

# UC Berkeley

## International Conference on GIScience Short Paper Proceedings

### Title

Semantically Refining the Groundwater Markup Language (GWML2) with the Help of a Reference Ontology

### Permalink

<https://escholarship.org/uc/item/8cz973mw>

### Journal

International Conference on GIScience Short Paper Proceedings, 1(1)

### Authors

Hahmann, Torsten  
Stephen, Shirly  
Brodaric, Boyan

### Publication Date

2016

### DOI

10.21433/B3118cz973mw

Peer reviewed

# Semantically Refining the Groundwater Markup Language (GWML2) with the Help of a Reference Ontology

Torsten Hahmann<sup>1</sup>, Shirly Stephen<sup>1</sup>, Boyan Brodaric<sup>2</sup>

<sup>1</sup>School of Computing and Information Science, University of Maine, Orono, ME 04469, USA  
Email: {torsten.hahmann | shirly.stephen}@maine.edu

<sup>2</sup>Geological Survey of Canada, Ottawa, Canada K1A0E9  
Email: boyan.brodaric@canada.ca

## Abstract

Reference ontologies are intended to aid domain ontology design, identify gaps and inconsistencies in representations of domain information, and facilitate data interoperability. The application of a reference ontology to the water domain is untested. We present findings from using a first-order logic reference ontology for the water domain, the Hydro Foundational Ontology (HyFO), to identify and remedy semantic gaps and inconsistencies in the Groundwater Markup Language (GWML2), a data model for groundwater information with less detailed formal semantics. We express GWML2 as a logical extension of HyFO, thereby improving GWML2's compatibility with other hydro data models. We derive general desiderata for a "good" domain reference ontology in the geosciences and discuss the benefits one can expect from their use for the ontological analysis of geoscience data models.

## 1. Introduction

Effective water management requires exchanging and integrating information about the location, quantity, and flow of water throughout the water cycle. The information is typically stored in multiple data stores based on different data structures, terminologies, and light-weight ontologies, subsequently summarily referred to as *data models*. Knowledge integration and querying across these data stores requires interoperability between their representations at the syntactic, schematic, and semantic (comprising differences in terminology and definitions) levels. To prepare for automated integration of geoscience knowledge across these levels, we explore the use of a reference ontology (Noy 2004) as a tool for increasing semantic precision and coherence in geoscience data models. We specifically test this idea within the hydro domain by using the Hydro Foundational Ontology (HyFO), a reference ontology for the hydro domain developed since 2011 (see e.g., Hahmann & Brodaric 2012, 2013; Brodaric & Hahmann 2014), to semantically analyze the Groundwater Markup Language (GWML2) (Boisvert & Brodaric 2012; Brodaric 2015) as one example of a hydro data model developed by domain scientists. The result is an improved version of GWML2 with (1) increased semantic precision through axiomatic constraints (i.e., constraints expressed in a logical language such as first-order logic) and definitions based on well-defined reference terms from HyFO, (2) a stratified formalization that separates concepts based on how broadly they apply (across geosciences, to the entire water domain, or only to groundwater), and (3) a completed taxonomy that fills gaps and renames classes to better reflect their position within the stratification.

The domain reference ontology HyFO is not yet another standard that restricts fairly generic scientific terms (classes and relations), such as *geologic unit*, *water body*, or *aquifer*, to a single interpretation. Instead, it provides a neutral but concise language for describing in a machine-interpretable format how terms are used in existing data models. Such formal descriptions subsequently allow data models to be semantically compared and integrated in a largely automated fashion for integrated querying and knowledge discovery as envisioned in semantic e-science (Brodaric and Gahegan 2010).

## 2. Background and Related Work

***State-of-the-Art Semantic Representations for the Hydro Domain*** A number of data models have emerged that standardize water data syntactically and, to some extent, semantically. However, they are fragmented in that they describe only disconnected subareas of the hydro domain, such as groundwater storage and flow (e.g., GWML2 (Boisvert & Brodaric 2012; Brodaric 2015), INSPIRE Geology (INSPIRE 2013)), surface hydrography and connectivity (e.g., USGS's NHDPlusV2, INSPIRE Hydrography (INSPIRE 2009), HY\_Features (Dornblut & Atkinson 2013)), water quality (e.g., WaterML2) or stream geometry (e.g., RiverML). Moreover, the meaning of classes in the existing data models are described only via subclass relationships, via generic UML associations, and via free-text descriptions, which are insufficient for machine-interpretability and incompatible across standards. This is especially problematic for central scientific terms that are used almost universally in all hydro data models, such as *geologic unit*, *water body*, *aquifer*, or *channel*. Other concepts central to modeling water storage on the Earth, such as *container spaces* and *voids*, are omitted altogether or alluded to only vaguely.

***Existing Approaches to Semantic Integration*** Existing *ontology mapping* and *alignment* techniques as surveyed, e.g., in (Kalfoglou & Schorlemmer 2003), aim to find similarities, equivalences, and subsumption relations between the contents of ontologies. These largely automated techniques are limited in ways that prevents their use for integrating the existing hydro data models: the ontologies must (1) be specified in a language from the OWL family (Noy 2004), (2) be already syntactically and schematically integrated, and (3) be of similar scope (i.e., describe the same part of the domain). Most problematically, automatic techniques to semantically integrate different data models or ontologies can do so only to the extent to which the semantics are already specified in a machine-interpretable way. The lack of formal specifications of the semantics of the existing hydro data models requires encoding them manually. In our work presented here, we manually construct a machine-interpretable version of GWML2 by expressing its semantics using HyFO's rigorous axiomatization.

## 3. Approach Using a Domain Reference Ontology

***Nature of a Domain Reference Ontology*** Generally, a domain reference ontology can support semantic integration by providing a formal language that provides a set of *neutral, formal terms* for concisely identifying, via axiomatic mappings, the differences and nuances in the interpretations of similar concepts across data models. The set of formal terms should be small, but each term should have tightly restricted semantics. These formal terms are not meant to capture a single, agreed-up meaning of inherently complex scientific terms, but instead serve as a machine-interpretable language to precisely describe the differences between alternative interpretations of scientific terms by providing fine-grained semantic targets to map to.

Thus, a domain reference ontology must: (1) *identify a core set of formal domain concepts*, and (2) *tightly constrain and relate them axiomatically* in sufficient detail. The second aspect requires specification in an expressive<sup>1</sup> machine-interpretable language, to ensure that the formal terms are interpreted unambiguously, and to permit automation of verification and subsequent integration among data models. A reference ontology must further (3) *cover the entire domain of interest* (e.g., the hydro domain) *broadly*, meaning it should omit concepts that are only relevant in specific subdomains or applications. To ensure that the formal terms are well-distinguished from one another, the reference ontology is (4) ideally *grounded in an upper ontology* that provides the philosophical underpinning for the distinctions between the different kinds of objects and processes relevant to the domain.

<sup>1</sup> Even the ontology languages from the OWL family have proven insufficiently expressive for the purpose of a reference ontology, preferable are first-order or higher-order logics.

**The Hydro Foundational Ontology (HyFO)** In prior work, we have laid the foundation for the *Hydro Foundational Ontology* (HyFO) (Hahmann & Brodaric 2012; 2013; Brodaric & Hahmann 2014) as a reference ontology for the hydro domain. HyFO is the result of formal ontological analysis of concepts and relations that play a role in water storage on and below the Earth's surface by geo-ontology experts, rigorously formalizing them in full first-order logic as specialization of the DOLCE upper ontology (Masolo et al. 2003). HyFO identifies four key concepts that form the *Hydro Ontological Square* (Brodaric & Hahmann 2014): (1) a physical container such as a rock formation; (2) a physical void such as a depression in the ground surface, or microscopic pores in the container's matter where water can be stored, (3) a body of water located in the void (and contained by the container), (4) and the rock and water matter that constitute the container and water body. These concepts are interrelated by the relations of *containment*, *constitution*, and *hosting a void*. As result of the presented work, we propose to add *hydro rock body* as a fifth concept that represents a complex physical object that consists of (i) a container body constituted by solid (e.g., rock) matter, (ii) a void hosted therein, and (iii) a body of water that is located in the void.

#### 4. Results and Discussion

**Stratified version of GWML2** The first-order logical axiomatization of GWML2 developed here adds semantic precision and clarity to GWML2's core concepts obtained from the GWML2 conceptual schema and accompanying textual descriptions. It results in a merged ontology that treats GWML2 as a consistent logical extension of HyFO and DOLCE. At the core, it consists of a refined and stratified taxonomy spanning four layers of increasing specificity (Fig. 1): (0) DOLCE concepts; (1) generic geological concepts (*geologic unit*, *earth material*, *fluid body*) that transcend the water domain; (2) hydro concepts that span surface and subsurface water (e.g., *hydro rock body*, *water body*, *hydro void*); and (3) groundwater specific concepts (e.g., *aquifer*, *well*, *subsurface water body*, *hydrogeo void*). This layering ensures that GWML2's groundwater concepts consistently specialize HyFO concepts, with HyFO also being able to anchor surface water concepts and thus being shareable across hydro ontologies.

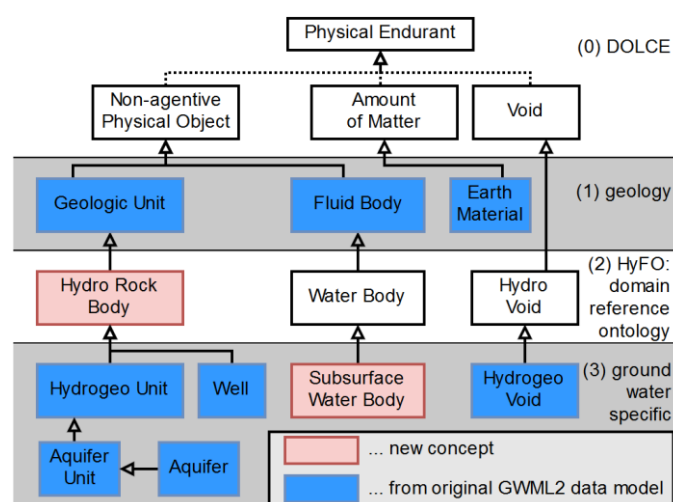


Fig. 1 Excerpt of the subclass hierarchy from the obtained stratified version of GWML2, with groundwater concepts (bottom layer) extending HyFO, general geology concepts and DOLCE.

**Ontological analysis of GWML2** The resulting revised and refined GWML2 ontology reduces barriers to interoperability with other hydro data models. A more concrete contribution is our detailed ontological analysis that clarifies what kind of objects GWML2 terms refer to, fleshing out their spatial, spatio-temporal, material, physical, and ontological characteristics and the relationships (e.g., physical containment, constitution, or spatial parthood) between them. Some represent 3D physical objects (*geologic unit*, *hydrogeo void*, *aquifer*) constituted partly or wholly of solid and/or fluid matter, others denote 2D surfaces (e.g., *water table*), and others purely spatial abstractions (e.g., *monitoring site*). In our ontological analysis we particularly examined borderline cases – more unusual *geologic units*, *aquifers*, *wells*, or *fluid bodies* – that deviate from the typical textbook schemata of water storage in order to test the general applicability of the proposed axioms. We thereby avoid including axiomatic constraints that are true

in typical “textbook” schemata about water storage but that are not true in a more unusual situations and thus should not be encoded as part of the ontology’s axiomatization.

The resulting ontology increases the semantic precision of GWML2 primarily by logically defining groundwater specific concepts (e.g., *subsurface water body* and *hydrogeo void* – the spaces where subsurface water bodies can be located) using a combination of HyFO’s general hydro concepts (e.g., *water body* and *hydro void* – all spaces where water can be located) and spatio-physical relations. Adding precise definitions and classes that are missing from GWML2 (e.g., *container solid body* and *water matter*) also semantically connects classes (e.g., *well*) that were isolated in GWML2’s original model. In addition, our analysis reorganizes GWML2 classes, moving those that are applicable beyond the groundwater domain to layers higher up in the taxonomy (e.g., *fluid body* to the general geology layer or *basin* to the HyFO layer). We specialize these concepts at more specific layer, for example, *water body* and *subsurface water body* are introduced as specializations of *fluid body* at the hydro and the groundwater layer.

**Summary** The following contributions are made: (1) Stratifying GWML2 classes, for a cleaner and more precise organization, and improved reusability and interoperability with other hydro ontologies. (2) Analyzing key GWML2 classes and proposing related axioms to add clarity, rigor, and detail. (3) Identifying a number of revisions to GWML2’s conceptual schema, to better reflect its domain. (4) Recognizing *hydro rock body* as an important concept missing from the Hydro Ontological Square. (5) Completing initial tests that demonstrates HyFO’s potential as a reference ontology for the water domain. More generally, our work exemplifies how a data model or lightweight ontology benefits from grounding in a deeply axiomatized reference domain ontology. Such grounding makes explicit subtle semantic differences between ontologies within a domain and thus enhances their semantic interoperability.

## Acknowledgements

The authors gratefully acknowledge support for this work obtained from the Maine Economic Improvement Fund and the Groundwater Program of Natural Resources Canada.

## References

- Boisvert, E. and Brodaric, B. (2012) GroundWater Markup Language (GWML) – enabling groundwater data interoperability in spatial data infrastructures. *Journal of Hydroinformatics*, 14(1):93–107.
- Brodaric, B. (2015): GroundWaterML2 – GW2IE Final Report. Technical Report Open Geospatial Consortium Engineering Report 15-082, version 2.1
- Brodaric, B. and Gahegan, M. (2010) Ontology use for semantic e-science. *Semantic Web Journal*, 1(1–2):149–153.
- Brodaric, B. and Hahmann, T. (2014) Towards a foundational hydro ontology for water data interoperability. In *Proc. of the 11th Int. Conference on Hydroinformatics (HIC-2014)*.
- Dornblut, I., Atkinson, R. (2013) HY\_Features: a geographic information model for the hydrology domain. Technical Report GRDC 43r1, Global Runoff Data Centre, November 2013.
- Hahmann, T. and Brodaric, B. (2012) The void in hydro ontology. In *Conf. on Formal Ontology in Inf. Systems (FOIS-12)*, IOS Press, 45–58.
- Hahmann, T. and Brodaric, B. (2013) Kinds of full physical containment. In: Tenbrink, T., Stell, J., Galton, A., Wood, Z. (Eds.) *Conference on Spatial Information Theory (COSIT-13)*. Springer, 397–417.
- INSPIRE Thematic Working Group Geology (2013) D2.8.II.4 INSPIRE Data Specification on Geology – Draft Guidelines, v3.0 rc3. Tech. report, INSPIRE, 369pp.
- INSPIRE Thematic Working Group Hydrography (2009) D2.8.I.8 INSPIRE Data Specification on Hydrography – Guidelines, v3.0. Tech. report, INSPIRE, 175pp.
- Kalfoglou, Y. and Schorlemmer, M. (2013) Ontology mapping: the state of the art. *Knowledge Eng. Review*, 18(1):1–31, 2003.
- Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. (2003) *Wonderweb Deliverable D18 – Ontology Library (Final Report)*. Technical report, CRN-ISTC, Trento, 349 pp.
- Noy, N. F. Semantic integration: A survey of ontology-based approaches. *SIGMOD Record*, 33(4):65–70, 2004.