2018 Spring Technology Summit

Project ORREO Foundry: Open Residential Real Estate Ontology Foundry

IT’S GAME ON FOR DATA STANDARDS!
Tavi Truman | CTO

BIO: Celebrating more than 25 years in Software sciences, Information, Cognitive and Applied Ontology Science and Commercial software development. Currently, Tavi leads all aspects of the RocketUrBiz Engineering. His focus is on Ontological Engineering, Real Estate Informatics, Automated Reasoning and Workflow processing.

Email: tavi@rocketurbiz.com
Phone: +1 408-214-0916
Website: http://www.rocketurbiz.com
The word ‘ontology’ can refer to a branch of Western philosophy—having its origins in ancient Greece with philosophers such as Parmenides, Heraclitus, Plato, and Aristotle—the concern of which is the study of what is, of the kinds and structures of objects, properties, events, processes, and relations in every area of reality.

From this philosophical perspective, ontology seeks to provide a definitive and exhaustive classification of entities in all spheres or domains of being. As a theoretical discipline concerned with accurately describing the taxonomy of all things that exist, philosophical ontology is synonymous with classical metaphysics.

ontology = def. a representational artifact, comprising a taxonomy as proper part, whose representations are intended to designate some combination of universals, defined classes, and certain relations between them.

- Universal = type, kind of thing or Entity
What is Ontology (continued)

taxonomy = def. a hierarchy consisting of terms denoting types (or universals or class-es) linked by subtype relations

By “types” or “universals” we mean the entities in the world referred to by the nodes (appearing here as boxes) in a hierarchy; in the case of figure 1.1, biological phyla, classes, and orders.

entity = def. anything that exists, including objects, processes, and qualities “Entity” thus comprehends also representations, models, images, beliefs, utterances, documents, observations, and so on.

A building, or edifice, is a structure with a roof and walls standing more or less permanently in one place.
What must an Ontology contain

Visualizations in Graph Theoretic form

- Terms (represent types in reality)
  - Preferred labels
  - Synonyms
- Unique IDs
  - Alphanumeric identifiers for each term
  - Namespace ID
- Nodes - terms
  - Complex collections of things
  - Noun or noun phrases
- Edges
  - relations
- Definitions
- Axioms
  - Governs how the terms are to be understood
The Vision of the Semantic Web powered by Ontology

Tim Berners-Lee, inventor of the internet: “sees a more powerful Web emerging, one where documents and data will be annotated with special codes allowing computers to search and analyze the Web automatically. The codes … are designed to add meaning to the global network in ways that make sense to computers”
Tim Berners-Lee: hyperlinked vocabularies, called ‘ontologies’ will be used by Web authors “to explicitly define their words and concepts as they post their stuff online.

“The idea is the codes would let software ‘agents’ analyze the Web on our behalf, making smart inferences that go far beyond the simple linguistic analyses performed by today's search engines.”
Working together: W3C Standards for the Semantic Web

In addition to the classic "Web of documents," W3C is helping to build a technology stack to support a "Web of data." The sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term "Semantic Web" refers to W3C's vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.

Linked Data
The Semantic Web is a Web of data — of dates and titles and part numbers and chemical properties and any other data one might conceive of. RDF provides the foundation for publishing and linking your data. Various technologies allow you to embed data in documents (RDFa, GRDDL) or expose what you have in SQL databases, or make it available as RDF files.

Vocabularies
At times it may be important or valuable to organize data. Using OWL (to build vocabularies, or "ontologies") and SKOS (for designing knowledge organization systems) it is possible to enrich data with additional meaning, which allows more people (and more machines) to do more with the data.

Query
Query languages go hand in hand with databases. If the Semantic Web is viewed as a global database, then it is easy to understand why one would need a query language for that data. SPARQL is the query language for the Semantic Web.

Vertical Applications
W3C is working with different industries — for example in Health Care and Life Sciences, eGovernment, and Energy — to improve collaboration, research and development, and innovation adoption through Semantic Web technology. For instance, by aiding decision-making in clinical research, Semantic Web technologies will enable many forms of biological and medical information across institutions.
The OBO Foundry

The Open Biological and Biomedical Ontology (OBO) Foundry is a collective of ontology developers that are committed to collaboration and adherence to shared principles. The mission of the OBO Foundry is to develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate. To achieve this, OBO Foundry participants voluntarily adhere to and contribute to the development of an evolving set of principles including open use, collaborative development, non-overlapping and strictly-scope content, and common syntax and relations, based on ontology models that work well, such as the Gene Ontology (GO).

The OBO Foundry is overseen by an Operations Committee with Editorial, Technical and Outreach working groups. The processes of the Editorial working group are modeled on the journal refereeing process. A complete treatment of the OBO Foundry is given in “The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration”.

On this site you will find a table of ontologies, available in several formats, with details for each, and documentation on OBO Principles. You can contribute to this site using GitHub OBOFoundry/OBOFoundry.github.io or get in touch with us at obo-discuss@sourceforge.net.
# Modern Information System Design moving to Ontology

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bfo</td>
<td>Basic Formal Ontology - The upper level ontology upon which OBO Foundry ontologies are built.</td>
</tr>
<tr>
<td>chebi</td>
<td>Chemical Entities of Biological Interest - A structured classification of molecular entities</td>
</tr>
<tr>
<td></td>
<td>of biological interest focusing on 'small' chemical compounds.</td>
</tr>
<tr>
<td>doid</td>
<td>Human Disease Ontology - An ontology for describing the classification of human diseases</td>
</tr>
<tr>
<td></td>
<td>organized by etiology.</td>
</tr>
<tr>
<td>go</td>
<td>Gene Ontology - An ontology for describing the function of genes and gene products.</td>
</tr>
<tr>
<td>obo</td>
<td>Ontology for Biomedical Investigations - An integrated ontology for the description of life-</td>
</tr>
<tr>
<td></td>
<td>science and clinical investigations.</td>
</tr>
<tr>
<td>pato</td>
<td>Phenotype And Trait Ontology - An ontology of phenotypic qualities (properties, attributes or</td>
</tr>
<tr>
<td></td>
<td>characteristics).</td>
</tr>
<tr>
<td>po</td>
<td>Plant Ontology - The Plant Ontology is a structured vocabulary and database resource that</td>
</tr>
<tr>
<td></td>
<td>links plant anatomy, morphology and growth and development to plant genomics data.</td>
</tr>
</tbody>
</table>
The Open Residential Real Estate Ontology (ORREO) Foundry is a collective of ontology developers that are committed to collaboration and adherence to shared principles. The mission of the Foundry is to develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate. To achieve this, ORREO Foundry participants voluntarily adhere to and contribute to the development of an evolving set of principles including open use, collaborative development, non-overlapping and strictly-scoped content, and common syntax and relations, based on ontology models that work well, such as the Gene Ontology (GO).

The ORREO Foundry is overseen by an Operations Committee with Editorial, Technical and Outreach working groups. The processes of the Editorial working group are modelled on the journal refereeing process. A complete treatment of the ORREO Foundry is given in "The ORREO Foundry: coordinated evolution of ontologies to support residential real estate data integration".
Project ORREO Foundry

Resources

- ORREO Foundry @ Github.com
  - MIT Commons License
- Website: [http://www.orreofoudry.org](http://www.orreofoudry.org)
  - Tools
  - Ontological Browser
  - Protégé Support
- OWL 2 DL
  - OWL/XML, RDF/XML, JSON-LD, Turtle serialization

- DL Reasoners
  - Common Logic Reasoner
  - DL Expression Explorer
- BFO and IAO Upper Ontologies
- Modular Design pioneered by Dr. Barry Smith (Ontological Engineering)
- Autodesk BIM Integration
  - Building Information Model Integration
- MS Azure IoT
- Smart City Collaboration
Project ORREO Foundry and RESO

- ORREO Foundry is about Real Estate and should be collaborative in its workings.
- Ontology Development effort is never considered to reach a complete state but Ontology do come to a “good-enough” state.
- RESO will play an important role providing guidance and experience and insights.
- RESO+ORREO will bring many more Real Estate professionals to the party; thus expanding reach of knowledge, learnings, experience and drive new business and technological innovation.

- People Ontology
- Email Ontology
- Emotion Ontology
- Material Ontology
- Geographical Ontology
- Many others
- ORREO will create new ontologies that span all aspects of reality where real estate is applicable.
Project ORREO Foundry: Multidisciplinary Teams

- Building Architects
- Home Builders
- Landscape Designers and Engineers
- Interior Decorators
- Appliance Manufacturers
- Power Utility (Gas, Water, Electric, Solar)
- Heating and Cooling
- Home & Property Insurance
- Mortgage Lending (FIBO – Financial Industry Business Ontology)
- Semantic Bank Compliance Ontology + Legal Knowledge Interchange Format (LKIF)
- Construction
- Home Security
- Whole-House Systems Approach Ontology @ Energy.gov
- More..
Start with BFO and IAO Upper Ontology
We will be able to use ontologies to help us share data only if

- they are ontologically coherent (intelligible to a human user)
- and logically coherent
- and computationally tractable
- and work well together
  - evolve together
  - created according to the tested rules
Data Model - Purpose

- To provide a consistent and efficiently functioning data store for a particular business application(s)
  - Represents specific business concepts in a way that determines organization of data in the store
  - Commonly used representations are relational and graph; they are supported by data management technologies, e.g. relational
    - Oracle and MySQL, graph – Neo4j, RDF/OWL stores.

- Efficiency requires
  - Application-specific representations
  - Store only data needed by the application
  - Objective (shared) representation of the domain is not the purpose – multiple data models for the same domain to accommodate different business applications
Ontology - Purpose

- Objectivity of representation of reality
- Commonly used representation is graph, it is supported by RDF-based semantic technologies
- Objective (shared) representation of the domain
  - one authoritative ontology for the domain of reality meant for re-use
- Storing vast volumes of data is not the purpose
Ontology - Organization

- Each type appears only once in the ontology hierarchy.
- The ontology view of reality is synoptic – it represents in non-redundant fashion an entire hierarchy of types at different levels of generality. Each term is associated in an intelligible way with its subsuming and subsumed terms (and thus with the ancestor and descendant types) in the hierarchy of more and less general.
- Representation is more flexible, changes are easier to make, and changes are not as disruptive.
Knowledge Designed for Human Understanding
Knowledge Designed for Machine Understanding

MKVSDRRKFKEKANFDEFESALNNKNDLVHCPSITLFESEIPTEVRSFY EDEKSGLIKKVFKRTAGMDKRKSFEKVVISVMVGKKNKKFLTFVED EPDFQGGPISKYLERKLKINLVMVYTLFQVHTKLKNRKYDTSLFLYNR GYNNLSFRVLERCHEIASAPNDSSTMRTFTDFVSGAPIVRSLQQK STIRKYGYNLAPYMLLHVDELTSIASAYQALQGEKKVDDTERLKRDL CPRKPIEIKYFSQICNDMNKRDRLGDILHIILRACALNFLGAGPRGG AGDEEDRSITNIEPSVDEHGLKVLKLSQFTPRLRKLTDVAKALLVSSC TARDLDDFDDANGVAMWIKIILLYHEAVETTLKDSYRIT LVPSSDGISSLAFAGPQRNVYVDTTRIRQLYTDYNKNGSSEPRKLT LDGLTSDYVFYFVTLRQMQLCALGNSYDAFROHDPWMDVGFEDP NQVTRNRDISRIVLYSYMFLNTAKGCLVEYATFQYMRELPKNAPQKL NFREMRQGLIALRHCGRVSFFETDLYESATSELMANHSVQTGRNIY GVDFSLSVSGLTATLLQERASERWIQWLGLSDDYHCSFSGTRNAE DVIDSRIVLYSYMFLNTAKGCLVEYATFQYMRELPKNAPQKLNFRE MRQGLIALRHCGRVSFFETDLYESATSELMANHSVQTGRNIYGVDF SLTSVSGLTATLLQERASERWIQWLGLSDDYHCSFSGTRNAEDV
Semantic Enhancement of Data Models by Ontology

- Semantic Enhancement (SE) is realized with the help of ontologies that are used to explicate data models and annotate data instances
  - Vocabulary of ontologies used for explications and annotations provides agile horizontal integration
  - Ontologies, by virtue of their nature and organization, provide semantic enhancement of data

<table>
<thead>
<tr>
<th>PersonID</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>Java</td>
<td>Programming</td>
</tr>
<tr>
<td>222</td>
<td>SQL</td>
<td>Database</td>
</tr>
</tbody>
</table>
The Meaning of ‘Enhancement’

- Semantic enhancement/enrichment of data = arm’s length approach (no change to data) – through simple explication we associate an entire knowledge system with a database field
  - enables analytics to process data, e.g. about computer skills, “vertically” along the Skill hierarchy, as well as “horizontally” via relations between Skill and Education.
  - and further… while data in the database does not change, its analysis can be richer and richer as our understanding of the reality changes

- For this richness to be leveraged by different communities, persons, and applications it needs to have the properties mentioned above and be constructed in accordance with the principles of the SE (see References)
Semantic Enhancement and Data Integration

- Traditional integration approaches involve creation of a new model in
  - A new physical store (data warehouse)
  - Another data store – rigid (potential data silo), interoperable with other stores
  - Querying the data sources via it
    - Fragile
  - Both entail loss and or distortion of data and semantics, and provide only ‘local’ integration (do not lead to interoperability with other sources)

- SE of a store
  - Does not require data reorganization and creation of another store
  - Changes to it are non-intrusive
  - Leads to integration of the store with other stores, enhanced previously or in the future
The Languages of Ontology, Knowledge and Reasoning

Figure 1. The Structure of OWL 2
The Languages of Ontology, Knowledge and Reasoning

OWL Language

- Three species of OWL
  - **OWL Full** is union of OWL syntax and RDF
  - **OWL DL** restricted to FOL fragment (≡ DAML+OIL)
  - **OWL Lite** is “simpler” subset of OWL DL
- Semantic layering
  - OWL DL ≡ OWL full within DL fragment
- OWL DL based on SHIQ Description Logic
- OWL DL Benefits from many years of DL research
  - Well defined semantics
  - Formal properties well understood (complexity, decidability)
  - Known reasoning algorithms
  - Implemented systems (highly optimised)
Reasoning with Description Logics

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Semantics</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$A' \subseteq A''$</td>
<td>atomic concept</td>
</tr>
<tr>
<td>$R$</td>
<td>$R' \subseteq R'' \times R'''$</td>
<td>atomic role</td>
</tr>
<tr>
<td>$T$</td>
<td>$T' \subseteq T''$</td>
<td>top (most general) concept</td>
</tr>
<tr>
<td>$\bot$</td>
<td></td>
<td>bottom (most specific) concept</td>
</tr>
<tr>
<td>$\neg A$</td>
<td>$A' \setminus A''$</td>
<td>atomic negation</td>
</tr>
<tr>
<td>$C \cap D$</td>
<td>$C' \cap D'' = C' \cap D''$</td>
<td>intersection</td>
</tr>
<tr>
<td>$\forall R.C$</td>
<td>$\forall b \in R.C (b \in C' \Rightarrow b \in C''')$</td>
<td>value restriction</td>
</tr>
<tr>
<td>$\exists R.T$</td>
<td>$\exists b \in R.T (b \in C' \Rightarrow b \in C''')$</td>
<td>limited existential quantification</td>
</tr>
</tbody>
</table>

$\mathcal{AL}$ (attributive language) logic syntax and semantics

<table>
<thead>
<tr>
<th>Name</th>
<th>Syntax</th>
<th>Semantics</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\cup$</td>
<td>$C \cup D$</td>
<td>$C' \cup D'' = C' \cup D''$</td>
<td>union of two concepts</td>
</tr>
<tr>
<td>$\exists R.C$</td>
<td>$\exists b \in R.C$</td>
<td>$\exists b \in C'$</td>
<td>full quantification</td>
</tr>
<tr>
<td>$\forall R.C$</td>
<td>$\forall b \in R.C$</td>
<td>$\forall b \in C'$</td>
<td>number restriction</td>
</tr>
<tr>
<td>$\neg C$</td>
<td>$\neg C'$</td>
<td></td>
<td>negation of arbitrary concept</td>
</tr>
</tbody>
</table>

Some further extensions of $\mathcal{ALC}$ logic that will be of interest for us as follows.

- $S$ - role transitivity $Trans(R)$ (asserting that role is transitive)
- $R^-$ - role hierarchy $RCS$ (asserting hierarchy of roles)
- $I$ - role inverse $R^-$ (creating inverse role)
- $F$ - functionality $\leq F$ (functional role in concept creation)
- $O$ - nominals $a_1, \ldots, a_n$ (concept declared by enumeration)
## Reasoning with Description Logics

### Complexity of reasoning in Description Logics

Base description logic: **Attributive Language with Complements**

$\mathcal{ALC} = \{ T, A, \neg C, O, D, U, C, D, \mathcal{SRI}, \mathcal{WKC} \}$

### Concept constructors:

- **$F$** - functionality ($\{ 1 \}$)
- **$N$** - (unqualified) number restrictions: $(\exists n \ R), (\exists n \ R)$
- **$Q$** - qualified number restrictions: $(\exists n \ R.C), (\exists n \ R.C)$
- **$O$** - nominals: $(a) \text{ or } \{a, \ldots, z\}$ ("one-of")
- **$\mu$** - least fixpoint operator: $\mu XC$

*Formulas with complex roles in number restrictions are in $\mathcal{ALC}^2$.*

### RBox (role axioms):

- $\mathcal{S}$ - role transitivity: $\text{Tr}(R)$
- $\mathcal{H}$ - role hierarchy: $R \sqsubseteq S$
- $\mathcal{R}$ - complex role inclusions: $R \sqsubseteq S \sqsubseteq R$
- $\mathcal{C}$ - some additional features (Click to see them)

### TBox (concept axioms):

- **Empty TBox**
- **Axiom** $A = C$, $A$ is a concept name, no cycles
- **General TBox** $C \sqsubseteq D$, for arbitrary concepts $C$ and $D$

### Complexity of reasoning problems

- Concept satisfiability: **PSpace-complete**
  - Hardness for $\mathcal{ALC}$: see [80].
  - Upper bound for $\mathcal{ALC}^2$ see [12, Theorem 4.6].
- ABox consistency: **PSpace-complete**
  - Hardness follows from that for concept satisfiability.
  - Upper bound for $\mathcal{ALC}^2$: see [12, Appendix A].

### Important properties of the Description Logic

- **Finite model property**
- **Tree model property**

$\mathcal{ALC}$ is a notational variant of the multi-modal logic $\mathcal{K}_m$ (cf. [77]), for which the finite model property can be found in [8, Sect. 2.3].

$\mathcal{ALC}$ is a notational variant of the multi-modal logic $\mathcal{K}_m$ (cf. [77]), for which the tree model property can be found in [4, Proposition 2.15].

---

Maintained by: Evgeny Zolin
Please see the list of updates

Any comments are welcome: E.Zolin@cs.man.ac.uk
The purpose of authoring ontologies is also reusing of knowledge. Once ontology is created for a domain, it should be (at least to some degree) reusable for other applications in the same domain. To simplify both ontology development and reuse, modular design is beneficial. The modular design uses inheritance of ontologies - upper ontologies describe general knowledge, and application ontologies describe knowledge for a particular application.
Questions?
Thank You, Everyone!