

# DESIDERATA FOR KNOWLEDGE GRAPHS

Michael K. Bergman<sup>1</sup>, Coralville, Iowa USA

February 21, 2018

*AI3::Adaptive Information* blog

I think the market has spoken in preferring the term of ‘knowledge graph’ over that for ‘ontology.’ I suppose we could argue nuances in differences for what the terms mean. We will continue to use both terms, more-or-less interchangeably. But, personally, I do find the concept of ‘knowledge graph’ easier to convey to clients.

As we see knowledge graphs proliferate in many settings — from virtual agents ([Siri](#), [Alexa](#), [Cortana](#) and [Google Assistant](#), among [others](#)) to [search](#) and AI platforms ([Watson](#)) — I’d like to take stock of the state-of-the-art and make some recommendations for what I would like to see in the next generation of knowledge graphs. We are just at the beginning of tapping the potential of knowledge graphs, as my recommendations show.

Going back for twenty years to Nicola Guarino in 1998 [\[1\]](#) and Michael Uschold in 2008 [\[2\]](#), there is a sense that ontologies could be relied upon for even more central aspects of overall applications. Both Guarino and Uschold termed this potential ‘ontology-driven information systems.’ It is informative of the role that ontologies may play by listing some of these incipient potentials, some of which have only been contemplated or met in one or two actual installations. Let me list nine main areas of (largely) untapped potential:

1. *Context and meaning* — by this, I mean the ability to model contexts and situations, which requires specific concepts for such and an ability to express gradations of adjacency (spatial and otherwise). Determining or setting contexts is essential to disambiguate meaning. Context and situations have been particularly difficult ideas for ontologies to model, especially those that have a binary or dichotomous design;
2. *A relations component* — true, OWL offers the distinction of annotation, object and datatype properties, and we can express property characteristics such as transitivity, domain, range, cardinality, inversion, reflexivity, disjunction and the like, but it is a rare ontology that uses any or many of these constructs. The `subProperty` expression is used, but only in limited instances and rarely in a systematic schema. For example, it is readily obvious that some broader predicates such as `animalAction` could be split into `involuntaryAction` and `voluntaryAction`, and then into specific actions such as `breathing` or `walking`, and so on, but schema with these kinds of logical property subsumptions are not evident. Structurally, we can use OWL to reason over actions and relations in a similar means as we reason over entities and types, but our common ontologies have yet to do so. Creating such schema are within grasp since we have language structures such as [VerbNet](#) and other resources we could put to the task;
3. *An attributes component* — the lack of a schema and organized presentation of attributes means it is a challenge to do [ABox](#)-level integration and interoperability. As with a *relations* component, this gap is largely due to the primary focus on concepts and entities in the early stages of semantic technologies. Optimally, what we would like to see is a well-organized attributes schema that enables instance data

---

<sup>1</sup>Email: [mike@mkbergman.com](mailto:mike@mkbergman.com)

characteristics from different sources to be mapped to a canonical attributes schema. Once in place, not only would mapping be aided, but we should also be able to reason over attributes and use them as intensional cues for classifying instances. At one time Google touted its Biperpedia initiative [3] to organize attributes, but that effort went totally silent a couple of years ago;

4. A *quantity units ontology* — is the next step beyond attributes, as we attempt to bring data values for quantities (and well as the units and labeling used) into alignment. Fortunately, of late, the [QUDT ontologies](#) (quantities, units and data types) has become an active project again with many external supporters. Something like this needs to accompany the other recommendations listed;
5. A *statistics and probabilities ontology* — the world is not black-and-white, but vibrantly colored with all kinds of shades. We need to be able to handle gradations as well as binary choices. Being able to add probabilistic reasoners is appropriate given the idea of continua (Thirdness) from [Charles Sanders Peirce](#) and capturing the idea of [fallibility](#). Probabilistic reasoning is still a young field in ontology. Some early possibilities include Costa [4] and the [PR-OWL](#) ontology using Multi-Entity Bayesian Networks (MEBN) [5] which are a probabilistic first-order logic that goes beyond Peirce's classic deterministic logic; as well as fuzzy logic applied to ontologies [6];
6. *Abductive reasoning and hypothesis generation* — Peirce explicated a third kind of logical reasoning, [abduction](#), that combines hypothesis generation with an evaluation of likelihood of success and effort required. This logic method has yet to be implemented in any standard Web ontologies to my knowledge. The method could be very useful to pose desired outcome cases and then to work through what may be required to get there. Adding this to existing knowledge graphs would likely require developing a bespoke abductive reasoner;
7. *Rich feature set for KBAI* — we want a rich features set useful to provide labeled instances for supervised machine learners. I addressed this need earlier with a rather comprehensive listing of possible features for knowledge graphs useful to learners [7]. We now need to start evaluating this features pool to provide pragmatic guidance for which features and learners match best for various knowledge-based artificial intelligence ([KBAI](#)) tasks;
8. *Consistent, clean, correct and coherent* — we want knowledge graphs that are as free from error as possible to make sure we are not feeding garbage to our machine learners and as a coherent basis for evaluating new additions and mappings; and
9. *ODapps* — ‘ontology-driven applications’ go beyond the mere templating or completions of user interface components to devise generic software packages driven by ontology specifications for specific applications. We have developed and deployed ODapps to import or export datasets; create, update, delete ([CRUD](#)) or otherwise manage data records; search records with full-text and faceted search; manage access control at the interacting levels of users, datasets, tools, and CRUD rights; browse or view existing records or record sets, based on simple to possible complex selection or filtering criteria; or process results sets through workflows of various natures, involving specialized analysis, information extraction or other functions. ODapps are designed more similarly to widgets or API-based frameworks than to the dedicated software of the past, though the dedicated functionality is quite similar. The major change in ODapps is to use a relatively common abstraction layer that responds to the structure and conventions of the guiding ontologies. We may embed these ODapps in a layout canvas for a Web page, where, as the user interacts with the system, the service generates new queries (most often [SPARQL](#)) to the various Web services endpoints, which produce new structured results sets, which can drive new displays and visualizations. As new user interactions occur, the iteration cycle is generated anew, again starting a new cycle of queries and results sets.

Fortunately, we are actively addressing multiple of these recommendations (#1 – #3, #6 – #9) with our [KBpedia](#) initiative. We are also planning to add mapping to QUDT (#4) in a near-future release. We are presently evaluating probabilistic reasoners and hypothesis generators (#5 and #6).

Realizing these potentials will enable our knowledge management (KM) efforts to shift to the description, nature, and relationships of the information environment. In other words, ontologies themselves need to become the focus of development. KM no longer needs to be abstracted to the IT department or third-party software. The actual concepts, terminology and relations that comprise coherent ontologies now become the explicit focus of KM activities, and subject to the direct control and refinement by their users, the knowledge workers, and subject matter experts.

We are still some months from satisfying our desiderata for knowledge graphs. Fortunately, we have already made good progress, and we are close at hand to check off all of the boxes.

## **Acknowledgements**

This article was originally posted on the *AI3::Adaptive Information* Web site at <http://www.mkbergman.com/2135/desiderata-for-knowledge-graphs/>. This version has been edited and reformatted slightly for PDF distribution. We thank Cognonto Corporation for making this content freely available.

- 
- [1] N. Guarino, “Formal Ontology and Information Systems,” in *Proceedings of FOIS’98*, Trento, Italy, 1998, pp. 3–15.
  - [2] M. Uschold, “Ontology-Driven Information Systems: Past, Present and Future,” in *Proceedings of the Fifth International Conference on Formal Ontology in Information Systems (FOIS 2008)*, Carola Eschenbach and Michael Grüninger, eds., IOS Press, Amsterdam, Netherlands, 2008, pp. 3–20.
  - [3] R. Gupta, A. Halevy, X. Wang, S.E. Whang, and F. Wu. “Biperpedia: An Ontology for Search Applications,” *Proceedings of the VLDB Endowment* 7, no. 7, 2014, pp. 505-516.
  - [4] P. C. Costa, “Bayesian Semantics for the Semantic Web,” Ph.D., George Mason University, 2005.
  - [5] K. B. Laskey, “MEBN: A Language for First-Order Bayesian Knowledge Bases,” *Artificial Intelligence*, vol. 172, no. 2–3, pp. 140–178, Feb. 2008.
  - [6] F. Bobillo and U. Straccia, “Fuzzy Ontology Representation Using OWL 2,” *International Journal of Approximate Reasoning*, vol. 52, no. 7, pp. 1073–1094, Oct. 2011.
  - [7] M.K. Bergman, “A (Partial) Taxonomy of Machine Learning Features,” *AI3::Adaptive Information* blog, November 23, 2015.