

# Linked Data and the Charm of Weak Semantics

## Introduction: The Strengths of Weak Semantics

by Thomas Baker and Stuart A. Sutton

### Linked Data and the Charm of Weak Semantics

#### EDITOR'S SUMMARY

Logic and precision are fundamental to ontologies underlying the semantic web and, by extension, to linked data. This special section focuses on the interaction of semantics, ontologies and linked data. The discussion presents the Simple Knowledge Organization Scheme (SKOS) as a less formal strategy for expressing concept hierarchies and associations and questions the value of deep domain ontologies in favor of simpler vocabularies that are more open to reuse, albeit risking illogical outcomes. RDF ontologies harbor another unexpected drawback. While structurally sound, they leave validation gaps permitting illogical uses, a problem being addressed by a W3C Working Group. Data models based on RDF graphs and properties may replace traditional library catalog models geared to predefined entities, with relationships between RDF classes providing the semantic connections. The BIBFRAME Initiative takes a different and streamlined approach to linking data, building rich networks of information resources rather than relying on a strict underlying structure and vocabulary. Taken together, the articles illustrate the trend toward a pragmatic approach to a Semantic Web, sacrificing some specificity for greater flexibility and partial interoperability.

#### KEYWORDS

linked data

semantic networks

RDF

ontologies

Thomas Baker, an organizer of the Dublin Core Metadata Initiative, is an associate professor at Sungkyunkwan University in Seoul, South Korea. He can be reached at [tb12<at>thbaker.org](mailto:tb12@thbaker.org).

Stuart A. Sutton, associate professor emeritus in the Information School of the University of Washington, is managing director of the Dublin Core Metadata Initiative. He can be reached at [sasutton<at>uw.edu](mailto:sasutton@uw.edu).

When the meme first emerged in the late 1990s, *Semantic Web* stood for logical data processing on the foundation of World Wide Web technology. One of its roots reached back to the 1955 meme of *artificial intelligence*, with its notion “that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.” [1, p. 12] The Semantic Web specifications developed by the World Wide Web Consortium from the late 1990s through the mid-2000s – the Resource Description Framework (RDF) and the Web Ontology Language (OWL) – were anchored in the notion of *ontology* as a “formal, explicit specification of a shared conceptualization” as supported by the field of ontology engineering.

Around 2006, Semantic Web was joined by the related, but more accessible and ultimately more popular meme of *linked data*. Starting with a cluster of databases linked to and from Wikipedia, the linked data movement took a more inclusive view of data technologies, with data serialized for Semantic Web-based interoperability as the five-star summit that providers of data in proprietary or application-specific document, database and record formats could by incremental steps ascend.

The contributions to this issue of the *Bulletin of the Association for Information Science and Technology* address, from five perspectives, how the shift to the idea of linked data at scale has changed the role of semantically precise ontologies.

As Oscar Corcho, María Poveda-Villalón and Asunción Gómez-Pérez see it, linked data has put the field of ontology engineering into a new context. Where tradition has favored heavyweight ontologies that demonstrate deep understanding of a domain and enable sophisticated inferences, the

linked data environment now favors vocabularies that are specified more lightly to maximize reusability and interoperability. With the rapid growth in dataset production, however, have come practices that are ontologically dubious. The casual recombination of terms from multiple and sometimes-disconnected sources can result in conceptually flawed, *Frankenstein ontologies*. Ontologies may also be designed as if they were schemas for checking data for conformance to application-specific constraints. Enabling providers to create linked data that is ontologically sound constitutes a key challenge to the field.

**Eric Prud'hommeaux** and **Jose Emilio Labra Gayo** note that the foundational RDF and OWL specifications published between 1999 and 2004 lacked the sort of constructs expected by everyday programmers more concerned with producing well-structured data than with logical inference. Popular RDF ontologies often describe entities without constraints tied to particular uses; the very constraints that make data so useful for a specific application can hinder reuse of that data by applications following different constraints. The RDF query language published in 2008, SPARQL, enormously improved search across RDF datasets but without providing a language for describing those datasets in validatable terms. Scattered moves in this direction, such as a Resource Shape specification from OSLC and a Description Set Profile constraint language from the Dublin Core community, were united in a 2013 workshop on RDF validation. The resulting W3C Data Shapes Working Group is currently working out how a language for data shapes, seen as constraint profiles, should relate to ontologically formal classes.

The BIBFRAME Initiative of the library world, as described by **Eric Miller** and **Uche Ogbuji**, shifts the focus away from precisely defined vocabularies and formal models towards more pragmatically defined profiles of the Dublin Core variety. Rich webs of links between resources contribute more to improving the quality of search results than highly specified data structures or descriptions based on formally perfect vocabularies. Rather than perfecting data models, they argue, the energies of subject experts and other practitioners are better spent adapting a starting

vocabulary to specific needs in local profiles and allowing such profiles to evolve iteratively through experimentation. In the linked data environment, partial interoperability is not only viable, but, given the massively diverse reality of the web, it is the only practical option.

**Karen Coyle** sees the influence of pre-Semantic Web technology in the development of library catalog models that distinguish generic works from specific editions and items. Technologies from physical card files through relational databases traditionally pushed programmers to define fixed, non-redundant data structures for efficiency of processing, and FRBR, a multilevel catalog model formulated in the 1990s, reflected such principles of design. Linked data, she argues, now renders such design constraints obsolete.

Linked data descriptions are just sets of statements, bundled as graphs. Overlap between graphs is not just the best we can expect; it is in fact good enough. Graphs can be decoupled from strongly defined classes in the sense of RDFS and OWL subsumption hierarchies and become recombinant parts of larger, ever-evolving knowledge graphs.

Looking at ontologies from the perspective of library, archive and museum (LAM) practice, **Antoine Isaac** and **Thomas Baker** note the functionally different role of property-and-class vocabularies, which specify relationships between the types of things described in datasets, in contrast to knowledge organization schemes (KOS), which typically define controlled vocabularies of values. Early efforts to translate pre-Semantic Web KOS into formal class hierarchies required extensive ontological debugging, while making it harder to re-use the resulting semantically complex data. The Simple Knowledge Organization Scheme (SKOS), published in 2009, provided a way to express hierarchical and association relationships without over-formalizing. Large-scale cultural heritage projects today commonly accept and mix data related to entities, such as persons, defined using either or both SKOS concepts and formal classes.

Linked data practice, in short, values pragmatic links alongside formal ontologies, prefers vocabularies specified with lightweight semantics for maximum reusability, defines overlapping profiles in place of monolithic data structures, sees data in terms of graphs and concepts more than formal

BAKER and SUTTON, continued

classes, shuns over-formalized semantics, embraces flexible and iterative evolution over static standardization and accepts partial interoperability as the only realistically attainable goal in today's massively diverse web. The linked data movement has invented useful new roles for constructs and languages that are, by design, semantically weak. ■

## Resources Mentioned in the Article

- [1] McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, 27(4). Retrieved from [www.aaai.org/ojs/index.php/aimagazine/article/view/1904](http://www.aaai.org/ojs/index.php/aimagazine/article/view/1904)