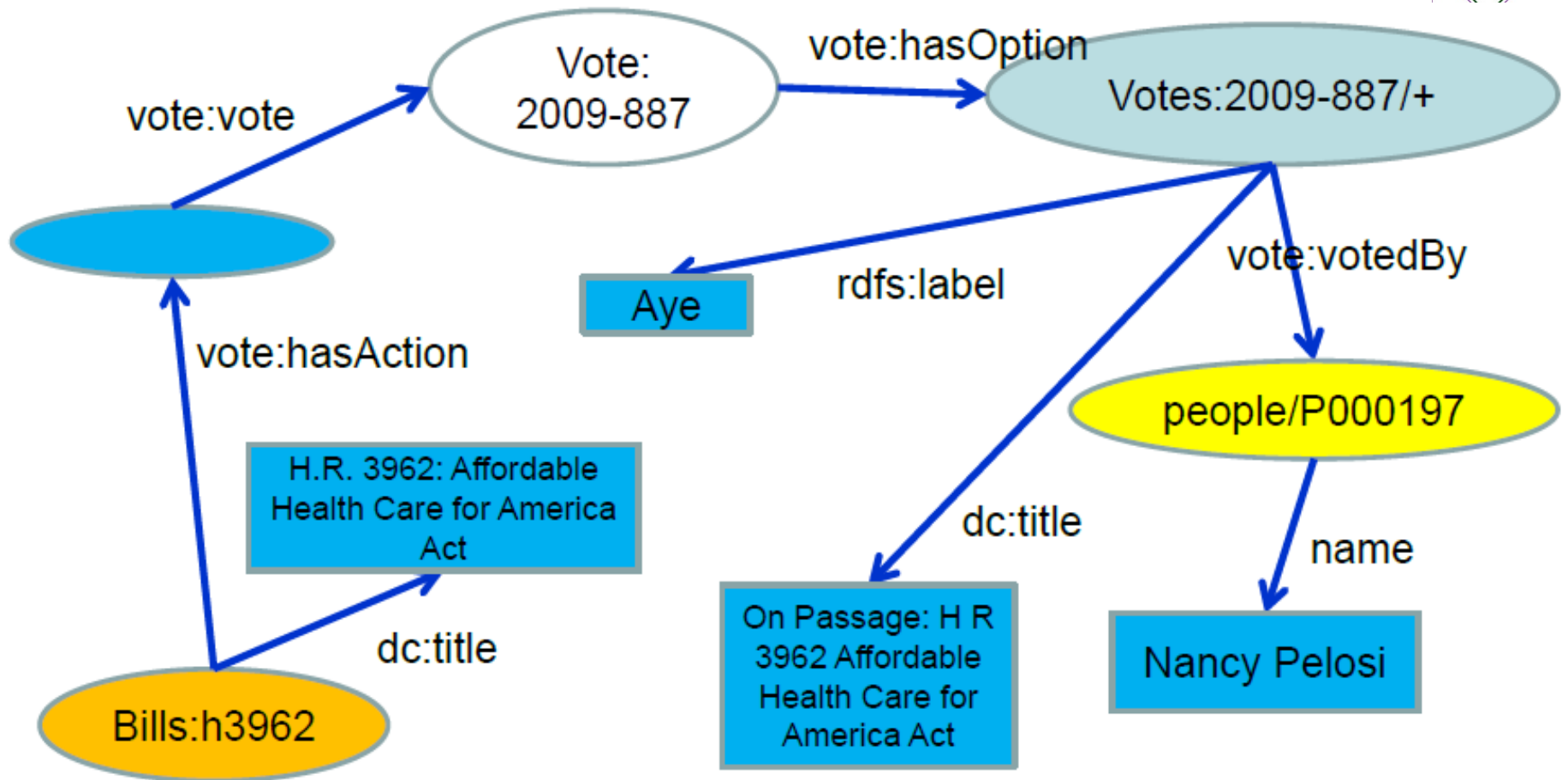
Linked Data

Geoindexed Linked Data – courtesy of Krzysztof Janowicz, 2012

http://stko.geog.ucsb.edu/location_linked_data



Linked Data Problems



Linked Data Problems



wd:Q56331940 wdt:P31 wd:Q5 ;
 wdt:P1416 wd:Q28942361,
 wd:Q63455745 ;
 wdt:P856 <<https://seco.cs.aalto.fi/u/eahyvone/>>,
 <<https://people.aalto.fi/eero.hyvonen>>,
 <<https://researchportal.helsinki.fi/en/persons/af1ac6a6-d9df-4233-bbd7-b31017d2c7d5>> ;
 wdt:P21 wd:Q6581097 ;
 wdt:P108 wd:Q300980,
 wd:Q28695 ;
 wdt:P2038 "Eero_Hyvoenen" ;
 wdt:P496 "0000-0003-1695-5840" ;
 wdtn:P496 <<https://orcid.org/0000-0003-1695-5840>> ;
 wdt:P1153 "8435405300" ;
 wdt:P185 wd:Q59247691 ;
 wdt:P106 wd:Q1650915,
 wd:Q82594 ;
 wdt:P735 wd:Q1295465 ;
 wdt:P734 wd:Q27890983 ;
 wdt:P27 wd:Q33 ;
 wdt:P2671 "/g/1hcnmkm1r" ;
 wdtn:P2671 <<http://g.co/kg/g/1hcnmkm1r>> ;
 wdt:P6304 "eero-hyv%C3%B6nen" ;
 p:P31 s:Q56331940-7E5C71F5-2C46-4766-BEB3-47256EB2265F .

In the early 2010s it dawned on the community that this is not going to work.

So what next?



Synthesis

Knowledge Graphs & Wikidata

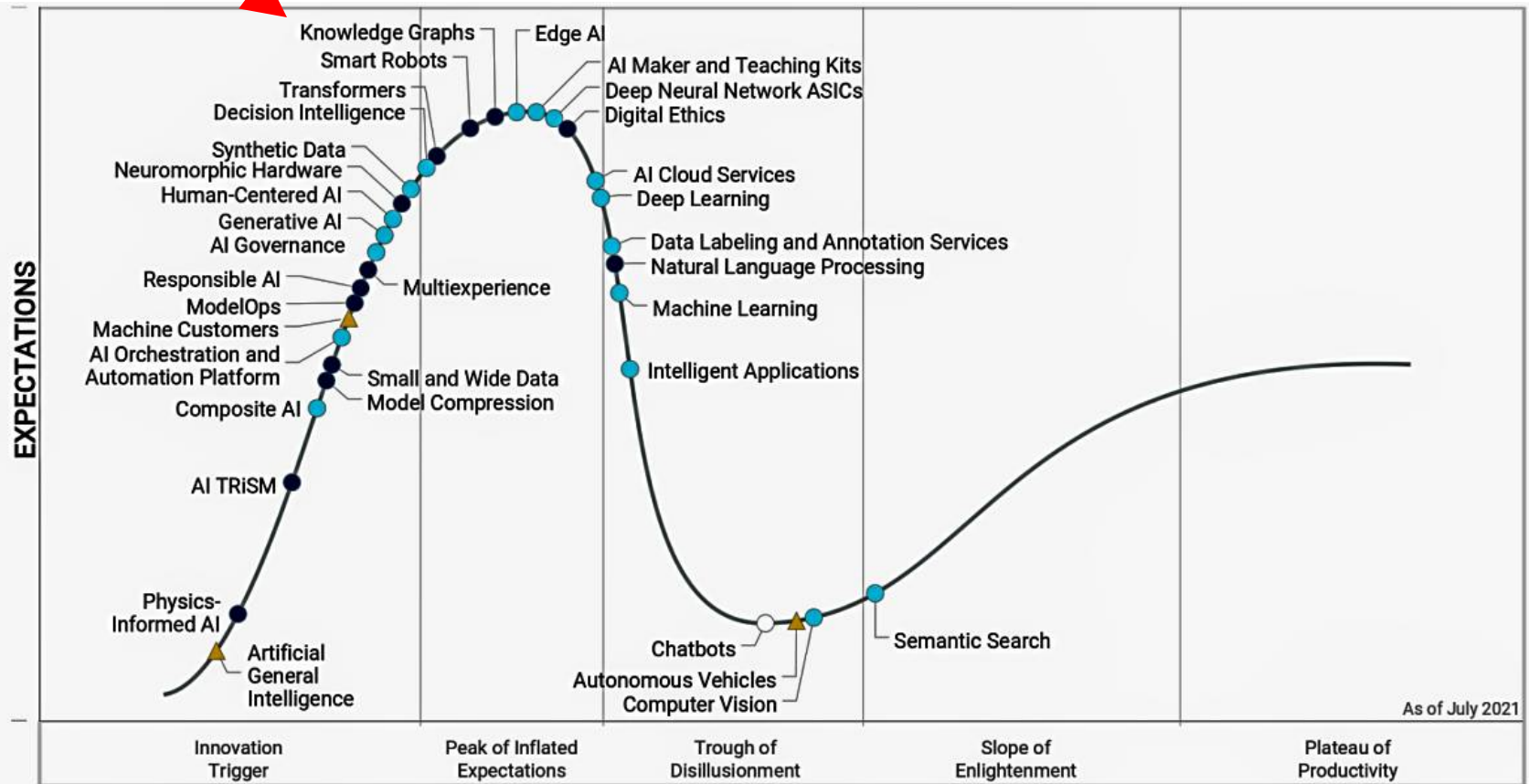
since 2012

HyperCrash



The Hype: **Google is doing it!**
(Google Knowledge Graph launched 2012)

July 2021 Gartner Hypecycle



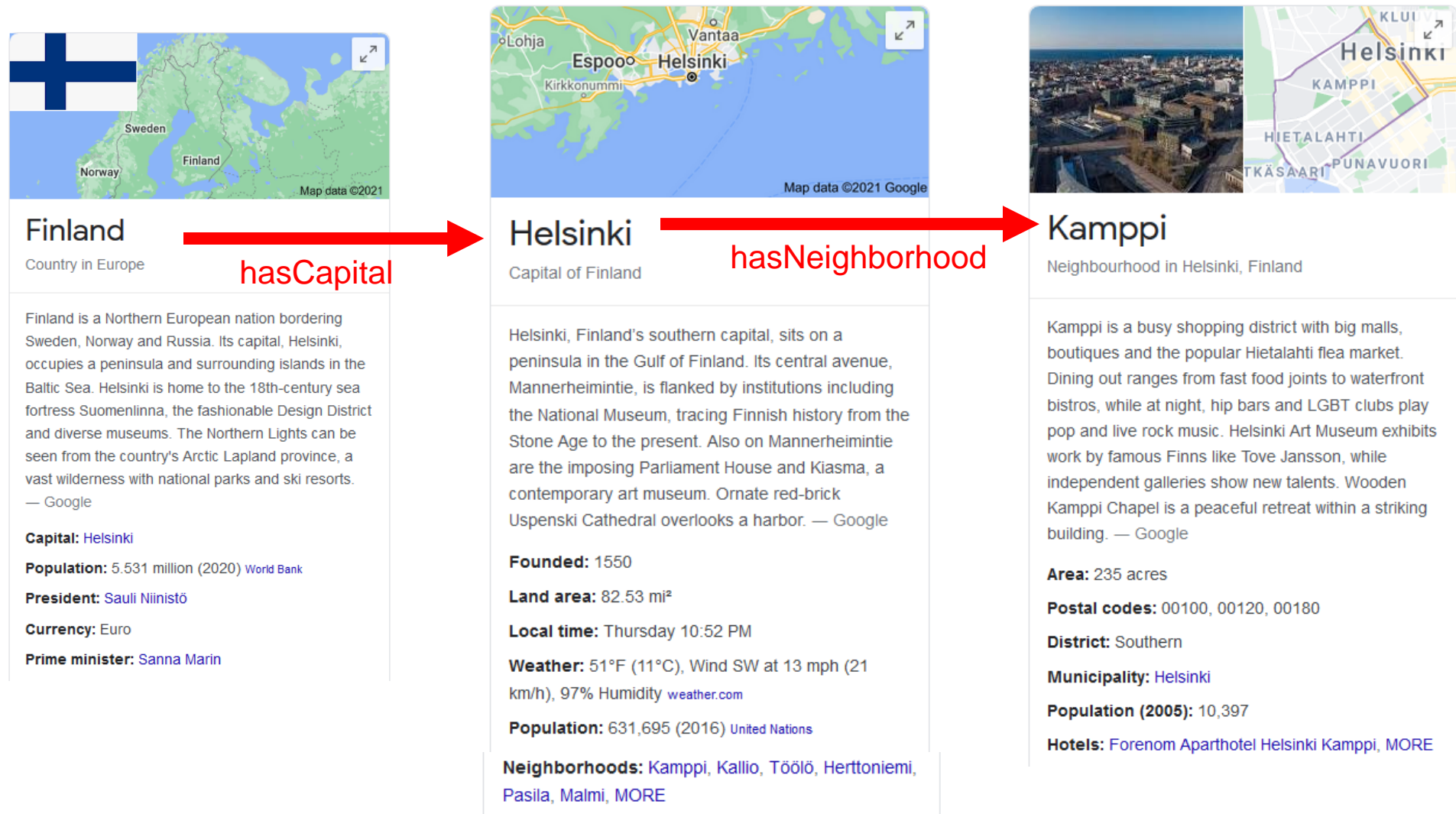
As of July 2021

TIME

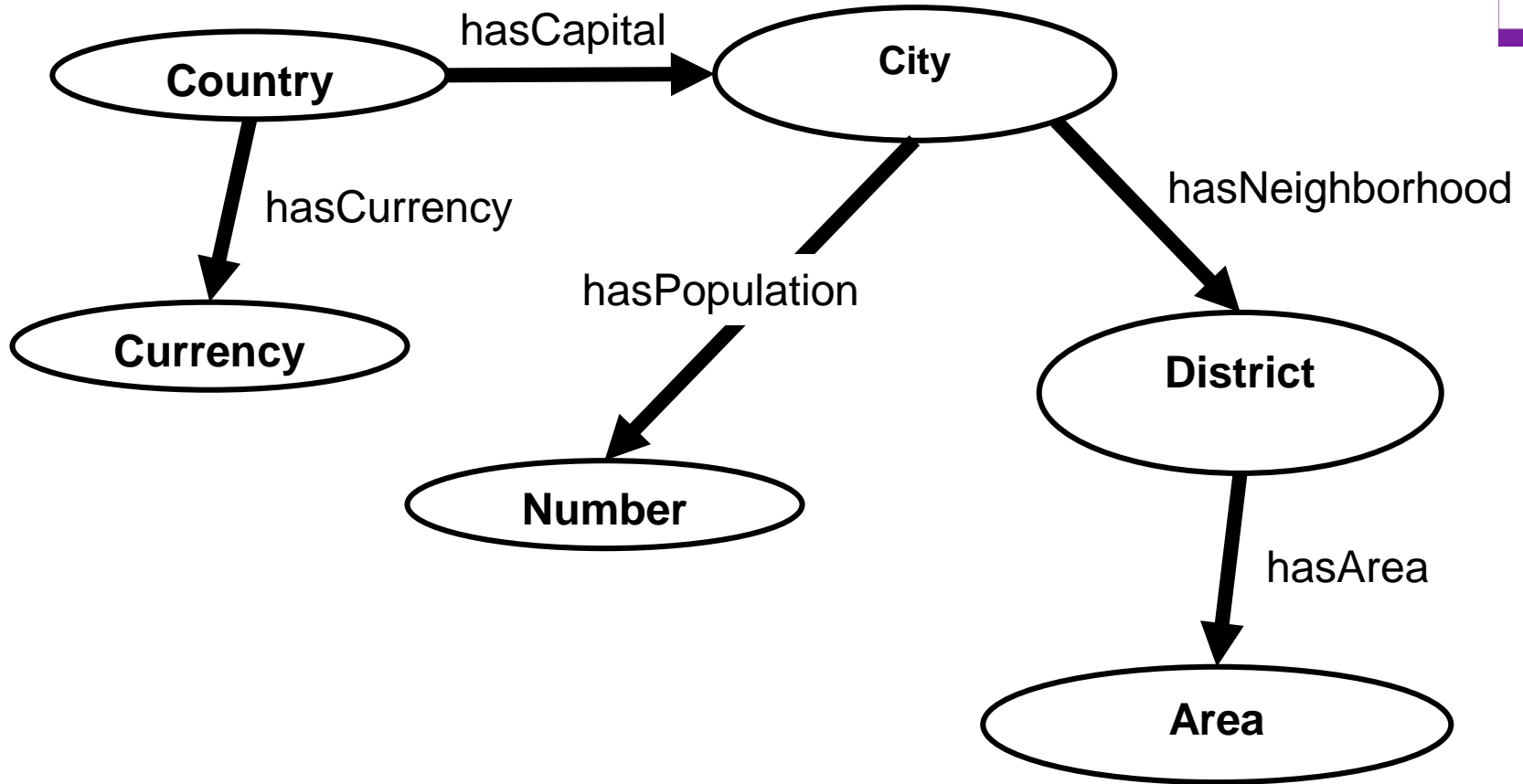
Plateau will be reached: ○ <2 yrs. ● 2-5 yrs. ● 5-10 yrs. ▲ >10 yrs. ✗ Obsolete before plateau

Ack: Gartner

Google Knowledge Graph



Schema (as diagram)



A good schema is critical for ease of reuse

PRACTICE

Industry-Scale Knowledge Graphs: Lessons and Challenges

By Natasha Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, Jamie Taylor

Communications of the ACM, August 2019, Vol. 62 No. 8, Pages 36-43

10.1145/3331166

[Comments](#)

VIEW AS:



SHARE:



Credit: Adempercem / Shutterstock

in knowledge graphs by defining a *schema* or *ontology*. For example, a link from a movie to its director must connect an object of type *Movie* to an object of type *Person*. In some cases the links themselves might have their own properties: a link connecting an actor and a movie might have the name of the specific role the actor



Knowledge graphs are critical to many enterprises today: They provide the structured data and factual knowledge that drive many products and make them more intelligent and "magical."

In general, a knowledge graph describes objects of interest and connections between them. For example, a knowledge graph may have nodes for a movie, the actors in this movie, the director, and so on. Each node may have properties such as an actor's name and age. There may be nodes for multiple movies involving a particular actor. The user can then traverse the knowledge graph to collect information on all the movies in which the actor appeared or, if applicable, directed.

Many practical implementations impose constraints on the links

SIGN IN for Full Access

» [Forgot Password?](#)

» [Create an ACM Web Account](#)

SIGN IN

ARTICLE CONTENTS:

[Introduction](#)

[What's In a Graph? Design](#)

[Decisions](#)

[Challenges Ahead](#)

[Other Key Challenges](#)

[Conclusion](#)

[References](#)

[Authors](#)

MORE NEWS & OPINIONS

MIT Robot Could Help People



Knowledge Graphs are like linked data, with

- **linking de-emphasized**
- **openness de-emphasized**
- **renewed consideration of schema quality**
- **tighter and central control**
- **clear transition from academia to industry**

Goal is to produce a flexible, extendable, end-user friendly but in the end rather tightly controlled repository of integrated, re-useable data.

<insert a ton of Eero's work>



Started 2012 at Wikimedia Deutschland, with funding by Google, Yandex, Allen Institute for AI.

**Crowdsourced knowledge graph,
like Wikipedia is a crowdsourced encyclopedia.**

Wikibase Software, Mediawiki interface.

Constrains graph shapes wrt. context information, references, datatype use.

Otherwise crowdsourced.



Eero Hyvönen (Q56331940)

Finnish professor

[E. Hyvönen](#) | [E Hyvönen](#) | [Hyvönen](#) | [Hyvönen E](#) | [Hyvönen E.](#) | [Eero Hyvonen](#) | [E. Hyvonen](#) | [Hyvonen E.](#)

▼ [In more languages](#)

[Configure](#)

Language	Label	Description
English	Eero Hyvönen	Finnish professor

doctoral student



Osma Suominen

 [edit](#)

[start time](#) 2010

[end time](#) 2013

▼ [1 reference](#)

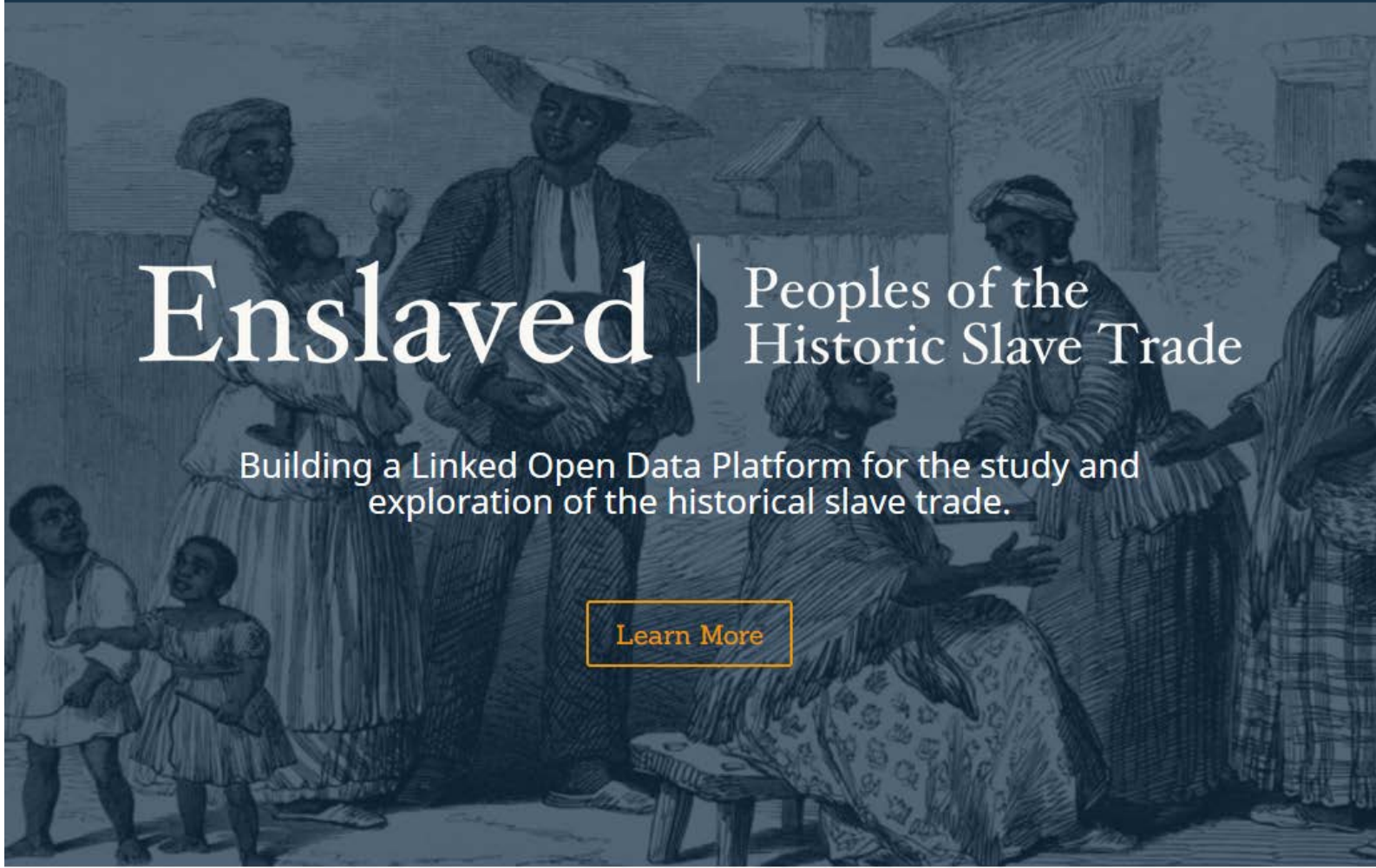
[stated in](#) [LinkedIn](#)

[reference URL](#) <https://www.linkedin.com/in/osmasuominen/>

[+ add reference](#)

[+ add value](#)

<https://www.wikidata.org/wiki/Q56331940>



Enslaved

 | Peoples of the
Historic Slave Trade

Building a Linked Open Data Platform for the study and exploration of the historical slave trade.

[Learn More](#)

- **Quality Ontology Design.**
- **Realization of ontology-based schema in Wikibase.**
- **Knowledge graph construction and interaction through Wikibase as engine.**

Similar:

The EU Knowledge Graph

Welcome to the EU Knowledge Graph!

This graph contains structured information about the European Union.

In particular, it contains information about:

- institutions of the European Union (like the [European Parliament](#) and the [Council of the European Union](#))
- countries in the world and in particular countries in the European Union (like [Hungary](#) and [Italy](#))
- capitals of European Countries (like [Athens](#) and [Tallinn](#))



- **Knowledge Graphs are currently everywhere.**
- **Wikidata / Wikibase increase in importance**
- **Semantic Web for Cultural Heritage has become a substantial community**
- **Explainable AI – explanations based on (knowledge graph) background knowledge attempts to merge Semantic Web and Deep Learning.**



Thanks!

References



- Pascal Hitzler, Semantic Web: A Review of the Field. *Communications of the ACM* 64 (2), 76-82, 2021.
- Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. *Scientific American* 284, 5 (May 2001), 34–43.
- <https://www.snomed.org/>
- Gruber, T. A translation approach to portable ontology specifications. *Knowledge Acquisition* 5, 2 (1993), 199–220.
- The Gene Ontology Consortium. The Gene Ontology Project in 2008. *Nucleic Acids Research* 36 (Database issue) (2008), D440–D444.
- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P., and Rudolph, S. (Eds.). *OWL 2 Web Ontology Language: Primer (2nd Ed.)*. W3C Recommendation 11 (Dec. 2012); <http://www.w3.org/TR/owl2-primer/>.

References



- Schreiber, G and Raimond, Y (Eds.). **RDF 1.1 Primer**. W3C Working Group Note (June 24, 2014);
<http://www.w3.org/TR/rdf11-primer/>.
- The W3C SPARQL Working Group (Ed.). **SPARQL 1.1 Overview**. W3C Recommendation (Mar. 21, 2013);
<http://www.w3.org/TR/sparql11-overview>.
- Hitzler, P., Krötzsch, M., and Rudolph, S. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2010.
- Bizer, C., Heath, T., and Berners-Lee, T. Linked Data—The story so far. *Int. J. Semantic Web Inf. Syst.*, 3 (2009), 1–22.
- Lehmann, J. et al. DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.
- Vrandečić, D. and Krötzsch, M. Wikidata: A free collaborative knowledgebase. *Commun. ACM* 57, 10 (Oct. 2014), 78–85.

References

- Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., and Taylor, J. Industry-scale knowledge graphs: lessons and challenges. *Commun. ACM* 62, 8 (Aug. 2019), 36–43.
- Cogan Shimizu, Pascal Hitzler, Quinn Hirt, Dean Rehberger, Seila Gonzalez Estrecha, Catherine Foley, Alicia M. Sheill, Walter Hawthorne, Jeff Mixter, Ethan Watrall, Ryan Carty, Duncan Tarr: The Enslaved ontology: Peoples of the historic slave trade. *J. Web Semant.* 63: 100567 (2020)
- Dennis Diefenbach, Max De Wilde, Samantha Alipio: Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph. *ISWC 2021*: 631-647
- Hitzler, P., Bianchi, F., Ebrahimi, M., and Sarker, M. Neural-symbolic integration and the Semantic Web. *Semantic Web* 11, 1 (2020), 3–11.
- Semantic Web Vol 12 No 2, Special Issue on Semantic Web for Cultural Heritage (2019)

