Data semantics, explanation and interoperability

Giancarlo Guizzardi

Semantics, Cybersecurity & Services (SCS), University of Twente, Enschede, The Netherlands

Abstract

It is well-known by now that, of the so-called 4Vs of Big Data (Velocity, Volume, Variety and Veracity), the bulk of effort and challenge is in the latter two: (1) data comes in a large variety of representations (both from a syntactic and semantic point of view); (2) data can only be useful if truthful to the part of reality that it is supposed to represent. Moreover, the most relevant questions we need to have answered in science, government and organizations can only be answered if we put together data that reside in different data silos, which are produced in a concurrent manner by different agents and in different points of time and space. Thus, data is only useful in practice if it can (semantically) interoperate with other data. Every data schema represents a certain conceptualization, i.e., it makes an ontological commitment to a certain worldview. Issue (2) is about understanding the relation between data schemas and their underlying conceptualizations. Issue (1) is about safely connecting these different conceptualisations represented in different schemas. To address (1) and (2), we need to be able to properly explain these data schemas, i.e., to reveal the real-world semantics (or the ontological commitments) behind them. In this masterclass, I discuss the strong relation between the notions of real-world semantics, ontology, and explanation. I will present a notion of explanation termed Ontological Unpacking, which aims at explaining symbolic representation artifacts (conceptual models connected to data schemas, knowledge graphs, logical specifications). I show that these artifacts when produced by Ontological Unpacking differ from their traditional counterparts not only in their expressivity but also on their nature: while the latter typically merely have a descriptive nature, the former have an explanatory one. Moreover, I show that it is exactly this explanatory nature that is required for semantic interoperability. I will also discuss the relation between Ontological Unpacking and other forms of explanation in philosophy and science, as well as in Artificial Intelligence. I will argue that the current trend in XAI (Explainable AI) in which "to explain is to produce a symbolic artifact" (e.g., a decision tree or a counterfactual description) is an incomplete project resting on a false assumption, that these artifacts are not "inherently interpretable", and that they should be taken as the beginning of the road to explanation, not the end. Finally, I will illustrate our approach with the unpacking of a prominent Viral Conceptual Model (VCM). This masterclass is based on the paper "Guizzardi, G., & Guarino, N. (2024). Explanation, semantics, and ontology. Data & Knowledge Engineering, 153, 102325."