

# GroundWater Markup Language (GWML) – enabling groundwater data interoperability in spatial data infrastructures

Eric Boisvert and Boyan Brodaric

## ABSTRACT

Increasing stress on global groundwater resources is leading to new approaches to the management and delivery of groundwater data. These approaches include the deployment of a Spatial Data Infrastructure (SDI) to enable online data interoperability amongst numerous and heterogeneous data sources. Often an important component of an SDI is a global domain schema, which serves as a central structure for the query and transport of data, but at present there does not exist a schema for groundwater data that is strongly compliant with SDI concepts, standards, and technologies. In this paper we present GroundWater Markup Language (GWML), a groundwater application of the Geography Markup Language (GML). GWML can be used in conjunction with a variety of web services to facilitate data interoperability in a SDI. We describe three common usage scenarios that motivate the design of GWML and a three-stage design methodology involving conceptual, logical and physical schemas. The resultant GWML has broad scope as demonstrated by its implementation in the Canadian Groundwater Information Network. Example uses include decision support in resource management, a scientific application for aquifer mapping, and a commercial application for drill site selection. These demonstrated uses suggest GWML can play a key role in emerging groundwater SDI.

**Key words** | geography markup language, groundwater, spatial data infrastructure

**Eric Boisvert**  
Geological Survey of Canada,  
490 rue de la Couronne,  
Quebec, G1K9A9,  
Canada

**Boyan Brodaric** (corresponding author)  
Geological Survey of Canada,  
234B – 615 Booth St.,  
Ottawa, K1A0E9,  
Canada  
E-mail: [brodaric@nrcan.gc.ca](mailto:brodaric@nrcan.gc.ca)

## INTRODUCTION

As in most parts of the world, Canadian groundwater is facing increasing stress due to growing water use, climate change, and contamination threats. It is abundantly clear that single agencies cannot address all the related challenges, and that intergovernmental cooperation is imperative (Rivera *et al.* 2003; CCA 2009). As part of this realization, recent expert reviews have identified the need for a more consolidated approach to groundwater data access and management (Rivera *et al.* 2003; CCA 2009). Such an approach is needed because Canadian groundwater data is difficult to find and obtain, and then hard to use. Its discovery is impeded because of variable online access and because the data is distributed amongst the private, public and academic sectors. By far the largest portion is collected and managed by several layers of government, particularly at the provincial and watershed

levels. Its use is impeded because each data provider organizes their data differently, and the data quality varies within and between providers. Importantly, this fragmented and heterogeneous state of groundwater data is not unique to Canada but typifies the global situation (United Nations 2008).

An increasingly popular solution to the consolidation problem is the deployment of a Spatial Data Infrastructure (SDI; Masser 2010). SDI concepts, standards and technologies have progressively matured, such that online data networks are now operationally feasible. Such networks exhibit interoperability amongst distributed and heterogeneous data sources, enabling large and complex data stores to be used as a single virtual online system. Data interoperability within SDI occurs at the systems, syntax, schema, semantic and pragmatic levels (Sheth 1999; Brodaric 2006). The systems level addresses

infrastructure heterogeneity concerned mainly with aligning data access protocols, the syntax level addresses the alignment of different data representation languages, the schema level addresses the alignment of diverse data structures, the semantic level addresses the alignment of differences in the meaning of both data content and schema, and the pragmatic level is concerned with aligning the differences in how data is collected and used. Open geospatial standards and technologies, largely driven by the Open Geospatial Consortium (OGC) and the International Standards Organization (ISO), provide solutions for the systems and syntax levels. These include system-level standards for web services, such as the Web Map Service (WMS; [de la Beaujardiere 2004](#)), Web Feature Service (WFS; [Panagiotis 2005](#)), and Sensor Observation Service (SOS; [Na & Priest 2007](#)), as well as the syntax-level standard Geography Markup Language (GML; [Portele 2007](#)). While a general schema-level standard is provided for Observations and Measurements (O&M; [Cox et al. 2007a, b](#)), domain specific solutions at the schema, semantic and pragmatic levels must be developed by domains. This is notably true for the groundwater domain, which lacks comprehensive and strongly OGC compliant data interoperability solutions for the schema and semantic levels.

In this paper we present GroundWater Markup Language (GWML) as a contribution to groundwater data interoperability solutions at the schema level. GWML is a specific application of GML, and can be used as a common query and transport schema in conjunction with standard OGC web services such as WFS and SOS. This allows queries to be issued against GWML to distributed data sources, and data to be delivered from the sources using the GWML structure, in SDI environments. Because groundwater is greatly constrained by geology, GWML is also developed as an extension to GeoSciML ([Sen & Duffy 2005](#)), a GML application for geology. As observational data play a key role in hydrogeology, GWML also extends Observation and Measurements and related sampling specifications ([Cox et al. 2007a, b](#)). To demonstrate operational feasibility, we show how GWML is used within the Groundwater Information Network (GIN), an emerging Canadian groundwater SDI. The remainder of the paper describes typical usage scenarios for GWML, related work and the methods used to develop GWML, the GWML structure as well as its GIN implementation. The paper concludes with some final comments and future directions.

## USAGE SCENARIOS

Groundwater and related data are vital to every sector of society. For example, about 25% of Canada's population relies on groundwater as a key source of water ([Statistics Canada 2010](#)). Schematic representations of groundwater data must therefore be sufficiently broad and deep to accommodate a diversity of usage situations. For the purposes of this paper we describe three typical usage scenarios that inform the design of GWML: a commercial scenario, a scientific scenario, and a water resource management scenario. Each scenario identifies critical data that must be represented in the GWML structure to enable its query and processing in a SDI. The scenarios mainly involve water well and compiled aquifer data, as these are a primary source of groundwater information; a complete accounting would also include scenarios for water budgets and other data relevant to hydrogeologic systems and their interface with surface water. A key subtext to these scenarios is the significance of GWML: prior to GWML either the required groundwater data could not be obtained in an OGC-GML compliant schema, or it could not be obtained in a single common schema because each source would serve the data using a different structure, creating significant barriers to the efficient scientific processing of the data.

### **Commercial scenario: Water well driller assesses local geology prior to bidding on a contract to drill a new well**

A driller zooms into an area of interest using an online portal, sequentially selects wells near the location of the planned well, and observes the rock material logs and reported water levels for each selected well. From this information, the driller can assess the nature and the thickness of geological formations that must be drilled before reaching the water table, which will enable an accurate cost estimate to be made for the drilling.

This is the simplest case where the only required information consists of rock material logs and water levels from wells within a specified geographic area.

### **Science scenario: Scientist delineates an aquifer's boundaries**

A scientist zooms into an area of interest using an online portal and requests a geologic map, some water wells, and

draws a line for a cross-section. The wells are plotted at the correct elevation and correct relative distance along the cross section path. Each well is selected, and the water level as well as a full description of the rock material is displayed. This display enables a geologist to begin identifying the physical characteristics and the probable limits of the aquifer, by inferring associated hydrogeologic properties of the rock materials, such as porosity and permeability, and interpolating boundaries between the materials using these properties.

This scenario expands the scope of the schema by additionally requiring aquifer information in the GWML schema.

### Water manager scenario: Emergency response team assesses a spill

A water manager zooms to the area of spillage using an online portal. Relevant aquifers and all the wells with the following characteristics are selected: the porous material above a certain threshold, the conductivity above a certain threshold, the associated rocks located below the water table, and the wells used to extract drinking water. From this information the manager can begin to determine the possible physical susceptibility of the aquifer to contamination, and begin to get an idea of which areas have potential for contaminant, to gauge the possible threat to humans.

This last scenario demonstrates a complex query where water wells in an area are selected using thematic criteria related to porosity, conductivity, water depth, and well usage. The scenario requires these key hydrogeological properties to be available in a format and unit of measure that can be queried seamlessly.

---

## RELATED WORK

To the best of our knowledge GWML is the first widely available GML application for groundwater data that is strongly aligned with OGC standards, thus ensuring early compatibility with emerging SDI environments for water. Related work includes the development of data structures for groundwater databases, groundwater ontologies, water related markup languages, and geology related markup languages.

## Groundwater database schemas

Databases containing groundwater data are abundant because their creation and maintenance is often mandated by governmental water management agencies. However, each database is designed to meet specific business requirements resulting in heterogeneous schemas (e.g. [Gogu \*et al.\* 2001](#); [Wojda 2008](#); [Oulidi \*et al.\* 2009](#)). To overcome this issue the Arc Hydro Groundwater data model is a standard database schema designed specifically for the ESRI ArcGIS environment ([Strassberg \*et al.\* 2007](#)), but at present there exists no standardized GML serialization of this schema, as variations proliferate with each implementation. This lack of standard GML serialization is also true for schemas designed to assist groundwater data transfer (e.g. [Federal-Provincial Working Group on Groundwater 1993](#); [National Groundwater Committee 1999](#)).

## Groundwater ontologies

Formal ontologies that contain groundwater elements include Semantic Web for Earth and Environmental Terminology (SWEET) ([Raskin & Pan 2005](#)), its groundwater extension ([Tripathi & Babaie 2007](#)), and DOLCE ROCKS ([Brodaric & Probst 2009](#)), which are expressed using the OWL-DL representation language. [Yang \*et al.\* \(2010\)](#) present an informal ontology for aquifers expressed using an UML-like diagram. These variably formal ontologies are concerned with the semantic level and particularly with the representation of the meaning of groundwater concepts. In data interoperability environments these concepts are typically associated with terms such as the labels of elements in groundwater data schemas and the vocabulary used to populate such schemas. Although ontologies could be used to transfer groundwater data, e.g. in instances of the OWL-DL concepts, such representation is not immediately compatible with SDI standards requiring GML, the approach is not supported by widely available OGC web service implementations, and the representation of numerics is typically cumbersome. These ontologies also contain only a relatively limited suite of groundwater concepts that need to be extended to satisfy the usage scenarios above: e.g. SWEET, DOLCE ROCKS, and Yang's do not explicitly represent water wells.

## Water related markup languages

There exist numerous XML-based schema for data primarily related to water measurements. These include WDTF, WQX, WaterML 1, X Hydro, UK-EA-TS, amongst others, and they are compared for unification under the OGC compliant WaterML 2.0 standards initiative by Taylor (2009). While parts of these schemas might be re-used to represent groundwater measurements related to groundwater levels, flows, and quality, they do not represent key groundwater features such as water wells and aquifers, and most are not fully compatible with OGC services. A promising exception is the schema from Oulidi *et al.* (2009), which does function with OGC services, but has limited scope, e.g. it does not model items such as observations and water flows, and it does not re-use other international standards such as GeoSciML.

## Geology related markup languages

Geological aspects must be included in groundwater data schema because physical groundwater systems are dependent on host geological frameworks. Emerging GML applications for geology show great benefit for geologic data interoperability (Lake 2005), with the leading candidate being GeoSciML (Sen & Duffy 2005), a development of the International Union of Geological Sciences. Although GeoSciML does not contain groundwater features such as water wells and aquifers, nor related properties such as water storage capacity, it does provide many of the important geological components required by a groundwater schema, such as geological formations and their constituent earth materials. GeoSciML also provides some hydrogeological properties such as porosity and permeability, requiring a groundwater schema to be carefully integrated with such geological efforts.

## METHOD – DEVELOPING GWML

The development of GWML is informed by a requirements analysis carried out as part of a wide review of existing groundwater information systems. The review uncovered both scientific and social aspects of groundwater data. The scientific aspects encapsulate basic data about the physical and chemical environment, while the social aspects encompass groundwater vulnerability, demand, and use. We focus on the scientific

aspects of groundwater data in this initial version of GWML, with the social aspects left to future work. The design itself follows three phases, each creating a distinct schema, analogous to the steps and products inherent in the design of a database schema: conceptual, logical, and physical.

## Conceptual schema design

The design of GWML includes development of a technology neutral conceptual schema, in which concepts and relationships are extracted from the requirements analysis and expressed as formally as possible. Feedback was elicited from hydrogeologists at key stages to ensure the accuracy of the conceptual model and its compatibility with the requirements analysis. Although no formalism is completely technology neutral, we primarily utilize UML classes and associations to describe some of the domain concepts and relations, primarily to capture ideas in interim working documents. UML was chosen because of its wide acceptance, broad applicability, and moderate expressivity. Figure 1 illustrates a fragment of the GWML conceptual schema in UML, in which hydrogeologic rock bodies are shown to have the capacity to host a reservoir for water. Many GWML concepts are in addition described informally in text, as well as formally in an ontology that is expressed using the OWL-DL representation language. Formal ontologies are an alternative representation for the conceptual schema, one that provides a rigorous description that can be directly used for reasoning in data interoperability environments. In practice the conceptual schema is often not developed, or if developed it is frequently not formally represented, but our experience with GWML suggests these omissions lengthen the overall development process as it leads to more iterations in the design of the next schema, the logical schema, due mainly to inadequate initial conceptualization.

## Logical schema design

The conceptual schema is manually converted into a logical schema that contains technological artefacts. These artefacts

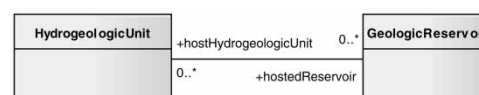


Figure 1 | GWML conceptual schema fragment in UML.

include the GML superstructure with which the conceptual schema must be aligned. Alignment primarily involves the assignment of a GML meta-class to each domain class, and secondarily the optional specialization of a domain class from an existing GML schema such as GeoSciML. We represent the logical schema in UML and use the FullMoon tool (<http://projects.arcs.org.au/trac/fullmoon/>) to automatically generate the next schema, the physical schema, after certain design patterns are followed (ISO 2006). Note that in our notation below we use fully qualified names, where a prefix refers to domain namespaces: *gml* refers to GML classes or meta-classes, *gwml* refers to GWML, *gsml* refers to GeoSciML, *om* refers to OGC Observations and Measurements (O&M; Cox 2007a), and *sa* refers to OGC Sampling Features (Cox 2007b). GML meta-classes become UML stereotypes in the logical schema, with the most important of these being <<FeatureType>>, <<Type>> and <<DataType>>. <<FeatureType>> refers to OGC features which we interpret to be entities that have identity, are typically located in geographic space, and have physical unity, such as water wells or aquifers. <<Type>> refers to entities that have identity and which we interpret as not necessarily having physical unity such as an amount of material, e.g. sand or water. <<DataType>> typically packages a group of properties, such as temperature or volume. Each logical class must contain a GML meta-class stereotype that is directly assigned, or which is inherited from

a more general parent class possibly from an external schema. For example, *gwml:HydroGeologicUnit* specializes the *gsml:GeologicUnit* feature type from GeoSciML, *gwml:WaterWell* specializes the *sa:SamplingPoint* feature type from OGC Sampling Features, and *gwml:WaterLevel* specializes the *om:Observation* feature type from OGC Observations and Measurements. Additional design guidelines include the requirement for each traversable association endpoint to be named and numbered, to enable correct nesting and sequencing of class attributes in a GML schema. Figure 2 illustrates a fragment of the GWML logical schema, in which GML stereotypes are illustrated with double angle brackets.

### Physical schema design

The physical schema denotes the final GWML schema. It contains the elements and constraints required for validation as a GML application, meaning it complies with W3C XSD and GML schema design requirements and principles (ISO 2006, 2007a, b), such that it can be used with OGC web services. In addition it must comply with requirements set by parent schemas such as GeoSciML. Compliance with these various requirements can be assured through careful deployment of tools for the automatic generation of a GML schema. The following code shows a fragment of the GWML schema for the hydrogeologic unit class.

```
<element name="HydrogeologicUnit" type="gwml:HydrogeologicUnitType" abstract="true"
substitutionGroup="gsml:GeologicUnit">
  <annotation>
    <documentation>Means any soil or rock unit or zone which by virtue of its porosity or permeability, or
lack thereof, has a distinct influence on the storage or movement of groundwater. (EPA). Any soil or rock unit or zone that because of
its hydraulic properties has a distinct influence on the storage or movement of ground water.
http://water.nv.gov/WaterPlanning/dict-1/PDFs/wwords-h.pdf</documentation>
  </annotation>
</element>
<complexType name="HydrogeologicUnitType" abstract="true">
  <complexContent>
    <extension base="gsml:GeologicUnitType">
      <sequence>
        <element name="waterQuantity"
type="gwml:WaterQuantityDescriptionPropertyType" minOccurs="0" maxOccurs="unbounded" />
        <element name="waterContent"
type="gwml:GroundwaterBodyPropertyType" minOccurs="0" maxOccurs="unbounded" />
        <element name="relatedReservoir"
type="gwml:GeologicReservoirPropertyType" minOccurs="0" maxOccurs="unbounded" />
        <element name="hydrogeologicClass"
type="gwml:CategorisationPropertyType" minOccurs="0" maxOccurs="unbounded" />
      </sequence>
    </extension>
  </complexContent>
</complexType>
```

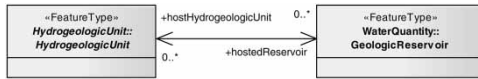


Figure 2 | GWML logical schema fragment in UML.

## Evaluation criteria

Criteria for evaluating the final GWML schema include validation, completeness, precision, usability, and efficiency. Validation refers to compliance with XSD/GML schema requirements; completeness refers to a measure of the adequacy of the breadth and depth of the hydrogeological concepts in the schema; precision refers to the degree of alignment with the conceptual schema (do the elements in the physical schema accurately reflect the intended meanings in the conceptual model?). Usability refers to the operational feasibility of the physical schema, and efficiency refers to how proficiently it can be used – can it be adequately implemented in operational systems? Evaluation of GWML is included below as part of the implementation description.

## RESULTS – GWML

The parts of the GWML schema that are related exclusively to the natural environment are designed around four water feature concepts, as shown in Figure 3: (1) a cavity in a physical object that can contain a body of water, (2) a body of water occupying part of the cavity, (3) the physical object hosting the cavity, and (4) the materials constituting the host physical object. Note these water feature concepts are quite general and could be applied, for example, to the representation of features such as lakes or aquifers.

The application of the general water feature concepts to groundwater goes as follows: the cavity typically consists of the aggregation of spaces within and between some earth materials, including the pores within consolidated rock, the openings and discontinuities between consolidated

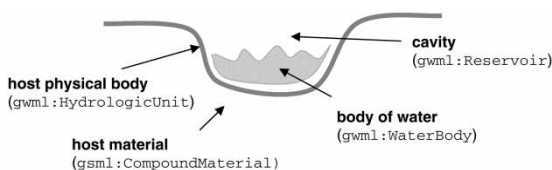


Figure 3 | Key water concepts.

rock bodies, as well as the spaces between the grains in unconsolidated materials such as sand or soil. A hydrogeologic unit, such as an aquifer, is the physical body that hosts the cavity and which is constituted by the earth materials. A groundwater body is then a discrete aggregated object contained in the spaces and composed of possibly several amounts of water and other materials. Although this conceptualization aligns with conventional definitions in which an aquifer is an earth material body capable of yielding obtainable water (Neuendorf *et al.* 2005), it is generally more specific about the various components and their interrelations, which commonly occurs with formal representations. Figure 4 illustrates the key water feature concepts applied to groundwater and Figure 5 shows the related GWML representation.

GWML also incorporates two additional water concepts related to human interaction with the natural environment: (5) hydrogeologic observation and measurement, which refers to the observation of a value for a specific hydrogeologic property at some time point, and (6) water well, which includes well characteristics, construction data, pumping data, and links to earth material logs taken along the length of the well. The six main components of the GWML schema are elaborated in the sections below. The components are illustrated with UML diagrams, while sample XML encodings are available at the GWML web site ([http://ngwd-bdnes.cits.rncan.gc.ca/service/api\\_ngwds/en/gwml.html](http://ngwd-bdnes.cits.rncan.gc.ca/service/api_ngwds/en/gwml.html)).

## Water reservoir

Although GWML focuses on geologic reservoirs, for purposes of completeness it also delineates other types of reservoirs, such as surface, biologic and atmospheric. A lake is an example of a surface reservoir hosted by a landscape depression – lakes are not included in GWML as they are deemed to better fit into a surface water schema. Other critical aspects of a reservoir modeled in GWML include

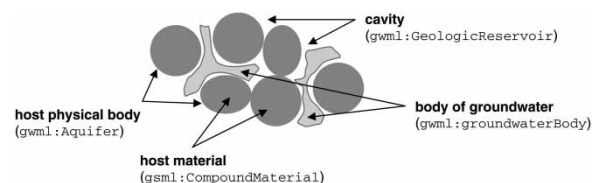


Figure 4 | Key water concepts applied to groundwater.

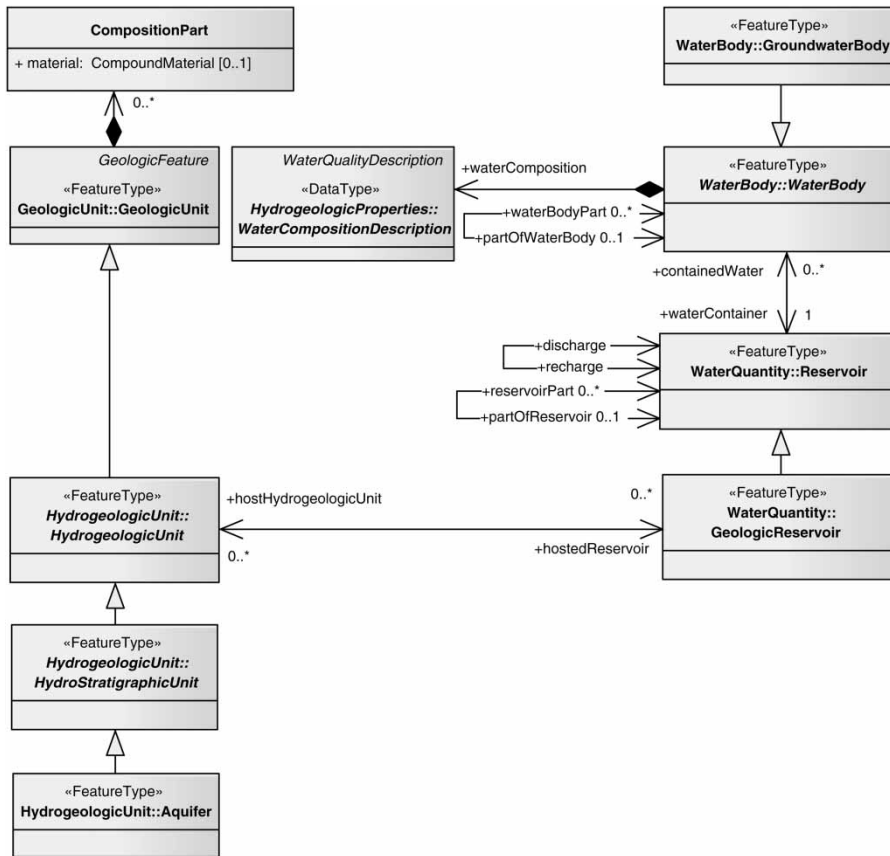


Figure 5 | GWML logical schema fragment denoting key water feature concepts.

elements for water accounting, such as inflow, outflow, and potential capacity, as well as adjoining hydraulic boundaries that constrain the flow of water. Flows are vital to recharge and discharge estimates and depend on the target and source reservoirs. For instance, evapotranspiration is the transfer of water from the biomass and soil to the atmosphere, and base flow is the transfer from a groundwater reservoir to a surface reservoir. Figure 6 depicts the GWML schema for reservoirs and their water flows.

## Water body

A water body is an amount of material composed mainly of H<sub>2</sub>O and some amounts of other materials in various states of solution and suspension, such as salts or sediments, or biologic components such as micro-organisms. A water body can retain identity even though its composition can change, for example, because of shifting water flow or pollution. Thus we can refer to a single groundwater body in an aquifer, even

though water flow might cause some of its components to be replaced over time and cause its volume to fluctuate. Water bodies are necessarily contained by a reservoir, which determines the water body's shape and connectivity, and they can have parts whose geospatial boundaries are delineated for example by temperature stratifications, salinity pools, and contamination plumes. Water body properties include structural characteristics, such as volume, thickness, and depth from surface, and compositional characteristics, such as degree of salinity and the phases of solution of constituents, as shown in Figure 7. Each compositional component is also identified as a type of constituent, such as arsenic or coliform, through reference to a user-supplied vocabulary.

## Hydrogeologic unit

A hydrogeologic unit is the principle physical body associated with groundwater systems. It consists of any package of consolidated rock, unconsolidated material, or their

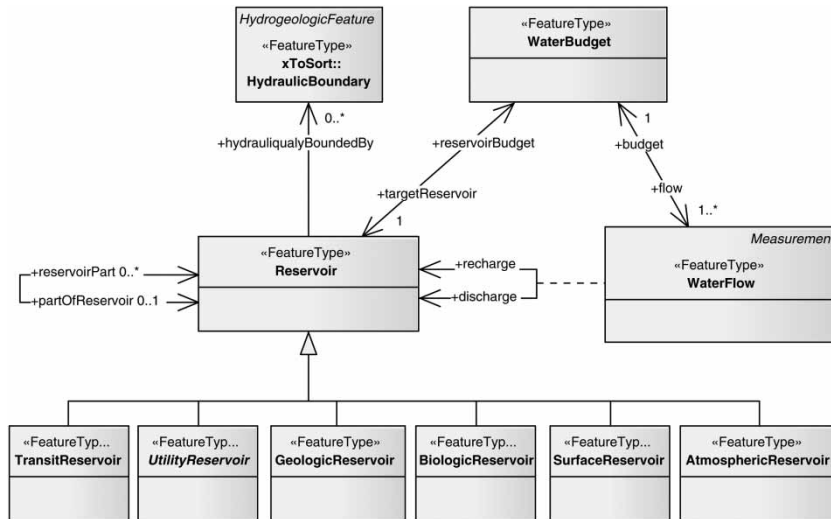


Figure 6 | GWML logical schema fragment for reservoirs and related water flows.

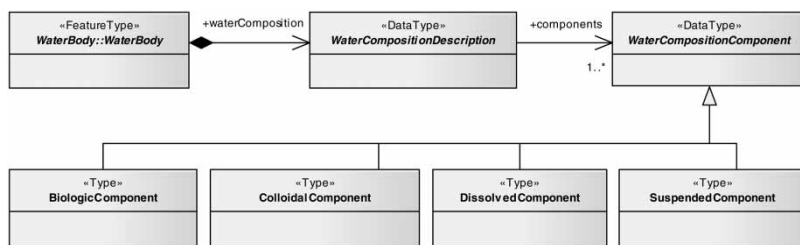


Figure 7 | GWML logical schema for water body and its composition.

mixture, that hosts reservoirs capable of storing groundwater and/or influencing its flow – as such it is a special type of geologic unit that possesses water-related properties. As shown in Figure 8, GWML explicitly recognizes this by specializing hydrogeologic unit from GeoSciML’s geologic unit. This enables a hydrogeologic unit to be fully described in terms of its geologic parts (e.g. formations), earth material composition (e.g. rock materials), and related properties. These interrelations imply the GWML schema is formally dependent on the GeoSciML schema – i.e. GWML cannot exist independently. Hydrogeologic units are further delineated into hydrostratigraphic units, which refer to geological formations capable of storing groundwater and/or impacting flow, and aquifer systems, which refer to an aggregation of hydrostratigraphic units. Hydrostratigraphic units are delineated into aquifer-related entities that can supply water to wells (`gwml:Aquifer`), impede water flow (`gwml:Aquiclude`), or permit low flow between aquifer

systems generally insufficient for drawing water into wells (`gwml:Aquitard`). Aquifers may be surrounded by impermeable barriers (`gwml:ConfinedAquifer`) or not (`gwml:UnconfinedAquifer`).

### Hydrogeologic unit materials

Because a hydrogeologic unit is modeled as type of geologic unit in GWML, its material composition is inherited from GeoSciML, enabling for example GWML aquifers to be described in terms of GeoSciML earth materials. This includes the inheritance of key geologic earth material properties, and highlights the need to add key hydrogeologic properties. The earth material properties supplied by GeoSciML that are particularly relevant to GWML include permeability and porosity, while the hydrogeologic earth material properties added by GWML include capacity, storativity, and conductivity.



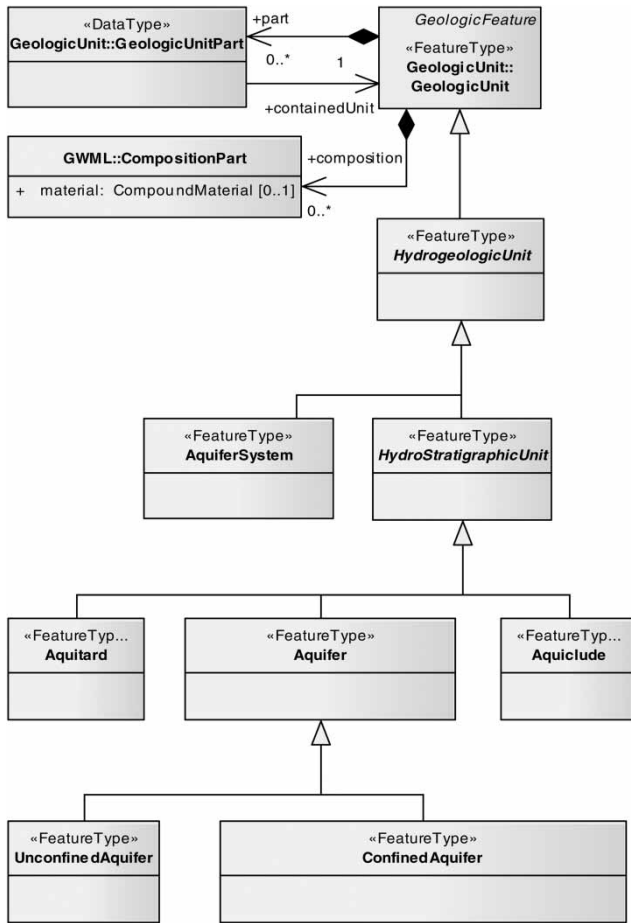


Figure 8 | GWML logical schema fragment for aquifers.

## Hydrogeologic observations and measurements

Groundwater systems are dynamic and adjust to short-term and long-term changes in the natural and social environments. Measurements taken by sensors in water wells are the principle source of data about the stresses acting on the systems. GWML provides classes for the measurement of groundwater flows, directions and levels, as specializations of `om:Observation`. Following the OGC Observations and Measurements schema, these classes refer to a suite of information about how a particular item of groundwater data is collected, including the collection procedure (e.g. via a gauge or even a human observer), the site (e.g. a water well), the property measured (e.g. water level), the resultant value (e.g. 5 m) as well as the targeted feature hosting the property (e.g. an aquifer). Such

observational data can be transmitted online using the OGC Sensor Observation Service as part of emerging sensor networks. It is anticipated this portion of the GWML schema will be replaced by the emerging international WaterML 2 standard (Taylor 2009).

## Water well

Water wells are significant not only because of their scientific role in the provision of sites for groundwater monitoring, but also for their social role as a key source of usable water. The water well concept in GWML is quite broad, encompassing any hole in the ground that is capable of yielding usable water. Data associated with water wells include elements pertaining to scientific investigation, such as logs of the mapped intervals of earth materials found along the length of the well, as well as elements pertaining to the use of the groundwater, such as data about its construction materials and processes, pumping, environmental protection, and administration. GWML spans the breadth of these data types, enabling a thorough description of water wells. The starting point for this description is `gwml:WaterWell`, which is a specialization of `sa:SamplingPoint`, meaning that a water well is interpreted to be a site where certain observations are taken and activities occur, emphasising the central role of the water well as a provider of groundwater data. Figure 9 displays some core elements of the GWML water well logical schema.

## IMPLEMENTATION OF GWML – THE GROUNDWATER INFORMATION NETWORK

GWML is the standard data schema for the Groundwater Information Network (GIN, [www.gw-info.net](http://www.gw-info.net)), an SDI for Canadian groundwater data. GWML is used to transport groundwater data between web clients that request data from GIN and web sources that serve data to GIN. Participating in GIN are one federal, one territorial, and several provincial data providers, who are collaborating to enable uniform online access to their distributed and heterogeneous data holdings. At present, the data holdings are primarily water well databases with associated web services. Web clients send queries using

GWML to central GIN web services (WFS, WMS, SOS), collectively called the GIN mediator, and data is retrieved from the distributed web services (via some of WFS, WMS, SOS) and returned to the clients also using GWML. The central GIN mediator provides translation to GWML and integrates the data into a single result. Translation involves both schematic and semantic alignment of both queries and data results, from/to the standard structure and content of GWML. The main item of standard content associated with GWML is a common vocabulary for rock types, a subset of the GeoSciML vocabulary. If translation from/to GWML is instead provided at the data

source, then the GIN mediator simply passes the queries to the data sources while continuing to integrate the results into a single response that is returned to the client. The GIN architecture enables the distributed and heterogeneous data sources to be used as a unified and normalized data repository. This encourages the development of clients that can be used to carry out the three usage scenarios.

### Commercial scenario – viewing water well logs

Satisfaction of the commercial usage scenario involves implementation of GWML for the viewing of water well

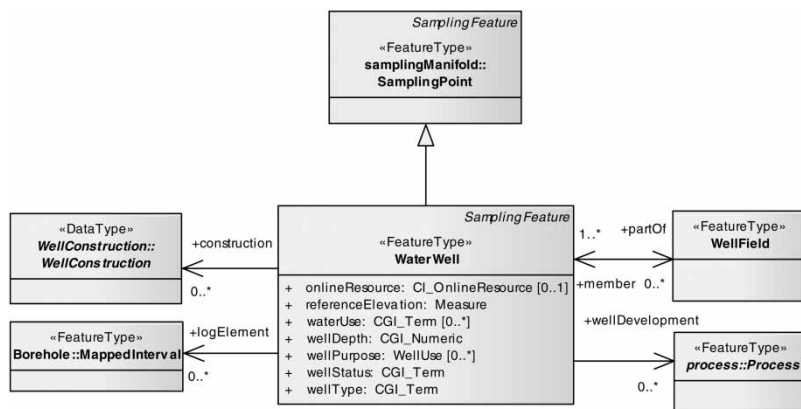


Figure 9 | GWML logical schema fragment for water wells.

Identity : ca.ab.gov.w\_well  
 Provincial id :  
 Source : Manitoba Water Stewardship  
 Online resource : [Manitoba Water Well Info](#)  
 Date of drilling : 1986-06-01  
 Length : 1.22m  
 Elevation (Google Maps) : 349m  
 Water use : Missing

DEPTH FROM (M)	DEPTH TO (M)	GIN LITHOLOGY	ORIGINAL LITHOLOGY	POROSITY*	HYDRAULIC CONDUCTIVITY*
0.00	0.30	Diamicton <a href="#">(More Info)</a>	TILL/DIAMICT		
0.30	1.07	Gravel <a href="#">(More Info)</a>	BOULDER(S)	[24,44]%	[3E-4,3E-2]m/s
1.07	1.22	Bedrock <a href="#">(More Info)</a>	UNDIFFERENTIATED ROCK		

Well log. (\*Notice: Values taken from literature)

Identity : ca.on.gov.w\_well  
 Provincial id :  
 Source : Ontario Ministry of Northern Development and Mines - Ontario Geological Survey  
 Online resource : [Ontario Borehole Database](#)  
 Date of drilling : 1970-02-27  
 Length : 31.09m  
 Elevation : 304.80m  
 Elevation (Google Maps) : 320m  
 Water level : 13.72m 1970-02-27  
 Water segment : 22.86, FRESH 1970-02-27  
 Water segment : 31.0896, FRESH 1970-02-27  
 Water yield : 4.55lpm  
 Water use : Domestic

DEPTH FROM (M)	DEPTH TO (M)	GIN LITHOLOGY	ORIGINAL LITHOLOGY	POROSITY*	HYDRAULIC CONDUCTIVITY*
0.00	4.57	Clay <a href="#">(More Info)</a>	CLAY	[34,57]%	[1E-11,4.7E-9]m/s
4.57	31.09	Granite <a href="#">(More Info)</a>	GRANITE		[3E-14,2E-10]m/s

Well log. (\*Notice: Values taken from literature)

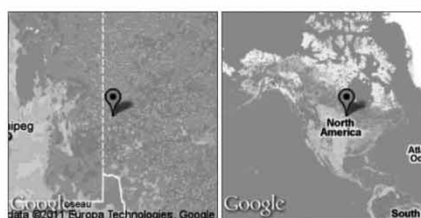
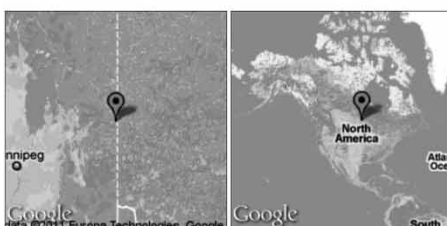


Figure 10 | Viewing water well records across provincial boundaries in GIN.

records, particularly related water levels and rock material logs. Figure 10 illustrates two water well records that are found and retrieved using a GIN web client. In the example, wells are selected from adjacent sides of the border of two Canadian provinces, Ontario and Manitoba, and their records are dynamically extracted from the corresponding web services. The water well records are returned to the client in the standard GWML structure using the common rock type vocabulary. The client transforms the GWML into a web page that is then presented to the end-user, enabling interpretations to be made about the nature of the geology and the water table in the selected area. This impacts decisions about whether to drill and where to drill.

### Science case study

Satisfaction of the science scenario requires additional information in GWML about wells and aquifers. This information can be used to produce cross-sections and estimate water flows, which helps delineate aquifer boundaries. For example, Figure 11(a) illustrates a cross-section view including water wells and a rock layer, generated by an online tool using GIN, and Figure 11(b) shows the information returned about the related aquifer, including the rock layer displayed in the cross-section.

### Water management case study

Satisfaction of the water management usage scenario requires use of GWML in a complex query. The query seeks water wells meeting thematic conditions in a geographic area. The conditions involve the rock material's porosity and conductivity, the depth of the water table, and the intended use of the well for extracting drinking water, as well as the geographic area. The query is specified in SQL-like pseudo-code as:

```
SELECT gwml:WaterWell WHERE
gwml:logElement/.../gsml:PhysicalDescription/gsml:porosity >= 34 AND
gwml:logElement/.../gwml:HydrogeologicDescription/gwml:hydraulicConductivity > 1E-11 AND
gwml:wellPurpose = "Domestic" AND
BELOW (gwml:logElement/.../gml:shape, gwml:staticWaterLevel) AND
BBOX (-111, 54 -110, 55)
```

The query is operationally stated as an OGC filter expression as shown below. It is moderately complex because the path from a water well to related hydrogeologic

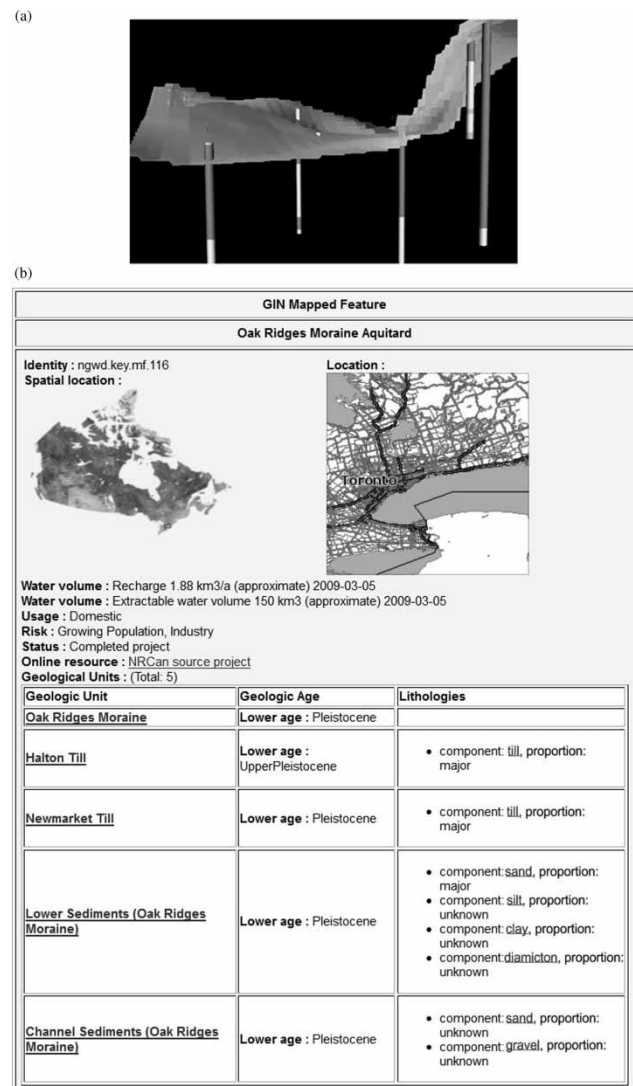


Figure 11 | (a) Cross-section view with water wells and a rock layer from GIN, (b) aquifer details in GIN.

properties is long and involves the traversal of multiple elements of the GWML schema – this complexity typically requires customization such as the BELOW function, which tests whether a rock segment in the well log is below the water level. The filter also relies on XPath (Clark & DeRose 1999) for scoping support, which is an advanced option being considered in pending WFS revisions. Without full XPath support the standard filter query of WFS 1.x, which only supports a subset of XPath, would return incorrect results because the query conditions would be incorrectly scoped. The XPath solution below correctly scopes the query so that every returned well meets *all* the thematic conditions. It can be sent to the GIN mediator in a WFS GetFeature request, which will invoke: (1) query

processors customized for each data source to enable the retrieval of wells from GIN, and (2) post-processors that ensure the conditions are fully met.

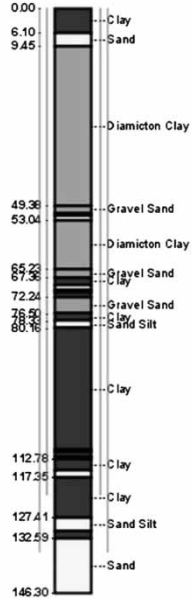
Figure 12 shows part of a water well record satisfying the query as returned by GIN. This demonstrates not only aspects of the range of the GIN structure, but also the feasibility of its application for complex data retrieval operations. Note at present the GIN portals do not provide a graphical user interface for the construction of such queries, but the GIN mediator will receive and process them. By examining various records such as the one shown in Figure 12, a water manager can begin to determine the areas of physical susceptibility of in a geographic region, which can help inform a response to some spill.

```
<wfs:GetFeature>
  <wfs:Query typeName="gwml:WaterWell">
    <ogc:Filter>
      <ogc:And>
        <ogc:PropertyIsGreaterThan>
          <ogc:Function name="gin:countMatchingNodes">

<ogc:PropertyName>gwml:logElement/gsml:MappedInterval[gin:below(gsml:sha
pe,
../../sa:relatedObservation/om:Observation[om:observedProperty/@xlink:hr
ef='urn:ogc:def:phenomenon:GIN:StaticWaterLevel']/om:result
)]/gsml:speci
fication/*/gsml:composition/gsml:CompositionPart/gsml:material/*/gsml:ph
ysicalProperty[gin:compareCgiValue(* /gwml:hydraulicConductivity, 'ge', 1e-
11) and
gin:compareCgiValue(* /gsml:porosity, 'ge', 34.0)]</ogc:Property
Name>
          </ogc:Function>
          <ogc:Literal>0</ogc:Literal>
        </ogc:PropertyIsGreaterThan>
        <ogc:PropertyIsEqualTo>
          <ogc:PropertyName>gwml:wellPurpose</ogc:PropertyName>
          <ogc:Literal>Domestic</ogc:Literal>
        </ogc:PropertyIsEqualTo>
        <ogc:BBOX>
          <ogc:PropertyName>sa:position</ogc:PropertyName>
          <gml:Box srsName="EPSG:4326">
            <gml:coordinates>-111,54 -110,55</gml:coordinates>
          </gml:Box>
        </ogc:BBOX>
      </ogc:And>
    </ogc:Filter>
  </wfs:Query>
</wfs:GetFeature>
```

**GIN Water Well Log**

Report an error



Source : Alberta Environment  
 Online resource : [Alberta Water Well Database](#)  
 Date of drilling :2000-08-26  
 Length : 146.30m  
 Elevation : Missing  
 Elevation (Google Maps) : 652m  
 Water level: 64.62m 2000-08-26  
 Water use : Domestic  
**Well casings :**  
 • from 0 to 136.25m Steel  
**Sealing components :**  
 • from 0 to 121.92m Cement/Grout

Well log. (\*Notice: Values taken from literature)

DEPTH FROM (M)	DEPTH TO (M)	GIN LITHOLOGY	ORIGINAL LITHOLOGY	POROSITY*	HYDRAULIC CONDUCTIVITY*
0.00	6.10	Clay (More Info)	Clay	[34,57]%	[1E-11,4.7E-9]m/s
6.10	9.45	Sand (More Info)	Sand	[26,53]%	[2E-7,6E-3]m/s
9.45	49.38	Diamicton (More Info) Clay (More Info)	Till and Clay	[34,57]%	[1E-11,4.7E-9]m/s
49.38	51.21	Gravel (More Info) Sand (More Info)	Sand and Gravel	[24,44]%	[3E-4,3E-2]m/s [2E-7,6E-3]m/s
51.21	51.82	Clay (More Info)	Clay	[34,57]%	[1E-11,4.7E-9]m/s
51.82	53.04	Sand (More Info)	Sand	[26,53]%	[2E-7,6E-3]m/s
53.04	65.23	Diamicton (More Info) Clay (More Info)	Till and Clay	[34,57]%	[1E-11,4.7E-9]m/s
65.23	67.36	Gravel (More Info) Sand (More Info)	Sand and Gravel	[24,44]%	[3E-4,3E-2]m/s [2E-7,6E-3]m/s
67.36	69.19	Clay (More Info)	Clay	[34,57]%	[1E-11,4.7E-9]m/s
69.19	70.71	Sand (More Info)	Sand	[26,53]%	[2E-7,6E-3]m/s

**Figure 12** | Water well record fragment, returned from a complex thematic query.

**GWML evaluation**

The GWML schema measures favourably against the evaluation criteria introduced above. It validates syntactically as a compliant XML schema (XSD) using commercial tools (XML Spy, [www.altova.com/xmlspy](http://www.altova.com/xmlspy)). Its generation using the Full Moon tool ensures full GML schema (XSD) compliance, and it has been successfully used with standard OGC services. Its degree of completeness is demonstrated by the satisfaction of the common usage scenarios, which further illustrate its wide applicability. Precision is difficult to determine accurately without direct comparison to a complete formal representation of the conceptual schema, however it might be inferred to some degree by adequate satisfaction of the usage scenarios and by the associated lack of overt conceptual conflicts. Usability is also demonstrated through satisfaction of the usage scenarios, and by the operational presence of clients that use the schema. The relatively high query complexity, due to the high schema complexity and associated limitations

in WFS 1.x querying, remains a concern as it requires sophisticated XML query languages and processors to be implemented. This might be alleviated in the next revision of WFS which will likely provide mechanisms for querying complex schemas. An alternative solution is the development of a simplified and highly constrained schema for query purposes, i.e. GWML-Lite. Finally, GWML is shown to be efficiently implemented in operational data interoperability systems: [Table 1](#) shows the uncompressed file sizes and times associated with the query, download, and integration of water well data within a geographic area from two GIN data providers.

**Table 1** | Performance measure of GWML download from two GIN data sources

Number of features	Time (sec)	Size (Mb)
50	2.17	1.08
500	15.01	7.74
2500	69.97	40.80
5000	142.27	80.41

The tests were performed on a Xeon Dual Core (2.80 Ghz) server with 3 Gb of memory, running Windows Server 2003 (SP2). GIN modules are written in Java (JDK 1.5) running under Tomcat 5.5 and Apache Cocoon 2.1.9. The data are provided using various WFS backend technologies, including GeoServer over SHP files, ArcIMS over Oracle, and Deegree over PostGIS/PostgreSQL. The time was measured at the client end, which was connected online to the services located in a different city. Planned enhancements to the mediator should further improve this performance.

## CONCLUSIONS

This work on GWML has shown that it is feasible to design and implement a schema for groundwater data that can be effectively and efficiently used in SDI settings, such as in the Canadian Groundwater Information Network. GWML is an integral component of such SDI, enabling data interoperability at the schema level by serving as a global schema for both query requests and data results. GWML has sufficient breadth and depth to be useful in a wide variety of groundwater applications, as demonstrated by its deployment in usage scenarios for commerce, science, and resource management. Because of this design it can potentially be re-used in other emerging groundwater SDI. Future work involves both schema design and implementation aspects. The schema needs to be harmonized with emerging international standards for water observation data (Taylor 2009), including water quality data, and with future surface water schema. It also needs to be extended to include social use of groundwater, and additional mechanisms for stating and processing complex queries need to be refined and tested. The GWML specifications are freely available online 'as is', including the physical schema, logical UML schema, and related documentation: [http://ngwd-bdnes.cits.rncan.gc.ca/service/api\\_ngwds/en/gwml.html](http://ngwd-bdnes.cits.rncan.gc.ca/service/api_ngwds/en/gwml.html).

## ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the Groundwater GeoSciences Program, of the Geological Survey of Canada, and the GeoConnections Secretariat in the development of

GWML. We are also indebted to Piotr Wojda (JRC), Peter Taylor (CSIRO), Nathaniel Booth (USGS), Alex Smirnoff (GSC), Heryk Julien (GSC), and François Létourneau (INRS) for their valuable insights, as well as to Harold Vigneault (GSC), Simon Cox (CSIRO), and two anonymous reviewers for their helpful comments.

## REFERENCES

- Brodaric, B. 2006 *Geo-pragmatics for the Geo-spatial Semantic Web*. *Transactions in GIS* 11 (3), 453–477.
- Brodaric, B. & Probst, F. 2009 *Enabling cross-disciplinary e-Science by Integrating Geoscience Ontologies with DOLCE*. *IEEE Intelligent Systems* 24 (1), 66–77.
- Canadian Council of Academies, Expert Panel on Groundwater 2009 *The Sustainable Management of Groundwater in Canada*. Canadian Council of Academies, Ottawa.
- Clark, J. & DeRose, S. (eds) 1999 *XML Path Language (XPath)*, Version 1.0. World Wide Web Consortium, November 16, 1999. Available from: <http://www.w3.org/TR/xpath/>
- Cox, S. (ed.) 2007a *Observations and Measurements – Part 1 – Observation Schema*, Version 1.0. Open Geospatial Consortium, OGC 07-022r1.
- Cox, S. (ed.) 2007b *Observations and Measurements – Part 2 – Sampling Features*, Version 1.0. Open Geospatial Consortium, OGC 07-002r3.
- de la Beaujardiere, J. F. (ed.) 2004 *Web Map Service*, Version 1.3.0. Open Geospatial Consortium, OGC 03-109r1.
- Federal-Provincial Working Group on Groundwater 1995 *Groundwater Data Management*. Environment Canada, Ottawa.
- Gogu, R. C., Carabin, G., Hallet, V., Peters, V. & Dassargues, A. 2001 *GIS-based hydrogeological databases and groundwater modeling*. *Hydrogeology Journal* 9 (6), 555–569.
- International Organization for Standardization 2006 *ISO 19118:2005, Geographic Information – Encoding*. ISO, Geneva.
- International Organization for Standardization 2007a *ISO 19136:2007, Geographic Information – Geography Markup Language (GML)*. ISO, Geneva.
- International Organization for Standardization 2007b *ISO 19139:2007, Geographic Information – Metadata – XML Schema Implementation*. ISO, Geneva.
- Oulidi, H. J., Ralf, L., Benaabidate, L. & Wachter, J. 2009 *HydrIS: An open source GIS decision support system for groundwater management (Morocco)*. *Geo-spatial Information Science* 12 (3), 212–216.
- Lake, R. 2005 *The application of geography markup language (GML) to the geological sciences*. *Computers and Geosciences* 31 (9), 1081–1094.
- Masser, I. 2010 *Building European Spatial Data Infrastructures*, Second Edition. ESRI Press, Redlands, California.

- Na, A. & Priest, M. (eds) 2007 *Sensor Observation Service*, Version 1.0. Open Geospatial Consortium, OGC 06-009r6.
- National Groundwater Committee, Working Group on National Groundwater Standards 1999 *The Australian National Groundwater Data Transfer Standard*, Release 1.0 July 1999. Commonwealth of Australia.
- Neuendorf, K. K. E., Mehl, J. P. & Jackson, J. A. (eds) 2005 *AGI Glossary of Geology*. American Geological Institute, Alexandria, Virginia.
- Panagiotis, A. V. (ed.) 2005 *Web Feature Service Implementation Specification*, Version 1.1.0. Open Geospatial Consortium, OGC 04-094.
- Portele, C. 2007 *Geography Markup Language (GML) Encoding Standard*, Version 3.2.1. Open Geospatial Consortium, OGC 07-036.
- Raskin, R. G. & Pan, M. J. 2005 [Knowledge representation in the Semantic Web for Earth and Environmental Terminology \(SWEET\)](#). *Computers and Geosciences* **31** (9), 1119–1125.
- Rivera, A., Crowe, A., Kohut, A., Rudolph, D., Baker, C., Pupek, D., Shaheen, N., Lewis, M. & Parks, K. 2003 *Canadian Framework for Collaboration on Groundwater*. Geological Survey of Canada, Ottawa.
- Sen, M. & Duffy, T. 2005 [GeoSciML: development of a generic Geoscience Markup Language](#). *Computers and Geosciences* **31** (9), 1095–1103.
- Sheth, A. P. 1999 Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In: *Interoperating Geographic Information Systems* (M. Goodchild, M. Egenhofer, R. Fegeas & C. Kottman, eds), Kluwer. Dordrecht, pp. 5–29.
- Statistics Canada 2010 Human Activity and the Environment (HAE), Freshwater supply and demand in Canada. *The Daily*, 13 September 2010, pp. 7–10. Available from: <http://www.statcan.gc.ca/daily-quotidien/100913/dq100913b-eng.htm>.
- Strassberg, G., Maidment, D. R. & Jones, N. L. 2007 A geographic data model for representing ground water systems. *Groundwater* **45** (4), 515–518.
- Taylor, P. 2009 *Harmonising Standards for Water Observation Data – Discussion Paper*, Version 1.0.0. Open Geospatial Consortium Discussion Paper, OGC 09-124r2.
- Tripathi, A. & Babaie, H. A. 2007 [Developing a modular hydrogeology ontology by extending the SWEET upper-level ontologies](#). *Computers & Geosciences* **34** (9), 1022–1033.
- United Nations 2008 *Our Waters: Joining Hands Across Borders – First Assessment of Transboundary Rivers, Lakes and Groundwater*. United Nations, Economic Commission for Europe, New York and Geneva.
- Yang, X., Steward, D. R., de Lange, W. J., Lauwo, S. Y., Chubb, R. M. & Bernard, E. A. 2010 [Data model for system conceptualization in groundwater studies](#). *International Journal of Geographical Information Science* **24** (5), 677–694.
- Wojda, P. 2008 *Hydrogeological Data Modeling in Groundwater Studies*. PhD Dissertation, University of Liege, November.

First received 11 May 2010; accepted in revised form 3 November 2010. Available online 22 April 2011