The Role of Ontology Engineering in Linked Data Publishing and Management: An Empirical Study

Markus Luczak-Rösch, University of Southampton, Southampton, UK

Elena Simperl, University of Southampton, Southampton, UK

Steffen Stadtmüller, Karlsruhe Institute of Technology, Karlsruhe, Germany

Tobias Käfer, Karlsruhe Institute of Technology, Karlsruhe, Germany

ABSTRACT

In this article the authors evaluate the adoption and applicability of established ontology engineering results by the Linked Data providers' community. The evaluation relies on a combination of qualitative and quantitative methods; in particular, the authors conducted an analytical survey containing structured interviews with data publishers in order to give an account of the current ontology engineering practice in Linked Data provisioning, and compared and expanded our findings with statistics on ontology development and usage provided by the Billion Triple Challenges datasets from 2012 (using the vocab.cc platform) and from 2014 and other related tools. The findings of the evaluation allow data practitioners and ontologists to yield a better understanding of the conceptual part of the LOD Cloud; and form the basis for the definition of purposeful, empirically grounded guidelines and best practices for developing, managing and using ontologies in the new application scenarios that arise in the context of Linked Data.

Keywords: *Analysis, Linked Data, Ontology, Ontology Engineering, Statistics, Survey*

INTRODUCTION

In the last years we have seen a continuous uptake of semantic technologies - most recently on the open Web, driven by the Linked Data movement, but in equal measure also in enterprise environments. The key distinct feature of semantic computing compared to other information management technologies is their use of Web standards - languages for knowledge representation, as well as protocols for exposing, accessing and exchanging this knowledge - to structure and formalize information in a way that enables computers to 'understand' com-

DOI: 10.4018/IJSWIS.2014070103

plex concepts and situations in a similar way as humans do. Ontologies, defined as reusable models capturing the knowledge in a given domain, are one of the core building blocks of the semantic-technologies stack. In combination with components for semantic data management, reasoning, search, as well as annotation and description of digital artifacts, they can facilitate the development of sophisticated - and economically feasible - solutions to many prevailing problems in today's information management.

Ontologies: Their Use in Information Systems and Linked Data Provisioning

Probably one of the best showcases for the added value of ontologies comes from the life sciences domain, which is often quoted as an important early adopters of semantic technologies. This becomes evident when inspecting the variety of ontologies and ontology-engineering environments built for and deployed in biology, genetics, biochemistry and medicine in order to describe domain-specific artifacts in a commonly accepted, uniform way, and to enable advanced search and knowledge organization capabilities for large (scientific) archives and data bases. To give an example, the BioPortal platform offers access to over 200 life sciences ontologies, created and continuously maintained collaboratively by dedicated editors and user organizations[1]. Some of the most sophisticated application examples demonstrating how ontologies could be leveraged as a knowledge management technology in life sciences emerged in the area of scientific publishing, most notably at Elsevier, for instance as part of their Grand Challenge, which is organized under the motto "Knowledge Enhancement in Life Sciences" (de Waard, 2010). In other areas such as public administration and media and publishing the prospects for using ontologies and other complementary semantic technologies such as Linked Data look equally promising, both in terms of the types of functionality, applications and services these technologies enable, and the maturity and economical feasibility of their deployment and maintenance.

The raise of Linked Data has marked the beginnings of a new era in the history of the Semantic Web, leading to a notable shift in the perception and understanding of the overall vision and the steps that are required in order to achieve it. Linked Data proposes a set of guidelines, supported by knowledge representation formats and technology, for publishing and accessing structured data on the Web. In its overly successful instantiation as the so-called "Linked Open Data Cloud" (LOD Cloud)[2], it introduced new application scenarios for the development, management and deployment of ontologies. These scenarios are concerned with the exposure and integration of data sets adhering to the four Linked Data principles, using the recommended life cycle and technology stack (Auer et al., 2012; Auer et al., 2013), and following emerging best practices from various vertical domains.

The creation, management and use of ontologies has been traditionally pursued by the ontology engineering community. The results achieved in the past decades of research and development in this discipline are acknowledged to be of incontestable value for the large-scale uptake of semantic technologies that we witness today. Nevertheless, at least prior to the advent of Linked Data, their range of application in real-world projects was comparatively limited, despite of a growing number of ontologies online available and gradual improvements of the accompanying technology. This state of affairs has been subject to several empirical surveys and case studies in the past (Cardoso, 2007; Hepp, 2007; Paslaru-Bontas & Tempich, 2006; Simperl, Mochol, & Bürger 2010). Today, the overall picture looks much more promising - as illustrated by applications as those briefly discussed earlier in this section that naturally capitalize on the groundwork undertaken in the ontology engineering community, and on the success of the Linked Data movement. Confronted with these bright prospects, one question that arises is about the directions the ontology engineering community needs to