As the core of the Semantic Web matures, we see parallel trends such as microformats, RDFa, and Linked Data evolve with it, all of which complement each other in what we may well call a machine-readable Web. Yet, scalable techniques for dealing with this Web of Data in its entirety, i.e. using the Web as a database, as the Semantic Web has been envisioned, still misses some important pieces of the puzzle. Scalability here does not mean solely the ability to handle amounts of data at Web scale in terms of actual data processing, but it also refers to the human scale in the form of user-friendly tools that open the Web of Data to the current Web user.

The present special issue of the Journal of Web Semantics brings together the latest research on several aspects of the Web of Data, from the efforts to surface data from the text Web, to algorithms for resolving keyword queries to structured queries, to novel data modeling techniques.

“DBpedia - A Crystallization Point for the Web of Data” (Bizer, Lehmann, Kobilarov, Auer, Becker, Cyganiak, and Hellmann), details the data extraction process that populates DBpedia as well as the current status of interlinking DBpedia with other data sources on the Web. It also gives an overview of applications that facilitate the Web of Data around DBpedia, which now consists of more than 2.6 million entities and links to 4.7 billion pieces of information, covering a variety of domains such as geographic information, people, companies, films, music, genes, drugs, books, and scientific publications.

Accessing this amount of structured data for useful information is not a trivial task for most users. “From Keywords to Semantic Queries - Incremental Query Construction on the Semantic Web” (Zenz, Zhou, Minack, Siberski, Nejdl) presents an algorithm that can translate keyword queries–which most web users are familiar with–into structured queries that take advantage of semantics within the data of a particular domain.

Tackling the same timely topic of semantic search, “A Scalable IR Approach to Search the Web of Data” (Wang, Liu, Penin, Fu, Zhang, Tran, Yu, Pan) also translates keyword queries into structured queries but using an IR-based approach for handling uncertainties. Their system, Semplore, allows the user to build up each resolved query further through faceted browsing features.

“Hermes: DataWeb Search on a Pay-As-You-Go Integration Infrastructure” (Tran, Wang, Haase) is a third paper focusing on semantic search. Hermes’ novelty lies in its ability to integrate several data sources and perform automatic schema mapping in order to resolve queries that span over multiple sources.

Dealing with data from multiple sources not only presents data integration challenges but also brings forth other issues such as authority and credibility. “Querying for Provenance, Trust, Uncertainty and other Meta Knowledge in RDF.” (Dividino, Sizov, Staab, Schueller) proposes an extension to SPARQL query processing in such a way that given a SPARQL query for data, one may request meta knowledge, such as authority, certainty, etc., without modifying the query proper. It achieves highly flexible and automatically coordinated querying for data and meta
knowledge, while completely separating the two areas of concern.

The task of modeling data itself must also be simplified so that more people can do it, and can produce high quality data to populate the Web of Data. Recognizing complexities in the original RDFS language design, “Simple and Efficient RDFS” (Muñoz, Pérez, Gutierrez) introduces a small fragment which, while preserving the normative semantics and the core functionalities of RDFS, avoids the complexities of the original specification, and captures the main semantic functionalities of RDFS. The paper also introduces a minimalist deduction system over this fragment, which by avoiding certain rare cases, obtains a simple deductive system and a computationally efficient entailment checking.

Together, this selection of papers illustrates the latest efforts toward making the Web of Data accessible to the typical web user by making Web data easier to model, easier to integrate from several sources, and easier to query using familiar interface mechanisms such as faceted browsing and keyword search.