Ontologies as applied to FAIR data

(The I in FAIR)

Ontologies provide a common structure to bring disparate data together – for this post I will refer to the definition of Ontology from Tom Gruber below – emphasis added by me. Note the last highlighted statement as a critical bit with significant implications in the implementation of systems in support of scientific processes. Having led and survived many data and systems integration efforts over the years, one of the most challenging aspects is hidden in this last statement. Changing data format, naming, etc… at the source is often met with almost religious fervor as change had wide-ranging implications to linked analysis, and multiple stakeholders have disparate needs or views of the data in question. The idea of an abstraction layer to bring these data together is nothing new, and this approach is a natural evolution in my mind. We are recognizing as an industry that isolated data is useful in context, but far more powerful when shared. To attain that goal, we need a common vocabulary and structure – enter the domain ontologies we can map to.

> In the context of computer and information sciences, an **ontology defines a set of representational primitives with which to model a domain of knowledge or discourse**. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. In the context of database systems, **ontology can be viewed as a level of abstraction of data models, analogous to hierarchical and relational models, but intended for modeling knowledge about individuals, their attributes, and their relationships to other individuals**. Ontologies are typically specified in languages that allow abstraction away from data structures and implementation strategies; in practice, the languages of ontologies are closer in expressive power to first-order logic than languages used to model databases. For this reason, ontologies are said to be at the "semantic" level, whereas database schema are models of data at the "logical" or "physical" level. **Due to their independence from lower level data models, ontologies are used for integrating heterogeneous databases, enabling interoperability among disparate systems,** and specifying interfaces to independent, knowledge-based services.
>
> HTTPS://TOMGRUBER.ORG/WRITING/DEFINITION-OF-ONTOLOGY.PDF

The ontology provides a navigable structure to the data relationships that will be consistent across all sources in scope of reference. This is the critical bit to derive value from the data – moving it from isolated to interoperable, and supporting the rest of the FAIR principles. Access control is often a critical bit when joining / sharing data, especially anything that can be used to form a conclusion that may be subject to challenge or reinterpretation absent context. Ontology based access can be used to support these access controls given the proper structure. While outside the scope of this surface level post, you can read more from MIT Press Direct here on that topic.

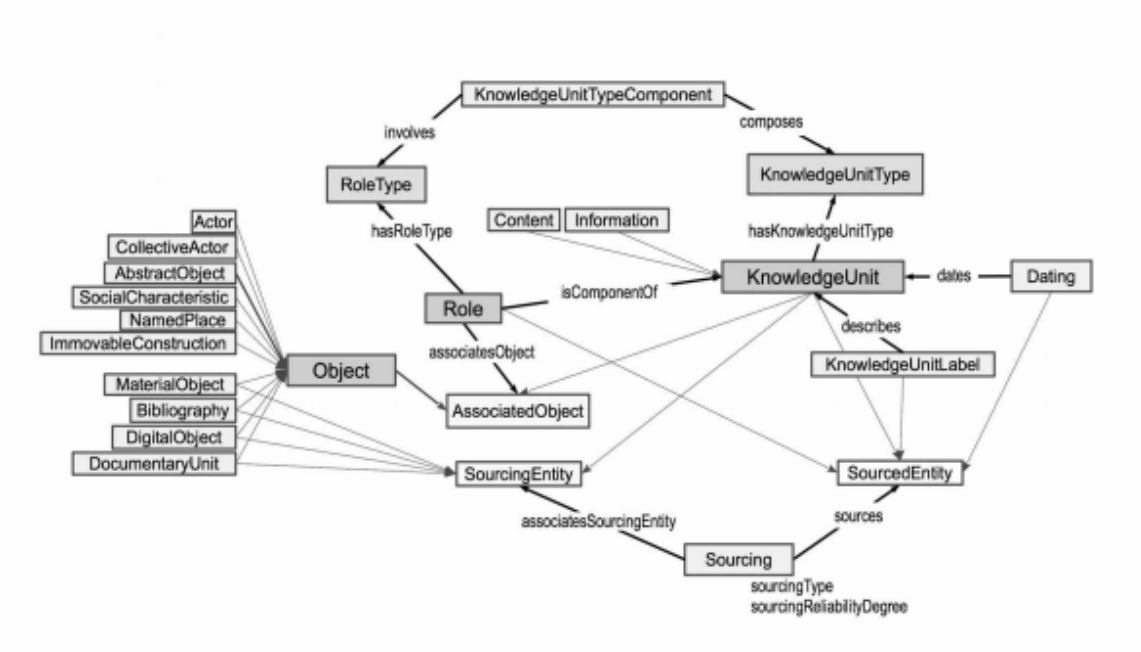

Fig. 1. Ontology of the symogih.org project -version 0.2.1

http://www.semantic-web-journal.net/system/files/swj2523.pdf

Mapping these ontologies and related data sets to a graph database and unlocking the power of the relationship hierarchy inferred through the ontology mapping, secured through the same, provides a rich foundation to build a query and interaction layer. There are challenges to be solved throughout this process, and this posts scratches the surface and provides some context / links, but it does help frame jumping off point for these ideas along with connections to papers and resources with the “rest of the story” as Paul Harvey would say.

References:

1. Tom Gruber (2008), Ontology. Entry in the *Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, 2009. https://tomgruber.org/writing/definition-of-ontology
2. Giancarlo Guizzardi; Ontology, Ontologies and the “I” of FAIR. Data Intelligence 2020; 2 (1-2): 181–191. doi: https://doi.org/10.1162/dint_a_00040
3. Poveda-Villalón, María & Espinoza-Arias, Paola & Garijo, Daniel & Corcho, Oscar. (2020). Coming to Terms with FAIR
   Ontologies. https://www.researchgate.net/publication/344042645_Coming_to_Terms_with_FAIR_Ontologies
   A. Direct link in case above link fails
4. Francesco Beretta, 06/30/2020. A challenge for historical research: making data FAIR using a collaborative ontology management environment
   (OntoME) http://www.semantic-web-journal.net/content/challenge-historical-research-making-data-fair-using-collaborative-ontology-management-0
   A. Direct link to paper in case above link fails
5. Christopher Brewster, Barry Nouwt, Stephan Raaijmakers, Jack Verhoosel; Ontology-based Access Control for FAIR Data. *Data Intelligence* 2020; 2 (1-2): 66–77. doi: https://doi.org/10.1162/dint_a_00029
6. Tim Berners-Lee, Date: 2006-07-27, last change: $Date: 2009/06/18 18:24:33 $, Status: personal view only. Editing status: imperfect but published. https://www.w3.org/DesignIssues/LinkedData.html