Is Your Data 6-Star?*

Cassia Trojahn

Institut de Recherche en Informatique de Toulouse, France
firstname.lastname@irit.fr

Abstract. Linked open data in general is poor in foundational distinctions and by consequence lacks semantics clarity. Such distinctions are however essential for many applications consuming the data, and have been since many years the subject of study in foundational ontologies. This paper argues that foundational distinctions have to be better taken into account in the process of construction, alignment and publication of linked open data, from a methodological perspective. It proposes to extend the well-known 5-star rating schema to a 6-star schema, with data getting a 6-star rating when equipped with foundational distinctions.

1 Introduction

The Berners-Lee's vision of a semantic web has been materialized with the *Linked Open Data* (LOD) initiative, where structured data are (ideally) exposed as instances of ontologies and linked across knowledge bases. In that perspective, the well-known 5-star incremental framework has guided the process of publishing data on the Web[1]. From there, the LOD cloud has become an extremely rich source of knowledge in several domains (such as Geography, Linguistics, Life Sciences, Social Networking) with about 1260 datasets with 16187 links[2].

While most (core) linked open datasets have been constructed from existing (encyclopedic) sources (such as DBpedia and BabelNet), the process of construction, alignment and publication of data (and the ontologies describing them) highly neglects the role of foundational ontologies. In consequence, linked open data in general is poor in foundational distinctions. Previous works have addressed, in particular, the challenge of matching domain and foundational ontologies [12,13]. More recently, the lack of foundational distinctions in the LOD has been highlighted in [2]: *distinctions such as whether an entity is inherently a class or an individual, or whether it is a physical object or not, are hardly expressed in the data, although they have been largely studied and formalised by foundational ontologies.* Such distinctions are essential in many applications consuming linked open data (and essential in Artificial Intelligence in general). They are however at the center of the formal and applied ontology field focusing

[1] https://www.w3.org/DesignIssues/LinkedData.html

[2] https://lod-cloud.net/ (August 2020)

on a large spectrum of foundational issues (types of entities, formal relations, space, time, etc.). Complementary to [2], the authors in [3] state that in the semantic web, there is an increasingly need for serious engagement with ontologies, understood as a general theory of the types of entities and relations making up their respective domains of inquiry. However, there is still little interaction between the communities, despite the fact that they share common ambitions in terms of knowledge understanding.

This paper argues that foundational distinctions have to be better taken into account in the process of construction, alignment and publication of LOD data from a methodological point of view. In that perspective, we propose to extend the 5-star data schema to a 6-star schema, with data getting a 6-star when equipped with foundational distinctions (as for instance, to be linked to appropriated foundational ontologies).

2 Which kinds of foundational distinctions?

Foundational distinctions are at the core of foundational (top-level or upper) ontologies. A foundational ontology is a high-level and domain independent ontology whose concepts (e.g., object, event, quality, disposition) and relations (e.g., parthood, participation, dependence, causality) are intended to be basic and universal to ensure generality and expressiveness for a wide range of domains. It is often characterized as representing commonsense concepts and is limited to concepts which are meta, generic, and philosophical. Diverse foundational ontologies have been developed so far (BFO, DOLCE, GFO, SUMO, UFO, PROTON, to cite a few), influenced by different philosophies and views on the reality [9,10]. One of the well-known foundational ontologies is DOLCE [5] (*Descriptive Ontology for Linguistic and Cognitive Engineering*), an ontology of *particulars* which adopts a *descriptive approach* with a clear cognitive bias, as it aims at capturing the ontological categories underlying natural language and human commonsense. DOLCE is based on a fundamental distinction between *endurant* (objects or substances) and *perdurant* entities (events or processes). The main relation between *endurants* and *perdurants* is that of participation. Under another perspective, GFO [6] (*General Formal Ontology*) considers distinctions between concrete individuals which exist in time or space whereas abstract individuals do not. While an *endurant* is an individual that exists in time, but cannot be described as having temporal parts or phases; a *process*, on the other hand, is extended in time. Complementary, BFO [1] (*Basic Formal Ontology*) represents the reality into two disjoint categories of *continuant* (independent and dependent continuants, attributes, and locations) and *occurrent* (processes and temporal regions). A comparison of foundation ontologies can be found in [10].

3 What has been done so far?

From a methodological point of view, LOD highly neglects the role of foundational ontologies. There are two approaches for the use of foundational ontologies

[14]. With a *top-down approach*, foundational ontologies are used as a reference for deriving domain concepts, taking advantage of the knowledge and experience already encoded in it. In a *bottom-up approach*, one usually matches an existing domain ontology to the foundational ontology. As reported in [8,9], methodologies for constructing ontologies should not neglect the use of foundational ontologies and should better address it in a *top-down* approach. In the absence of systematic adoption of foundational ontologies within the domain ontology development process (in general), *bottom-up* approaches have to be applied instead. In this task, matching foundational and domain ontologies plays a key role. Most state-of-the-art matching systems however fail in the task [13], with few dedicated approaches been developed so far [7,12].

With respect to aligning and equipping LOD datasets with foundational ontologies, in [7], the approach has been used to align the PROTON foundational ontology to LOD datasets. It uses Wikipedia to construct a set of category hierarchy trees and then determines which classes to align using different similarities. One proposal analysing the foundational coverage of DBPedia is the one by [11], where correspondences between DBpedia ontology and DOLCE-Zero [4], a module of DOLCE, are used to identify inconsistent statements in DBpedia. The authors focus on finding systematic errors or anti-patterns in DBpedia. They argued that by aligning these ontologies and by combining reasoning and clustering of the reasoning results, errors affecting statements can be identified with minimal human workload. More recently, in [2], automatic classification of foundational distinctions (class vs. instance or physical vs. non-physical objects) of LOD entities is done with two strategies: an (unsupervised) alignment approach and a (supervised) machine learning approach. The alignment approach, in particular, relies on the structure of alignments between DBpedia, DOLCE, and external lexical linked data. They use the paths of alignments and taxonomical relations in these resources and automated inferences to classifying whether a DBpedia entity is a physical object or not.

4 The 6-star data rating schema

We propose a 6-star rating schema for data that expresses foundational ontological distinctions. As briefly introduced in Section 2, such distinctions include clear semantics on, for instance, concrete and abstract individuals, events and processes, time and temporal regions. They take into account as well formal relations, such as parthood, dependence, constitution, causality, instantiation. The proposed rating schema, revising the well-known 5-star schema, is as in the following:

⋆	Available on the web (whatever format) but with an open licence
⋆⋆	Available as machine-readable structured data
⋆⋆⋆	Available with non-proprietary format
⋆⋆⋆⋆	All the above plus using open standards from W3C
⋆⋆⋆⋆⋆	All the above, plus data linked to other data to provide context
⋆⋆⋆⋆⋆⋆	All the above, plus data equipped with foundational distinctions

5 Final remarks

Foundational distinctions guarantee data consistency and improves semantics clarity. Linked open data equipped with such distinctions should be rewarded with a 6-star. In the future, we plan to extend our previous work [12] in order to take into account data (instance) matching. Complementary to what has been proposed in [2], such approach could be then applied for helping improving existing datasets with foundational distinctions.

References

1. R. Arp, B. Smith, and A. Spear. *Building Ontologies with Basic Formal Ontology*. MIT Press, 2015.
2. L. Asprino, V. Basile, P. Ciancarini, and V. Presutti. Empirical analysis of foundational distinctions in linked open data. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 3962–3969, 2018.
3. M. Bennett and K. Baclawski. The role of ontologies in linked data, big data and semantic web applications. *Applied Ontology*, 12(3-4):189–194, 2017.
4. A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. Sweetening WORDNET with DOLCE. *AI Magazine*, 24(3):13–24, 2003.
5. A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening Ontologies with DOLCE. In *13th Conference on Knowledge Engineering and Knowledge Management*, pages 166–181, 2002.
6. H. Herre, B. Heller, P. Burek, R. Hoehndorf, F. Loebe, and H. Michalek. General Formal Ontology (GFO): A Foundational Ontology Integrating Objects and Processes. In *Basic Principles, Research Group Ontologies in Medicine*, 2007.
7. P. Jain, P. Yeh, K. Verma, R. Vasquez, M. Damova, P. Hitzler, and A. Sheth. Contextual Ontology Alignment of LOD with an Upper Ontology: A Case Study with PROTON. In *Proceedings of the 8th Extended Semantic Web Conference*, pages 80–92, 2011.
8. C. Keet. The use of foundational ontologies in ontology development: An empirical assessment. In *Proceedings of the 8th Extended Semantic Web Conference*, pages 321–335, 2011.
9. Z. Khan and C. Keet. ONSET: Automated Foundational Ontology Selection and Explanation. In *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management*, pages 237–251, 2012.
10. V. Mascardi, V. Cordì, and P. Rosso. A Comparison of Upper Ontologies. In *8th AI*IA/TABOO Workshop on Agents and Industry*, pages 55–64, 2007.
11. H. Paulheim and A. Gangemi. Serving dbpedia with dolce – more than just adding a cherry on top. In *ISWC 2015*, pages 180–196, 2015.
12. D. Schmidt, R. Basso, C. Trojahn, and R. Vieira. Matching domain and top-level ontologies exploring word sense disambiguation and word embedding. In *Emerging Topics in Semantic Tech.*, pages 27–38, 2018.
13. D. Schmidt, C. Trojahn, and R. Vieira. Analysing Top-level and Domain Ontology Alignments from Matching Systems. In *Proc. of the Workshop on Ontology Matching*, pages 1–12, 2016.
14. S. Semy, M. Pulvermacher, and L. Obrst. Toward the use of an upper ontology for U.S. government and U.S. military domains: An evaluation. Technical report, MTR 04B0000063, The MITRE Corporation, 2004.