Driving the next generation of spatial standards: examples from hydro ontology

Torsten Hahmann

Dept. of Computer Science, University of Toronto

from Nov.: National Center for Geographic Information & Analysis (NCGIA)
School of Computing and Information Science
University of Maine

October 17, 2013

Why bother standardizing geospatial terms?

1. Badly needed
   - Basis for data & information integration and exchange in the earth sciences (compare the goals of EarthCube)
   - Supports other disciplines: much information has a spatial component
2. Many intuitive conceptualization, concepts, and relations, each of which comes with various interpretations that slightly differ
   - Conceptualizations: raster/grid- vs. vector- vs. graph/network-based, discrete vs. continuous, flat vs. spherical, ...
   - Concepts: *boundary*, *surface*, *curve*, *region*, *hole*, *water body*, *lake*, toponyms, ...
   - Relations: *in contact with*, *is part of*, *contains*, ...

The role of standards: reference + implementation guide

Reference that defines a reusable terminology with shared semantics

- **standardized terms**
- **standardized definitions of the terms' meanings**

Expressive formal ontologies can help achieve both:

- **Terminology:** concepts and relations in a logical language
- **Shared semantics:** axioms that constrain the interpretation of the terms and help disambiguate the terms

Ontologies are *heavy* vs. *light* analogous to *reference manual* vs. *user guide*

- A reference ontology is necessarily **heavy**: complete, formal, rigorous
- Implementation/user guide is usually light

Current geospatial standards

ISO/OGC Simple Features, OGC GeoSPARQL, Spatial Schema (ISO 19107), Ordnance Survey Spatial Relations, GML, hydro ontologies (GWML, INSPIRE, SWEET)

- Specified using UML, RDF Schema, or lightweight OWL (OWL-DL)
  - **Light: standardize the terms** (vocabulary)
  - **Don't formalize the terms' meaning**: not a formal reference
  - Only the beginning of exploiting the benefits of ontologies for standards
  - Even the expressive power of OWL-DL not fully exploited yet
- Relations (between concepts) are less emphasized than concepts
  - Relations tie concepts together: need relations to describe how certain concepts relate to one another (incl. behaviour)
  - OWL language is less expressive with respect to relations
- Many concepts and relations are already formalized
  - e.g. mereotopological relations (RCC and Egenhofer's 9-intersections) are included in GeoSPARQL, Simple Features, and Ordnance Survey Spatial Relations, but **neither use the known logical formalizations**

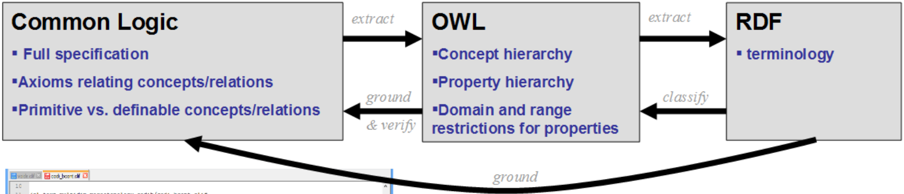
How can expressive ontologies help improve standards?

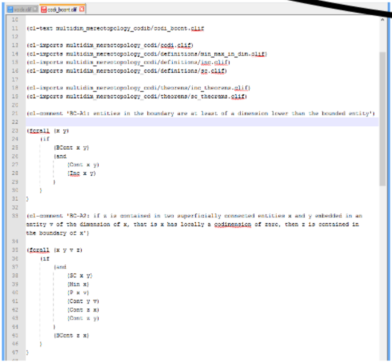
General idea:

1. Identify key concepts and relations (terminology)
2. Axiomatize them in an expressive logic (e.g. Common Logic)
   - Identify primitive vs. definable concepts and relations
   - Constrain primitive concepts/relations
   - Define definable concepts/relations
3. Extract concept and relation hierarchies and verify
   - Use automated theorem provers
   - Verify consistency and that concepts and relations can be non-empty
   - Extract, e.g., `subclass` and `subproperty` relationships
   - Extract `DisjointWIth` and `DisjointUnionOf` conditions
   - Essentially extract a lightweight ontology

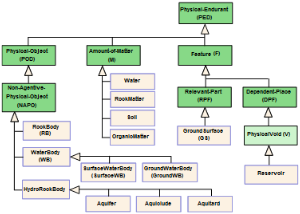
Will use examples from current work on hydro ontology as examples

Goal: Multiple consistent representations of a standard





```
(cl-text multidim_mereotopology_codib/codi_bcont.clif

(cl-imports multidim_mereotopology_codi/codi.clif)
(cl-imports multidim_mereotopology_codi/definitions/min_max_in_dim.clif)
(cl-imports multidim_mereotopology_codi/definitions/inc.clif)
(cl-imports multidim_mereotopology_codi/definitions/sc.clif)

(cl-imports multidim_mereotopology_codi/theorems/inc_theorems.clif)
(cl-imports multidim_mereotopology_codi/theorems/sc_theorems.clif)

(cl-comment 'BC-A1: entities in the boundary are at least of a dimension lower than the bounded entity')

(forall (x y)
    (if
        (BCont x y)
        (and
            (Cont x y)
            (Inc x y)
        )
    )
)

(cl-comment 'BC-A2: if x is contained in two superficially connected entities x and y embedded in an entity v of the dimension of x, that is x has locally a codimension of zero, then z is contained in the boundary of x')

(forall (x y v z)
    (if
        (and
            (SC x y)
            (Min x)
            (P x v)
            (Cont y v)
            (Cont z x)
            (Cont z y)
        )
        (BCont z x)
    )
)
```



Concepts:

PED, POB, M, F, RPF, DPF, V, WaterBody, RockBody

Relations (Properties):

Cont (contains), inside, surrounds, hosts, hosts-v, DK (constitutes), P (part_of), PO (overlap), Inc (incidence), BCont (boundary contained), etc.

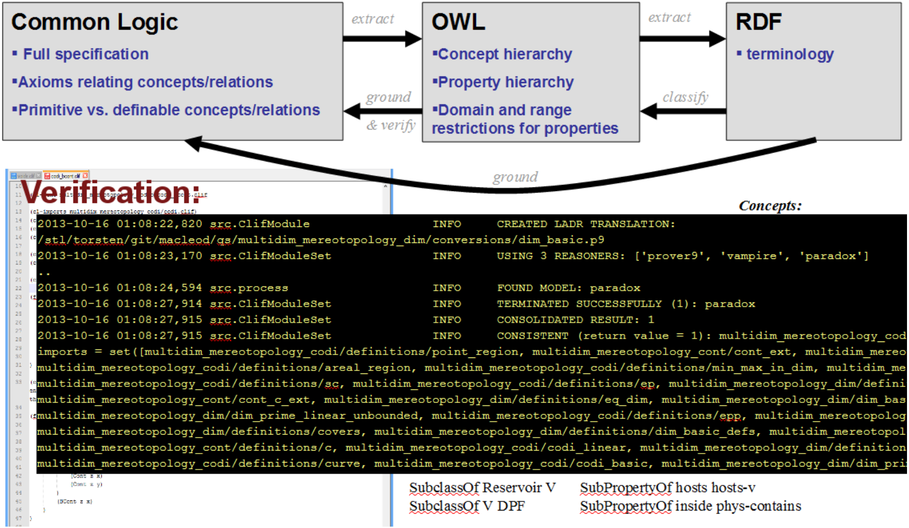
Concept hierarchy:

SubclassOf Reservoir V
SubclassOf V DPF

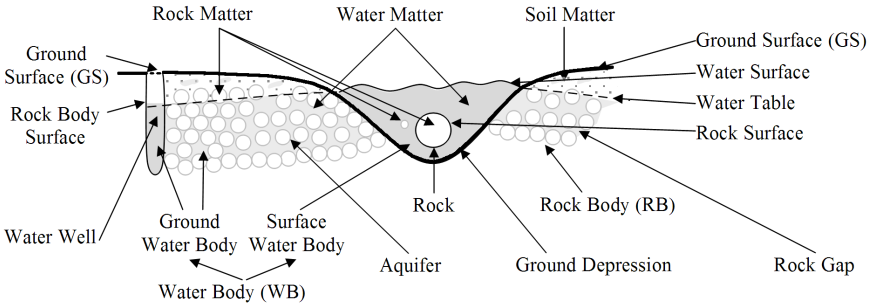
Property hierarchy:

SubPropertyOf hosts hosts-v
SubPropertyOf inside phys-contains

Goal: Multiple consistent representations of a standard

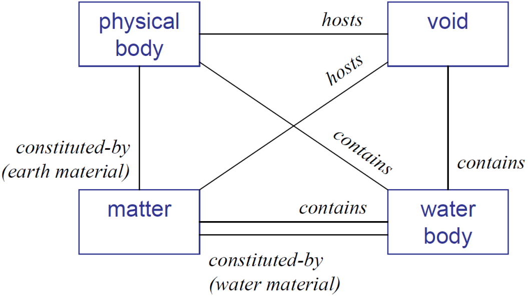


Our work: Develop **formally grounded** hydro ontology



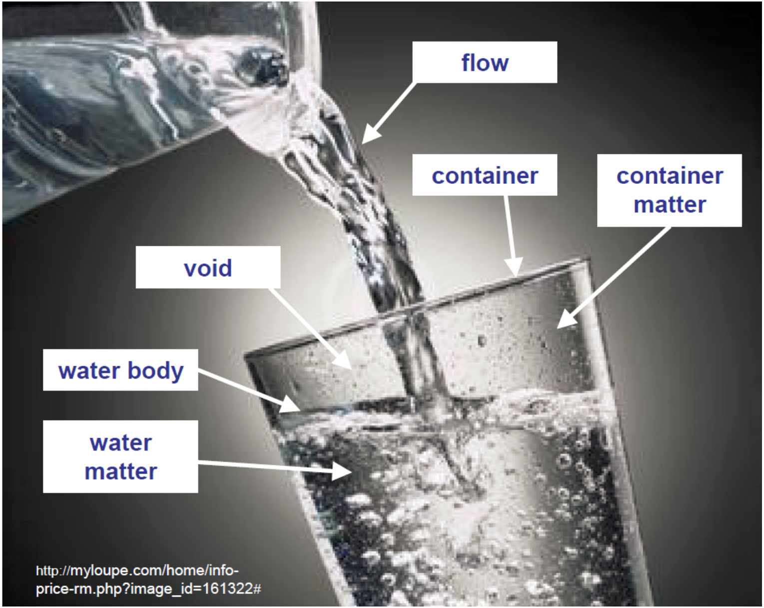
1. Rigorous formalization in Common Logic
2. Verification: assisted by first-order theorem provers; partly automated
   - prove consistency
   - prove coverage: exhaustiveness of concepts/relations
   - prove intuitive intended relationships (‘theorems’)
3. Extract taxonomies to extend OWL version of DOLCE upper ontology

Starting point: Basic elements of a hydro ontology



- Develop a rigorous formalization of these concepts and relations in a formal logic & extract a consistent lightweight vocabulary

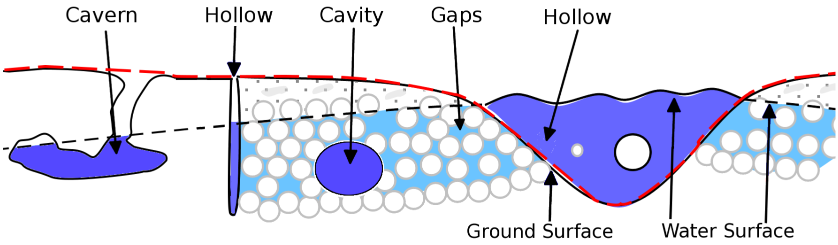
Basic elements of a reference hydro ontology: Analogy



Water and physical bodies (Hahmann & Brodaric, 2012)

Define water bodies by their physical containers' voids

- Lake or River WB: in a hollow of the ground surface
- Water Well WB: in a hollow below the ground surface
- Aquifer WB: in gaps in the rock matter and in holes below the ground surface



Example axioms: Water and rock bodies

A WaterBody may only be constituted by water if it has constituents:

$$WB(x) \rightarrow NAPO(x) \wedge \forall y[DK_1(y,x) \rightarrow Water(y)]$$

A RockBody is constituted by rock matter and only by rock matter:

$$RB(x) \equiv NAPO(x) \wedge \exists y[DK_1(y,x)] \wedge \forall y[DK_1(y,x) \rightarrow RockMatter(y)]$$

GS denotes a ground surface (not fully defined):

$$GS(gs) \rightarrow RPF(gs) \wedge \exists o[NAPO(o) \wedge hosts(o,gs)]$$

$WB, RB, GS, Water, RockMatter$ ....Domain theory (Hydrogeology)

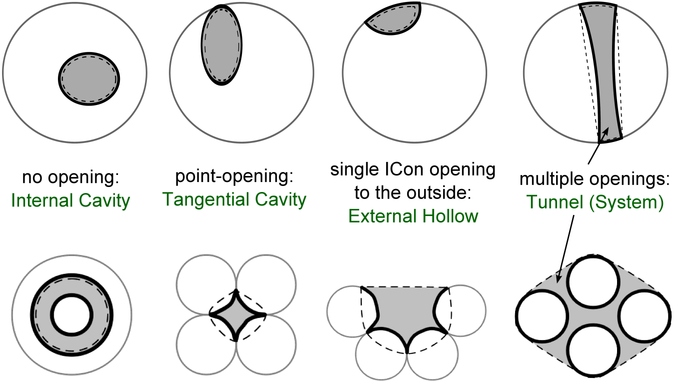
$NAPO, RPF, DK_1, hosts$ ....DOLCE concepts/relations

Voids (Hahmann & Brodaric, 2012)

Holes vs. Gaps: based on whether the host is internally self-connected

Cavities vs. Tunnels vs. Depressions: based on the void's opening

Opening to the outside vs. opening to other voids only



Example axioms: Water and rock bodies (contd.)

Surface- vs. Ground-WaterBody:

$$SurfaceWB(wb) \rightarrow WB(wb) \wedge \exists gs[\mathbf{hol_e}(wb, gs) \wedge GS(gs)]$$

$$GroundWB(wb) \rightarrow WB(wb) \wedge \exists rb, gs\big[RB(rb) \wedge hosts(rb, gs) \wedge GS(gs) \wedge r(wb) \subseteq \mathbf{voidspace}(rb) \wedge \forall v[\mathbf{hol_e}(rb, v) \rightarrow \neg PO(wb, v)]\big]$$

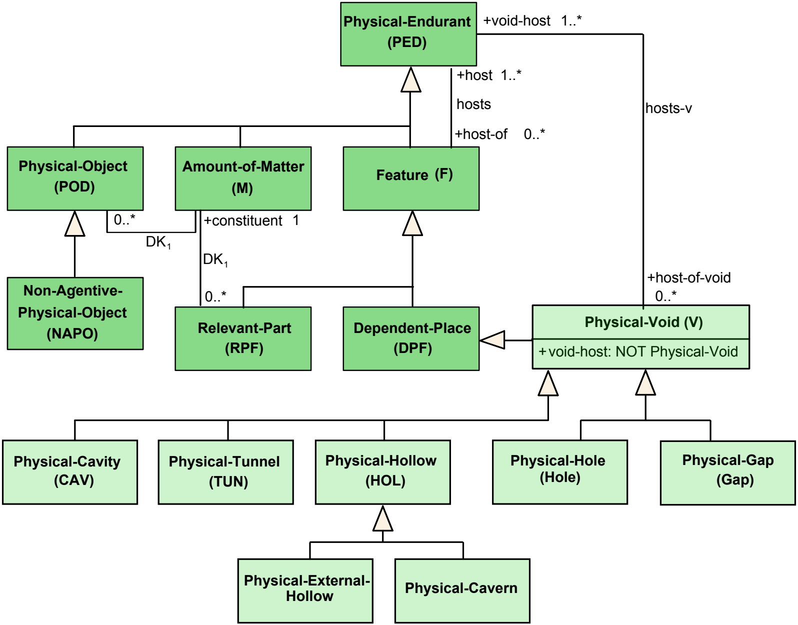
A HydroRockBody consists of a RockBody and a GroundWaterBody with the GroundWaterBody located in Voids of the RockBody:

$$HydroRockBody(aq) \rightarrow NAPO(aq) \wedge \exists rb, wb\big[r(aq) = r(rb) + r(wb) \wedge RB(rb) \wedge GroundWB(wb) \wedge r(wb) \subseteq \mathbf{con\text{-}voidspace}(rb)\big]$$

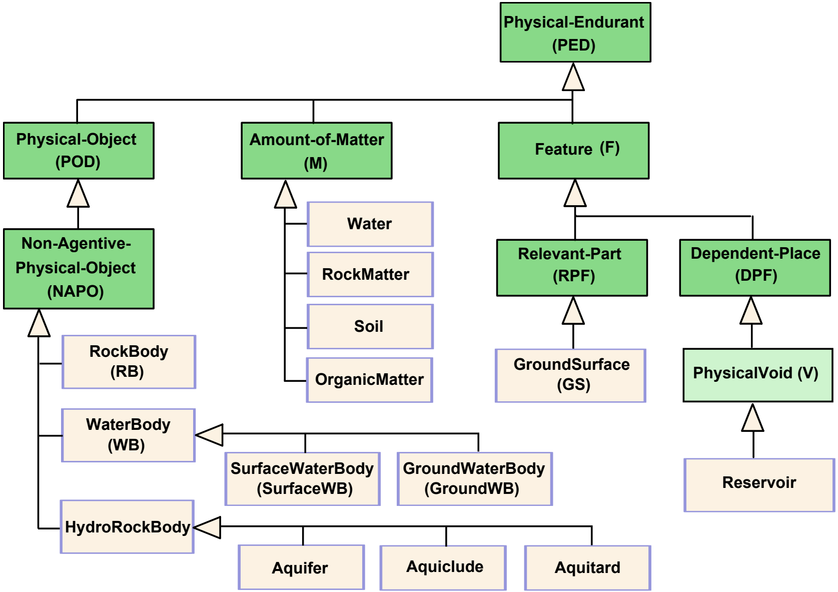
A Reservoir is the voidspace of some RockBody:

$$Reservoir(wr) \equiv \mathbf{V}(wr) \wedge \exists rb[RB(rb) \wedge r(wr) = \mathbf{voidspace}(rb)]$$

DOLCE with voids – OWL version

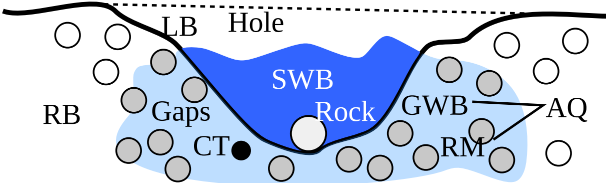


DOLCE with hydrogeology concepts – OWL version



Containment relations (Hahmann & Brodaric, 2013)

Relate voids, water bodies, and other physical bodies through containment relations



openly-surrounds-mat$(\mathrm{RB}, \mathrm{SWB})$
hosts-v$(\mathrm{RB}, \mathrm{Hole})$
mat-inside$(\mathrm{Rock}, \mathrm{Hole})$
materially-contains$(\mathrm{AQ}, \mathrm{GWB})$
encloses-mat$(\mathrm{AQ}, \mathrm{CT})$
mat-inside$(\mathrm{Gaps}, \mathrm{GWB})$
encloses-mat$(\mathrm{GWB}, \mathrm{CT})$

openly-surrounds-mat$(\mathrm{RB}, \mathrm{Rock})$
mat-inside$(\mathrm{SWB}, \mathrm{Hole})$
openly-surrounds-mat$(\mathrm{SWB}, \mathrm{Rock})$
materially-contains$(\mathrm{AQ}, \mathrm{RM})$
hosts-v$_{\mathrm{any}}(\mathrm{AQ}, \mathrm{Gaps})$
mat-inside$(\mathrm{Gaps}, \mathrm{CT})$

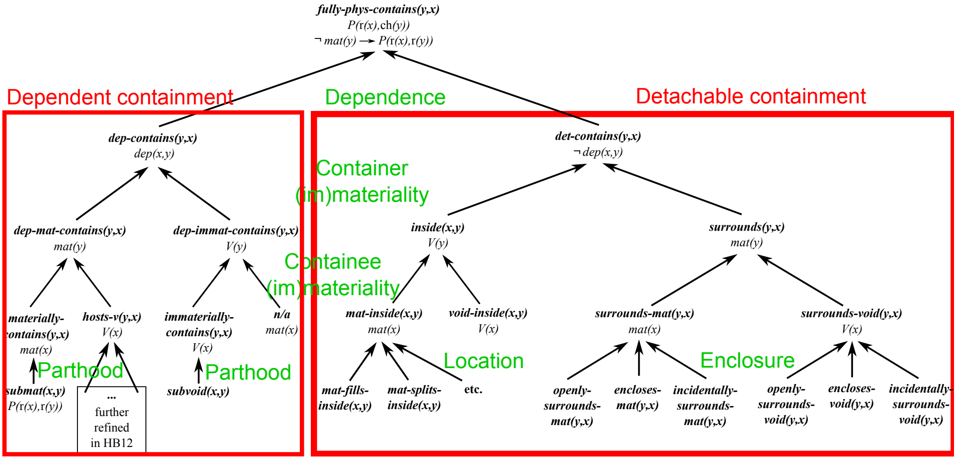
Containment relations: Heavy approach first

- Precise definitions based on topological-geometric containment relations, physical constraints and DOLCE concepts:

$$\textit{fully-phys-contains}(y,x) \leftrightarrow PED(x) \wedge PED(y) \wedge P(\mathrm{r}(x), \mathrm{ch}(y)) \wedge \left[\neg mat(y) \rightarrow P(\mathrm{r}(x), \mathrm{r}(y))\right]$$

- Classify physical containment relations based on
  1. whether container and containee are in a **physical dependency**
  2. whether the **container** is a material or a void endurant
     - **inside** (a void) vs. **surrounded** (by a material endurant)
  3. whether the **containee** is a material or void endurant
  4. other spatial relations: enclosure, contact, spatial parthood
- The resulting "leaf" relations are exhaustive and pairwise disjoint

Containment relations: Light version follows



Taxonomy expressible in OWL using `subproperty` **and** `DisjointWith` **relationships; cannot express exhaustiveness in OWL**

Conclusions

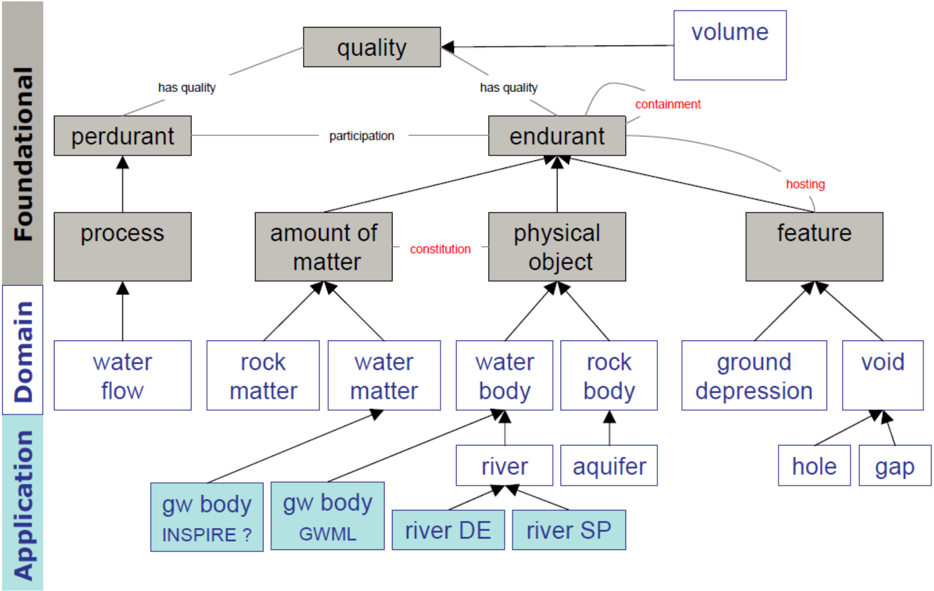
Critical to ground any lightweight implementation representation ("user guide") in a formal reference ("technical specification")

- Formally grounds and disambiguates geospatial concepts
- Serves as basis for **(semi-)automated** extraction of lightweight versions (OWL, RDF) that can be used as terminological reference for annotation or implementation in a triple store
- Formal specification helps automated verification

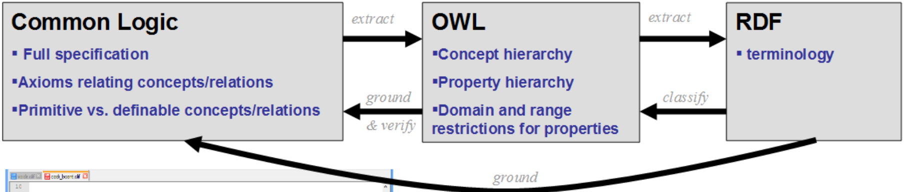
Standards should be flexible in two ways

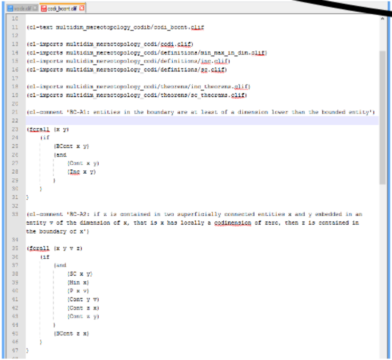
1. Amendable to various applications or domains, i.e., not too specific
2. Offer various degrees of formality
   - Most formal: for reference, verification, and heavyweight reasoning
   - Least formal: as terminology for annotating data ('Linked Data')

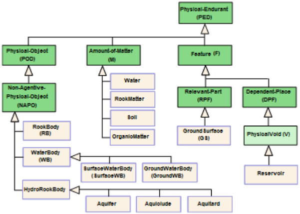
1) **Formally grounded** tiered standards



2) Various degrees of formality







Concepts:

PED, POB, M, F, RPF, DPF, V, WaterBody, RockBody

Relations (Properties):

Cont (contains), inside, surrounds, hosts, hosts-v, DK (constitutes), P (part_of), PO (overlap), Inc (incidence), BCont (boundary contained), etc.

Concept hierarchy:

SubclassOf Reservoir V
SubclassOf V DPF

Property hierarchy:

SubPropertyOf hosts hosts-v
SubPropertyOf inside phys-contains

Two general observations

- Much work on upper ontologies, less on the middle layer
  - upper ontologies can only be formalized to a certain degree
  - narrow application-specific ontologies are often too tedious to formalize
  - missing the middle layer: can be formally standardized
    - specific enough but also not too many concepts and relations
    - that's the level where information integration and exchange happens
- Relations are often still neglected: less understood?
  - Relations define how concepts relate to one another

Publications

T. Hahmann, B. Brodaric: **The Void in Hydro Ontology**. In: Proc. of the 7th Int. Conference on Formal Ontology in Information Systems (FOIS-2012), 2012. IOS Press.

T. Hahmann, B. Brodaric: **Kinds of Full Physical Containment**. In: Proc. of the 11th Int. Conference on Spatial Information Theory (COSIT-2013), 2013, Springer.

T. Hahmann: **Reconciliation of Logical Theories of Space: from Multidimensional Mereotopology to Geometry**, PhD thesis, University of Toronto. Feb. 2013.

Full formalizations of the ontologies (in progress), COLORE repository, `http://stl.mie.utoronto.ca/colore/org.html`, Section "*Space*"

- ontologies discussed here are named `multidim_mereotopology_XXX` and `multidim_space_XXX`

Acknowledgements: Boyan Brodaric, Michael Gruninger

Thank you!