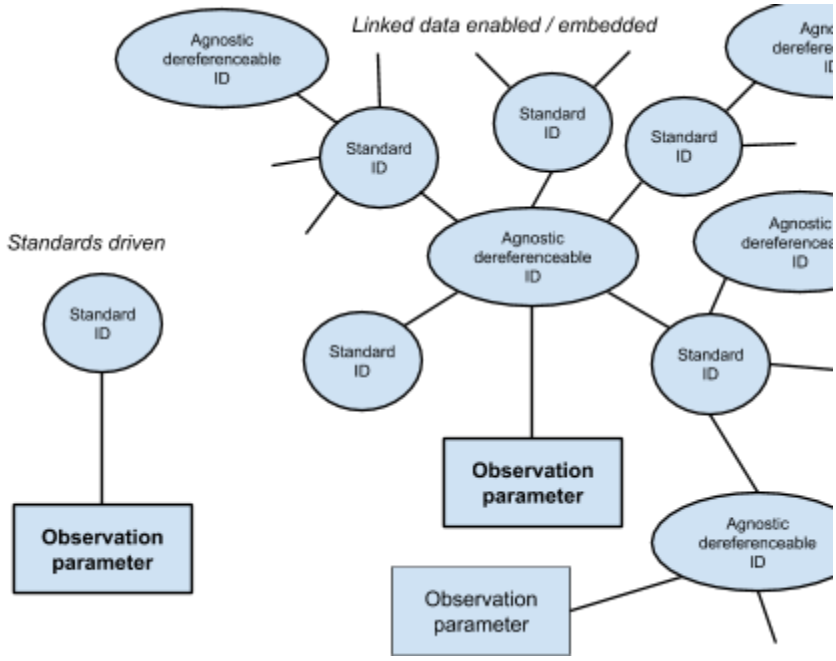
Jonas Almeida, April 21, 2020

The argument that "a little semantics goes a long way"[1] is often misrepresented as being against data standards, instead of being about assessing the value of a semantic annotation by the improvements in interoperability. The misrepresentation is puzzling to many in Data Science, because of the paradox that it comes from James Hendler, one of the authors of the original "semantic Web" paper, who is also the editor of the reference textbook in the field[2]. Until recently, this confusion reflected an academic argument between Knowledge Engineering (declarative) and the Science of the Web[3] (executable). However, with a growing trend towards using AI to annotate data elements, we stand at the shores of a new world where annotations are made and acted upon by AI engines and researchers alike, configured to identify multiparametric signals defining association manifolds. Clarification is, therefore, in order.

Own your own relationships

When a concept is being instantiated by observations it is critical that it is assigned a dereferenceable unique identifier that is semantically neutral. This allows the owner of the concept ID to document the original context in the observation, and do so without compromising subsequent enrichment/evolution by anyone, including by AI. This is the AAA principle - "anyone can say anything about anything" driving semantics as linked data. This concept is discussed in the three references listed above and is illustrated below. It is also the principle driving modern data exchange catalogs backed by search engines, as documented by schema.org.



It is important to note that this is very mature technology, and is at the root of "run-away" standards (AAA user-driven) such as Gene Ontology (Genomics), Loinc (Pathology), and FHIR (Health Care). It is also critical to note that linked data expands, scales and loosens the standards-driven approach. It is an update, not a replacement. It is also much easier to use, to serialize and track: json-ld.org. So how does one start? Simply by creating a JSON (LD) document in a (versioned) github repository with an agnostic (random, incremental, etc) dereferenceable, unique ID. Something as simple as this.

---

[1] https://www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html
[2] https://www.amazon.com/Semantic-Web-Working-Ontologist-Effective-dp-0123859654
[3] https://science.sciencemag.org/content/313/5788/769.full