

Upper-Ontologies a closer look

Fausto Giunchiglia and Mattia Fumagalli

University of Trento



- *Benefits of using an Upper Ontology*
- *Criteria for choosing*
- *Academia vs. Industry*
- *Upper Ontologies as part of a broader set of resources for building enterprise ontologies*
- *Overviews of some well known Upper Ontologies*
- *Discussion: Choices and Tradeoffs*

Why not start from scratch?

- **Build your ontology faster:**
 - reuse other people's thinking
- **Build a better ontology**
 - starting with a solid foundation
 - if highly axiomatized, can catch many errors
- **Enhanced interoperability from sharing**

- **But which upper ontology to choose?**
Cyc, SUMO, DBpedia, DOLCE, BFO, shema.org ...

Build your ontology faster

- Don't reinvent the wheel
- Do you really want to have to think about how to model "Person", "Organization" or "Agreement"?
- A well-thought out UO will have most of the bugs worked out so you don't have to.
- Most things in a given organization have things in common with other organizations
- Attach things specific to your organization to the more general things

- *Enhanced clarity (removing ambiguity)*
- *Improved accuracy*
- *Reduced complexity*

- Inherit *carefully crafted meaning* from the UO
- The *discipline of deciding where in the UO a given concept in your organization belongs, forces you to ask probing questions to get deeper understanding of your concept*
- Example: if you have to model “Country” you will be forced to decide if you mean *spatial region on the earth, or if you mean a social construct.*

- Improved clarity helps reduce errors
- Connect to the upper ontology enables you to use inference to check consistency
- Example:
 - you might put *Country* as a subclass of both *Organization* and *GeoRegion*
 - If *Organization* is disjoint from *GeoRegion*, this will be flagged

- *Less is more*
- *Using the discipline of attaching every class or property to an existing class or property in the upper ontology helps reduce proliferation of new concepts*
- *It does so by highlighting “undiagnosed similarity”.*
- *Reduced complexity also contributes to clarity*

What to look for in a Upper Ontology

1. Easy to learn, understand and use
 1. Manageable scope
 2. Well-structured with visualization aids
 3. Unambiguous definitions
2. Supports required inferences
3. Ability to agree and commit to the ontology
4. Maturity, Stability, Evolution and Support
5. Business roots (v. academia)
6. User base and community
7. Mapped to other ontologies
8. Standards-based



User and
Situation
Specific

Ontological Commitment



Tradeoffs: Academic vs. Industry

- Academia: emphasis on expressive power vs. inference
- Industry: understandability & usability are the initial and primary concerns
- Inference important, but secondary

	Abstraction	Inference	Understandable/Usable
Academia	More	More	Less
Industry	Less	Less	More

- *Upper Ontologies: broad coverage*
 - *e.g., gist, DOLCE, SUMO, CYC, BFO, Schema.org*
- *Linguistic resources:*
 - *WordNet, FrameNet, VerbNet*
- *Broadly reusable, domain specific ontologies*
 - *Usually small scale, narrow coverage*
 - *e.g., Foaf, Dublin Core, SIOC, SKOS, many more*
- *Data Sets: e.g. GeoNames*

Purpose: understanding and processing natural language.

- WordNet
- FrameNet
- VerbNet
- Commercial variants of same (proprietary)

WordNet (see next class)

- Ontologically structured Thesaurus
- Contains 170,000 English words and their equivalents in many languages
- Uses “word sense” to disambiguate homonyms
- Has “broader/narrower” “partOf/part” and “hyper/hyponym” relations
- The hypernym hierarchy is a taxonomy of synonym sets, and is often used as an ontology.
- OWL: <http://www.w3.org/2006/03/wn/wn20/>

Broadly Reusable:

- **vCard, Foaf: (Friend of a Friend)**
 - People, Contact information, Who you know
- **SKOS: Simple Knowledge Organizing System**
 - For representing thesauri in OWL
- **SIOC: Semantic Interlinked Online Communities**
 - A social media ontology
- **Dublin Core: describing resources for discovery**
(dublincore.org)
- **GoodRelations, geoNames and many more...**

- *Library Science ontology for categorizing documents*
- *15 core concepts, no restrictions or axioms*
- *Very widely used*
- *Available in OWL*

- *By all means use them, if they fit your needs.*
- *But they are very different from an upper ontology*
- *Much narrower in scope*
- *Most of the criteria still apply when selecting these kinds of ontologies for reuse.*

- **Upper and Middle:**
 - Cyc: Gigantic scale
 - SUMO: Large scale
 - schema.org: recent, lightweight, growing fast
- **Upper Only: small scale**
 - DOLCE: academic, philosophical underpinning
 - Basic Formal Ontology (BFO): philosophical & scientific
 - gist: business focus

- Most extensive upper ontology
- 20 year effort, over a person-millennium invested
- Over 1,000,000 assertions (60,000 in openCyc)
- Written in CycL
- Goal is to mimic human common sense
- Licensable; should be considered for cases where you need to read unstructured text in an untrained environment.
- OWL:
<http://mkrmk.net/knowledge/opencyc/latest.O2.owl>

- *Rigorous definitions*
 - *Good source of inspiration*
- *Fairly mature and supported*
- *Mapped to some other ontologies*
- *Available in standard languages (e.g., OWL)*

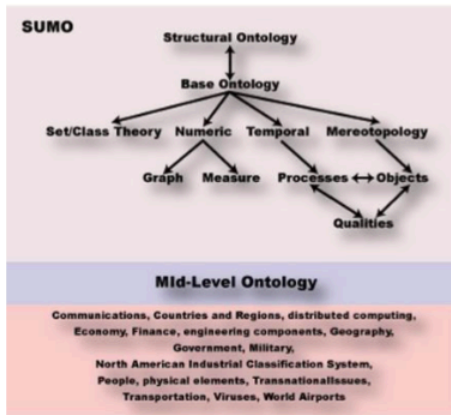
- *Hard to learn, understand and use*
 - *Huge scope*
 - *Poor visualization aids*
- *Academic roots, some business application*
- *Unknown user base*
- *Proprietary language*

For decades, someone is paying the bills, who?

- *Government research?*
- *Government applications for the agencies?*
- *Commercial clients?*
 - *I hear about them*
 - *I have seen some cases of companies having Cyc in house*
 - *Hard to tell how much value add*
- *The knowledge representation is a good source of inspiration*

- *Designed to be a mapping ontology to link many ontologies and vocabularies*
- *Carefully selected approx. 2% of all of OpenCyc*
- *Mapped to several common resources*
- *Recently published a new version (Sept 14)*
- *Unsure of its uptake*

- Consensus initiated by IEEE
- Based on C.S. Peirce and Pragmatism
- Several thousand concepts
- Nice organization into mid-level and domain-specific ontologies
- Written in Common Logic
- Owl version exists (lossy translation)



- Rigorous definitions, good at middle levels
- Possible to learn, understand and use
 - Large scope
 - Rigorous definitions
- Respectable user base
- KIF / Common Logic
- Academic roots, defense funded
- Mapped to WordNet
- Higher levels have been criticized
- One man show (Adam Pease), is evolving with ongoing projects, new open source modules are released

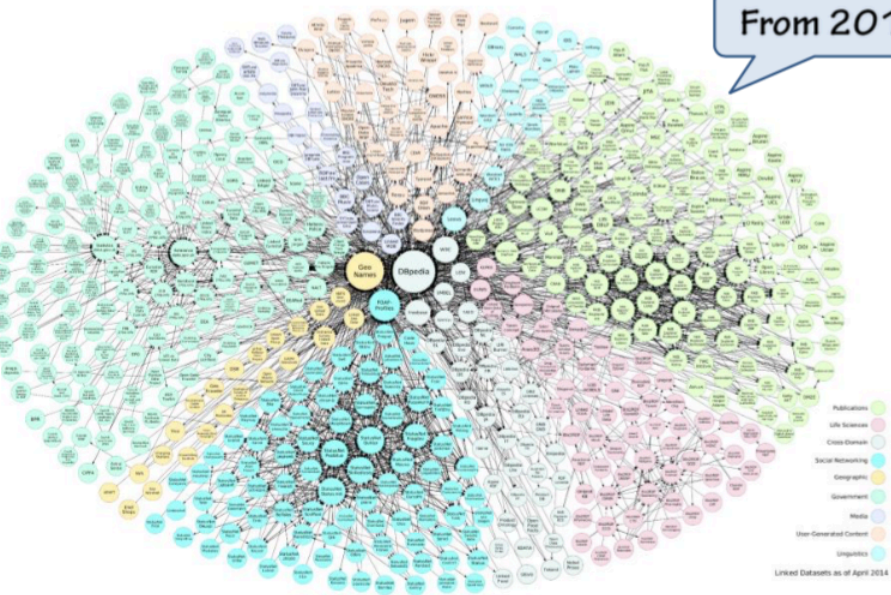
- Mostly backed by one individual, Adam Pease of Articulate Software
- Numerous contracts over the years, he gets agreement to release each extension as open source
- Web page: <http://www.ontologyportal.org/>
- A Book to describe SUMO

- *Semi-automatically extracted from Wikipedia. Originally by the YAGO project team.*
- *Backbone taxonomy is intersection of:*
 - *Wikipedia category hierarchy*
 - *WordNet synset hierarchy*
- *Properties derived from property boxes*
- *In effect, a “crowd sourced ontology”*
- *Not sure if it has a permanent home.*

- Manageable scope
- Aligned with schema.org and everything under the sun
- Few axiomatized definitions
- No visualization aids
- Crowd sourced
 - Spotty unpredictable coverage of domains
 - More on sports and entertainment than serious business topics
- Not clear how DBpedia ontology is being used
- Support subject to vagaries of funding

DBpedia: Industrial Relevance

From 2014



- Lightweight ontology for the Web
- Some prefer to call it a dictionary, but it has classes, properties, domains and ranges
- Very broad scope, growing all the time
- Mainly for web marketing and advertising
- Targeted at everyday web developers
- Available in 3 syntaxes:
 - RFDa, Microdata, JSON-LD

- Joint effort among major search engines
 - Microsoft/Bing, Yahoo, Google, Yandex
- An effort to get an agreed vocabulary for the web
- Discussion hosted by W3C,
- Implementation hosted on github

- Strong focus on usability
(given mainstream web developer user base)
- Grounded in every day terms
- No abstractions, they diminish understandability and usability
(e.g. “Agent” is too abstract)
- It is paying off
- Astonishingly fast uptake (15% of web sites)
- Anyone can participate, many are

- **Examples of collaboration**
 - WorldCat (World Catalog) from bibliographic world (Nobody uses library catalogs any more, they use Google)
 - Martin Hepp spearheading a current effort on the Automotive Industry
- Many other groups creating/evolving extensions to schema.org, as we speak

- Roles
- Actions:
<http://searchenginewatch.com/article/2340996/Schema.org-Introduces-New-Elements-for-Actions>
- Schema in email
- Details:
 - *added sameAs,*
 - *broadened domain of citation property*

- People find their own way to collaborate,
- Some prefer on web, others face to face
- public vocabularies work well
(see public-vocabs@w3.org)
- github a bit daunting for non coders
- machine readable schema critically important
- JSON-LD starting to see search engine adoption

Process:

- still making it up as we go along
- important to keep strong connection to search engines
- open cross-community dialog and collaboration
- incoming proposals are getting bigger and bigger: need new way to handle them

- **QUESTION:** Change happens. How can you avoid everything breaking that uses old way?
- **ANSWER:** (Dan Brickly, Google UK)
It's kind of messy anyway, if something used on 10s of 1000s of sites, we don't mess with it.
- **UPSHOT:**
 - It seems there is no formal process
 - A bad idea used by many will stay around ☹️

Schema.org vs Google Knowledge Graph?

QUESTION: what is the relationship between schema.org and whatever the metadata is that is behind Google's knowledge graph?

ANSWER: (Guha, Google) schema.org is a small and simple vocabulary for commerce for everyday web developers.

UPSHOT: Behind the knowledge graph, there is probably the biggest honkingist ontology on the planet that has widespread impact.

Cyc is probably bigger, but does not impact 100s of millions every day.

Schema.org is an Upper Ontology

- It was never intended to be an upper ontology
- It has no academic or enterprise roots
- It is for the web!
- Not suitable for use as an upper ontology
- BUT: it may end up being the defacto [lightweight] ontology of the world, from an every day perspective.
- SO: it could make sense to align with this vocabulary unless there are reasons not to

- European initiative, funded through EU, as part of the WonderWeb initiative
- Rigorous philosophical and mathematical underpinning
- Dolce-lite has 500 concepts
- Hard to understand
- Weaker version available in OWL

DOLCE: Fundamentals

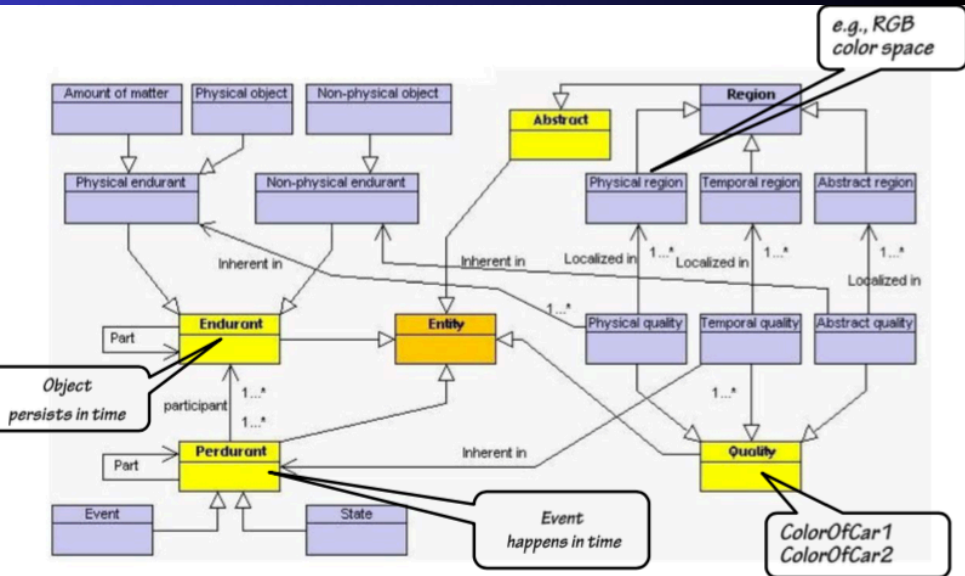


Fig. 2. A UML class diagram showing the basic classes and relations of DOLCE

DOLCE: Core Restriction

1. Is it *agentive* or not?
2. Does it happen in time (e.g., Event occurs) or persist over time? (e.g., PhysicalObject)
3. Does it occupy space/time?
If not, then DUL:Abstract
4. Are its parts the same kind of thing?
YES for Water & Sand; NO for Person & Rock
5. Does it have independent existence or is it parasitic on host?
(e.g., *edge of a table vs. leg of a table;*
the color of my car vs. a particular RGB value)

- Manageable scope
 - Impressive piece of work, rigorous definitions
 - Some visualization aids (PowerPoint)
 - OWL version
-
- Very hard to understand
 - Academic roots
 - Not supported or evolving
 - Hodgepodge of versions, not clear what to use

BFO (Basic Formal Ontology)

- Built using a consistent set of principles
- Focus on science
- Substantial user community
- Ontology evolving, albeit slowly, possibly due to large user base
- See: [BFO_Tutorial_2012.pptx](#) for detailed history and background of BFO with specific comparisons to DOLCE
- See: [BFO2-Reference.docx](#) for full reference documentation

Roots and History of BFO

- Aristotle,
- Husserl's *Logical Investigations*
- Patrick Hayes, "Naïve Physics Manifesto"
- Qualitative spatial reasoning
- DOLCE, GO, FMA
- 2004 BFO 1.0
- 2005 OBO (relations)
- 2006 BFO 1.1
- 2012 BFO 2.0

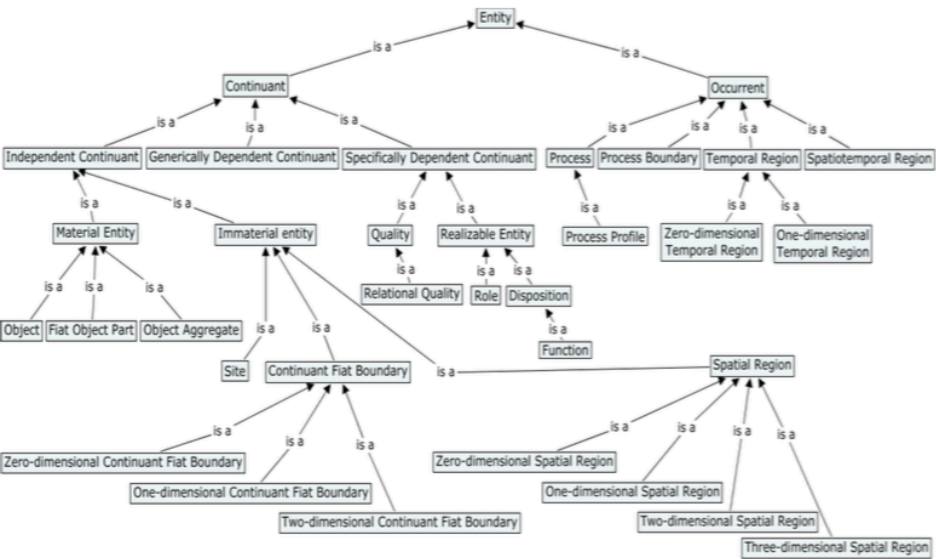
Principal features of BFO

- Quite small
- Focus on supporting integration of scientific data
- Integration works better with consistent ontologies
- Designed to support the consistent representation of different domains of reality
- Program of project-based testing, feedback & training

BFO and DOLCE share common philosophical roots

- Distinction between *continuants* and *occurrents*
- Distinction between *independent entities* and *dependent entities* ('qualities' in DOLCE)
- Just two levels, *instances* and *universals*, with no meta-level
i.e.: There is no universal universal of which first-level universals would be instances

BFO:TOP Level Class Hierarchy



Users of BFO (Consortia)

OBO Foundry	Open Biological and Biomedical Ontologies
NIF Standard	Neuroscience Information Framework
eagle-I Ontologies	eagle-I, VIVO and CTSAconnect
IDO Consortium	Infectious Disease Ontology
CROP	Common Reference Ontologies for Plants

- Manageable scope
- Rigorous definitions
- Consistent principles for development
- OWL version

- Fairly abstract concepts, a barrier to understandability and take-up in industry
- Academic roots, philosophical foundations
- Moderate amount of practical use (it seems)
- State of support and evolution unclear, last news was in July 2012
- Version 2.0 incompatible with version 1.1, some guidelines provided

What kind of top-level ontology?

Comparing Upper Ontologies

U & U: Usability & Understandability

	Scope / Focus	Scale	Rigor	Representation	Community	U & U	Evolving
Cyc	Upper & Middle	Huge	High	FoL + Meta	Unclear	Very low	Yes
DOLCE	Upper	Small	High	FoL,	Small	Low	No
BFO	Upper	Small	High	FoL, OWL	Yes	Medium	?
DBpedia	Upper & Middle with deep bits	Medium	Medium	Rdf? OWL?	Unclear	Medium	?
Schema.org	Upper & Middle with deep bits	Growing Large	Low	RDFS- (?)	Very large	High	Very active
SUMO	Upper & Middle with deep bits	Med / Large	High	FoL, KIF Common Logic	Some traction	Medium?	Yes, slowly
WordNet,	Linguistic	Large	Low	?, OWL	Yes	Medium	?
SIOC	Social	V. Small	?	?, rdf? OWL?	?	High	?
foaf	People	V. Small	Low?	Rdf? OWL?	Yes	High	no
gist	Upper Business	Small	Medium	OWL	Small	High	yes
Dublin Core	Library references	V. Small	Low?	Rdf? Owl?	?	Med-High?	no

Expressivity and Inference

- Cyc: super-expressive, super-complicated, inference just about impossible.
- SUMO: uses first order logic, takes several days to run full inference.
- DOLCE: fairly expressive, until it was dumbed down into OWL
- BFO: fairly expressive, some dumbing down into OWL
- schema.org: very lightweight and informal, limited inference

Which do you think would be easier for a business technical person to grasp? Which one would you like to teach?

1. specifically/generically dependent
continuant, fiat object part, continuant fiat
boundary, realizable entity, relational quality
2. Person, Time Interval, Organization,
Agreement, Landmark, GeoRegion, Content,
Event, PhysicalThing

- You can **understand** it
- You can **agree** with it
- You can **use** it
efficiently and
reliably
- Everything else is gravy

Question for Discussion

- With huge success and uptake of lightweight ontologies like schema.org, what is the role of more formal & rigorous ontologies like DOLCE, SUMO, BFO?
- What will be unique about the demands of any niche communities that require more rigorous ontologies?
- What are these niches communities going to be, that demonstrate the relevance of rigorous ontologies?

Acknowledgments

These slides have been inspired by (or reuse) (possibly adapted) content included in the following material:

“Upper Ontologies a brief tour of what is available by Michael Uschold (www.semanticarts.com)”