

Groundwater data network interoperability

Boyan Brodaric, Nate Booth, Eric Boisvert and Jessica Lucido

ABSTRACT

Water data networks are increasingly being integrated to answer complex scientific questions that often span large geographical areas and cross political borders. Data heterogeneity is a major obstacle that impedes interoperability within and between such networks. It is resolved here for groundwater data at five levels of interoperability, within a Spatial Data Infrastructure architecture. The result is a pair of distinct national groundwater data networks for the United States and Canada, and a combined data network in which they are interoperable. This combined data network enables, for the first time, transparent public access to harmonized groundwater data from both sides of the shared international border.

Key words | data network, groundwater, groundwater level, interoperability, time series, water well

Boyan Brodaric (corresponding author)
Geological Survey of Canada,
Ottawa, ON,
Canada
E-mail: brodaric@nrcan.gc.ca

Nate Booth
US Geological Survey,
Reston, VA,
USA

Eric Boisvert
Geological Survey of Canada,
Quebec, QC,
Canada

Jessica Lucido
US Geological Survey,
Middleton, WI,
USA

ABBREVIATIONS

| | | | |
|----------|--|---------|---|
| API | Application Programming Interface | SOA | Service-Oriented Architecture |
| CSA | Client–Server Architecture | SOS | Sensor Observation Service |
| GeoSciML | Geoscience Markup Language | Turtle | Terse RDF Triple Language |
| GEOSS | Global Earth Observation System of Systems | URI | Uniform Resource Identifier |
| GIN | Groundwater Information Network | VVG | Visualizing Victoria’s Groundwater |
| GML | Geographic Markup Language | WaterML | Water Markup Language |
| GWML | GroundWater Markup Language | WFS | Web Feature Service |
| HTML | HyperText Markup Language | WMS | Web Mapping Service |
| HTTP | Hypertext Transfer Protocol | WOA | Web-Oriented Architecture |
| INSPIRE | Infrastructure for Spatial Information in the European Community | XML | Extensible Markup Language |
| JSON | JavaScript Object Notation | XSLT | EXtensible Stylesheet Language Transformation |
| LOD | Linked Open Data | | |
| NGWMN | National Ground-Water Monitoring Network | | |
| O & M | Observations and Measurements | | |
| OGC | Open Geospatial Consortium | | |
| OWL | Web Ontology Language | | |
| RDF | Resource Description Framework | | |
| REST | Representational State Transfer | | |
| RiverML | River Markup Language | | |
| SDI | Spatial Data Infrastructure | | |
| SKOS | Simple Knowledge Organization System | | |

INTRODUCTION

Canada and the United States share the world’s longest international border. It is crossed by many aquifers supplying water to both sides of the border, causing cross-border groundwater management to be a significant issue. This issue, as well as other cross-border science issues related to climate, energy, public health and safety, can be greatly facilitated by the efficient sharing of groundwater data.

Such sharing is generally complex, because the data are segmented among multiple data providers at the watershed, State, Provincial, and Federal levels, on each side of the border. This greatly amplifies the overall heterogeneity of the data, making cross-border groundwater data difficult to find and use. To ease data access, and overcome segmentation and heterogeneity, Federal agencies in Canada and the United States are developing data networks that provide a united view over the distributed data sources: the Canadian Groundwater Information Network (GIN) and the US National Ground-Water Monitoring Network (NGWMN). After mutual experimentation (Brodaric & Booth 2010; Brodaric et al. 2013), these two data networks have become increasingly interoperable: as a result, the data holdings of one network can be seen as an extension of the other network, such that data are exchanged seamlessly and automatically between networks, and any of the total pool of data can potentially be accessed from either network. Such interoperability is achieved through the implementation of a variety of international open geospatial data standards, and the development of a novel solution for North American groundwater data at five levels of interoperability: systems, syntax, structure, semantics, and pragmatics (Sheth 1999; Brodaric 2007). While the technical approaches adopted by the data networks at each level are generally well known, architectures for hydro interoperability have not yet been framed using the five levels, and the depth of implementation at each level is also new for hydro interoperability. GIN and NGWMN also represent the first national and international data networks for groundwater data. This work therefore makes the following original contributions: (1) it establishes a five-level architecture for hydro data interoperability in which each level is developed extensively; (2) it links and harmonizes national groundwater data from the United States and Canada, thereby representing a significant new online resource for the data.

The paper is organized into sections that describe: (1) two examples of data heterogeneity encountered by the data networks; (2) background material on data interoperability, including prevailing levels, architectures, infrastructures, and networks; (3) related work on geospatial and hydrologic data networks; (4) the solution adopted by GIN and NGWMN for groundwater data; (5) implementation examples; and (6) a summary and brief indication of

possible future directions. To avoid disruption of the text, common technical abbreviations are expanded at the beginning of the paper.

GROUNDWATER DATA HETEROGENEITY

Prominent requirements and usage scenarios for groundwater data interoperability are described in Boisvert & Brodaric (2012). An additional key usage scenario is trans-border aquifer management in which it is essential for international neighbors to seamlessly exchange information about aquifers located on both sides of the shared border, and where data are provided by Federal, State and Provincial sources. These usage scenarios require a groundwater data network to return data in response to queries for a variety of groundwater features, from multiple data sources. These features include, for example, water wells, aquifers, water bodies, and management areas, in addition to observations related to these features, such as time-indexed groundwater levels. The main obstacle to be overcome is data heterogeneity, as the representation of features and observations often varies greatly between and within data networks. Data heterogeneity is a by-product of the multiplicity of groundwater data providers in a data network, inasmuch as each provider typically utilizes different database systems, data structures, transfer formats, etc. Such heterogeneity is inherent in the water well records shown in Figure 1, which originate from different sides of the USA–Canada border, specifically from Alberta and Montana (Government of Alberta 1978; Montana Bureau of Mines & Geology 1999), and which intersect the same cross-border aquifer (i.e., the Milk River Aquifer). A comparison of these records reveals that the tabular organizations differ, with the use of one or two columns for depth intervals (structure difference); the lithology vocabularies vary, for example, ‘sand’ and ‘sand-fine’ (semantics difference); and the units of measure for depths vary, i.e., meters vs feet, due to dissimilar measurement practices (pragmatic difference). In addition, although not illustrated, the underlying database systems are not the same (systems difference) and the characters can vary as Canadian data often include French letters (syntax difference). Data heterogeneity is also shown in Figure 2, where groundwater level

| Depth from ground level (ft) | Water Bearing | Lithology Description | To | Description |
|------------------------------|---------------|-------------------------|----|---|
| 4.00 | | Sand | 4 | SOIL |
| 19.00 | | Sandy Clay & Gravel | 6 | CLAY- YELLOW |
| 29.00 | | Brown Sandy Shale | 10 | CLAY- YELLOW- CONTAINS COBBLES |
| 120.00 | | Gray Sandy Clay | 15 | CLAY- YELLOW- CONTAINS THIN LAYER OF SAND |
| 170.00 | | Sandy Clay | 30 | CLAY- YELLOW- SOFT |
| 174.00 | | Coal | 35 | CLAY- GRAY BROWN- SOFT |
| 183.00 | | Silty Sandstone | 41 | CLAY- GRAY BROWN- CONTAINS SOME SAND |
| 194.00 | | Silty Sandstone | 45 | CLAY- GRAY |
| 240.00 | | Sandy Shale & Siltstone | 52 | CLAY- GRAY- AND FINE SAND. CONTAINS SOME COAL |
| 314.00 | | Shale | 55 | GRAVEL |
| | | | 65 | SAND- FINE CONTAINS FRAGMENTS OF COAL |
| | | | 95 | SAND- FINE |

Figure 1 | Data heterogeneity between Canadian (Alberta, left) and US (Montana, right) water well records.

measurements exhibit noticeable differences, particularly in semantics where diverse vocabulary is used, e.g., for observation methods, and in structure where the groundwater level observations are stored either in a distinct table (Montana), or are intermingled by GIN with other types of observations in a single table (Alberta). As with the wells in Figure 1, these measurements originate at sites located on either side of the international border and within the same cross-border aquifer.

While groundwater data interoperability exhibits significant similarities to interoperability in other hydro and geoscientific domains, notable differences are also evident. The similarities suggest surface and subsurface water data networks might adopt common systems and syntax, and the differences in key entities suggest a disparity in at least data structure, semantics, and pragmatics. The similarities include a common need to interconnect massive observational networks with the features being observed, such as the water bodies in rivers, lakes, or aquifers. The types of

observations are also similar across the surface and subsurface water domains, and include properties such as water level or temperature, as well as a vast array of chemical, biological, and material constituents. Despite these significant resemblances, differences can exist in the methods used to make measurements, as well as in the selection of properties and constituents to measure. Another major difference involves the features that contain a water body. For example, inland surface water containers include water channels and their various parts, as well as features such as watersheds and catchments. In contrast, related groundwater containers include aquifers and other geological bodies, such as geological formations. Features related to sampling are another difference, as water wells are unique to the groundwater domain.

An interoperability scenario suggested by the data heterogeneity in Figures 1 and 2, and motivated by cross-border aquifer management through co-location of the data within the same cross-border aquifer, requires an

| time_stamp | numeri | feature_of | offering_id | procedure_id | | | | | |
|------------------------|---------|------------|-------------|----------------------|-------------------------|-------|------|------|---------|
| 2009-05-31 22:00:00-04 | 944.362 | ab.mon.100 | GW_LEVEL | urn:ogc:object:featu | 2013-11-18 13:42:00.000 | 16.05 | NULL | MBMG | SOUNDER |
| 2009-05-31 23:00:00-04 | 944.365 | ab.mon.100 | GW_LEVEL | urn:ogc:object:featu | 2013-08-13 10:27:00.000 | 16.00 | NULL | MBMG | SOUNDER |
| 2009-06-01 00:00:00-04 | 944.347 | ab.mon.100 | GW_LEVEL | urn:ogc:object:featu | 2013-05-06 08:50:00.000 | 16.25 | NULL | MBMG | SOUNDER |
| 2009-06-01 01:00:00-04 | 944.343 | ab.mon.100 | GW_LEVEL | urn:ogc:object:featu | 2012-10-19 15:07:00.000 | 16.82 | NULL | MBMG | SOUNDER |
| 2009-06-01 02:00:00-04 | 944.34 | ab.mon.100 | GW_LEVEL | urn:ogc:object:featu | 2012-05-19 16:28:00.000 | 16.45 | NULL | MBMG | SOUNDER |
| 2009-06-01 03:00:00-04 | 944.335 | ab.mon.100 | GW_LEVEL | urn:ogc:object:featu | 2012-01-06 13:11:00.000 | 16.48 | NULL | MBMG | SOUNDER |
| 2009-06-01 04:00:00-04 | 944.331 | ab.mon.100 | GW_LEVEL | urn:ogc:object:featu | 2011-06-12 16:43:00.000 | 16.05 | NULL | MBMG | SOUNDER |
| 2009-06-01 05:00:00-04 | 944.326 | ab.mon.100 | GW_LEVEL | urn:ogc:object:featu | 2010-09-22 19:03:00.000 | 17.08 | NULL | MBMG | SOUNDER |
| 2009-06-01 06:00:00-04 | 944.32 | ab.mon.100 | GW_LEVEL | urn:ogc:object:featu | 2010-05-01 10:22:00.000 | 17.57 | NULL | MBMG | SOUNDER |

Figure 2 | Data heterogeneity between Canadian (Alberta, left; from GIN cache) and US (Montana, right) groundwater level observations.

application to retrieve the well and groundwater level data in a common format. Specifically, because the GIN web portal only displays data that conform to certain international data standards, the heterogeneous Alberta and Montana data must be transformed to these standards so that the groundwater information can be viewed in the GIN web portal during aquifer assessment. This scenario is resolved in the section below on demonstration of results.

BACKGROUND

Interoperability has been studied widely and diversely, with mature approaches demonstrated in many domains by major research initiatives and operational systems. Interoperability is also defined variously although, in general, it refers to the working together of diverse autonomous entities to achieve a common goal, requiring the entities to possess the ability to exchange information about system function and domain content (Manso et al. 2009). Data interoperability then specifically refers to a collaboration among data providers in which their goal is to exchange, deliver, or use data in a coordinated way. This involves the ability to send and receive messages that are mutually understood (Brodeur et al. 2003), including requests for data as well as responses that contain either retrieved data or control messages such as error descriptions. Such messages must typically be transformed at each interoperability level, either by the sender or receiver, to a construct that can be readily consumed and thus understood by the receiver – this process is often referred to as alignment. The components involved in attaining alignment can be arranged in various ways. Thus, in addition to describing levels of transformation, we also distinguish here between architectures, infrastructures, and networks to progressively describe the arrangement of the components that enable message alignment, hence interoperability, between data providers.

Interoperability levels

Interoperability transformations are typically conceptualized as nested multilevel stacks, and key stacks include those for information system interoperability, the Semantic Web, simulation interoperability, and computational

semiotics (Sheth 1999; Stamper et al. 2000; Horrocks et al. 2005; Tolk et al. 2007). Although the type and number of levels can vary among stacks, there exist significant commonalities (Brodaric 2007; Manso et al. 2009), and each stack has been applied variously to geospatial data (Manso et al. 2009). The levels adapted here follow primarily the information system interoperability stack (Sheth 1999), consisting of systems, syntax, structure, and semantics, additionally topped with pragmatics (Brodaric 2007), as shown in Figure 3. These levels equally apply to any form of data interoperability, including within a data network or between data networks.

- **Systems:** refers to hardware or software elements required for core functions such as message passing or data manipulation, and largely involve platform aspects such as operating systems, transmission protocols (e.g., Hypertext Transfer Protocol (HTTP)), or particular database limits (e.g., maximum record size or maximum number of points in a polygon). Systems interoperability involves overcoming heterogeneity in such aspects.
- **Syntax:** refers to the language used to encode a message, including requests for data as well as the actual data content. Syntax includes a language's alphabet, words, and grammar, and it can be abstract or concrete: for example, the Resource Description Framework (RDF) language has an abstract syntax composed of nested triples, as well as several concrete encodings such as RDF/XML or Turtle. Syntax interoperability thus involves

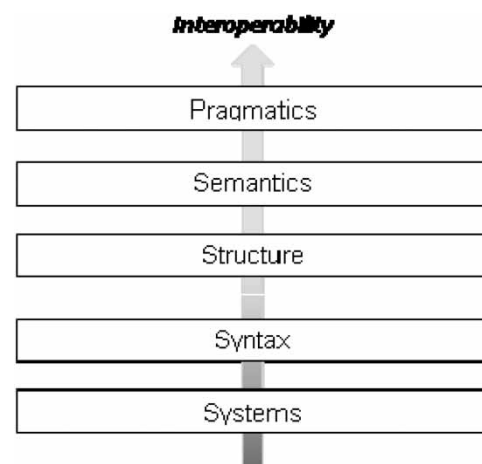


Figure 3 | Levels of interoperability: transformation of messages to a common construct proceeds from bottom to top.

overcoming differences in abstract or concrete syntaxes. It is achieved in Open Geospatial Consortium (OGC) geospatial web service standards primarily through the common deployment of the Geographic Markup Language standard (GML; [Portele 2012](#)) for the encoding of geographical features; alternatively, a standard RDF option has also lately emerged via the GeoSPARQL standard ([Perry & Herring 2012](#)), alongside other non-OGC syntaxes such as GeoJSON ([Butler et al. 2008](#)).

- **Structure:** refers to the arrangement of the parts of a message, such as representing something as one entity or many (e.g., the diverse structures for well intervals in [Figure 1](#)). Structures are typically manifest as schemas that provide a pattern for arranging the message. They are realized in the OGC suite of standards as domain-specific schema that apply OGC languages such as GML and GeoSPARQL. Hydro-related schema are primarily GML applications such as WaterML for water measurements ([WaterML2; Taylor 2012](#)), GroundWaterML for subsurface features like wells or aquifers ([GWML1; Boisvert & Brodaric 2012](#)), HY_Features for a broad suite of hydrologic features including river networks ([Dornblut & Atkinson 2014](#)), and RiverML for river channel descriptions ([Jackson et al. 2014](#)). Structure interoperability is also manifest by standard OGC web service structures, which provide a canonical interface to geospatial operations, including many for data access. Structural interoperability involves overcoming diverse structures for data or related web services, via alignment of associated schemas.
- **Semantics:** refers to the meaning inherent in some component of a schema or data content. Such meanings are typically represented as digital definitions, and their form can vary from highly unstructured, e.g., free-form text, to highly structured, e.g., inter-related logic statements. The latter forms the basis for machine-readable ontologies, while vocabularies (e.g., glossaries, thesauri) are considered here to be semi-structured and thus positioned somewhere in middle of this range. An ontology element, often referred to as a concept, consists of a label, a logic-based definition, and its connection to other concepts via formal relations such as IS-A or PART-OF. In contrast, a vocabulary element, here called a term, might consist of a label, a text definition, and its connection to other terms via simple linguistic relations

such as SYNONYM. For data interoperability, both ontologies and vocabularies are often used to overcome semantic heterogeneity, primarily by aligning concepts and/or terms to deal with meaning differences. For example, synonymy occurs when multiple words refer to the same meaning, and polysemy occurs when multiple meanings refer to the same word. A typical solution then maps the words to a canonical ontology or vocabulary containing distinct concepts or terms, respectively, for each meaning. OGC provides minimal support for semantics in relation to GML data delivery. This occurs primarily through feature catalogs, which have moderately expressive techniques for defining geospatial feature types (i.e., concepts), such as ‘river’ or ‘road’, and through standards for deriving, storing, and using ontologies developed from schemas for data or web services ([ISO/TS 19150-1 2012](#)). However, these efforts are not tightly integrated with GML for encoding and transmitting data, requiring integration of GML with ontologies and vocabularies to occur via ad hoc techniques often normalized within specific communities. In contrast, the Semantic Web enables geospatial feature types and data to be tightly coupled via languages such as RDF/OWL, which is leveraged by the OGC GeoSPARQL standard. Integration of RDF/OWL approaches with GML-based standards and related systems is ongoing and represents a fertile research domain ([Schade & Smits 2012; Harvey et al. 2014](#)).

- **Pragmatics:** refers to contextual factors. While such factors can be quite diverse, including legal, organizational, economic, and other such factors, pragmatic interoperability frequently involves information about data provenance as well as guidelines about data use including best practices – how the data have been created and how they should be used or accessed ([Brodaric 2007](#)). Such factors are important because interoperability can be impeded even if all remaining levels are aligned. For example, the same type of data (semantically) could be collected using a variety of scientific methods that might be incompatible, due to application of discordant instruments or procedures, thus prohibiting data integration and the unified use of the data. OGC supports pragmatics through best practice specifications that further constrain how web services are to be deployed within specific domains or communities, typically by placing restrictions on aspects of a web service

interface (e.g., [GEOWOW Consortium 2014](#)). Although such specifications could document how data are to be used, they typically describe how data should be accessed within a particularly constrained environment.

Interoperability architectures

A conceptual interoperability architecture describes how functional components are arranged to enable transformation at each level. Such architectures are orthogonal to the interoperability levels, as any such component can provide functions that cross levels; for example, a component might carry out syntactic, structure, and semantic translation, or it might provide infrastructure support for other core functions such as data discovery. Such architectures, in general, can be described in terms of the type of functional components, connectivity, or layers ([Heitmann et al. 2014](#)). Functional interoperability components include those for the storage of data, metadata, and knowledge artifacts (e.g., schemas, ontologies, vocabularies, queries, and mappings), as well as for their access, caching, cache update, discovery, translation, integration, distribution, orchestration, mediation (i.e., via a mediator), and display. Connectivity approaches describe how functional components are organized, and include configuration aspects such as: (1) centralized or distributed, to describe the logical and physical location of components within a data network; (2) monistic or pluralistic, to describe the use of single or multiple knowledge artifacts; and (3) static or dynamic, to describe the degree of change among components, data, or knowledge artifacts. Unlike component or connectivity approaches, layer-based architectures focus on the role of a component within the data network, and prominent examples include client-server, service-oriented, and web-oriented architectures.

- **Client-server architecture (CSA):** is a two layer approach to distributed computing in which tasks are segmented between resource requesters (clients) and providers (servers). In interoperable data networks, data management components (e.g., storage, cache, access, and update) are typically server responsibilities, display is a client responsibility, and the rest are deployed variously. As clients and servers are tightly coupled, components tend to be platform-specific resulting in high system heterogeneity.

- **Service-oriented architecture (SOA):** is a three layer and loosely coupled approach in which platform-neutral and often standardized middleware is introduced between servers and clients. For example, the OGC standard web services represent canonical middleware primarily for accessing geospatial data, metadata, and executable codes. Component interactions follow a publish-find-bind strategy whereby server resources, such as data, are published via metadata descriptions and discovered by clients who request them from servers. The use of SOA primarily reduces system heterogeneity, due to the focus on platform independence.
- **Web-oriented architecture (WOA):** extends SOA, mainly by adopting principles for web service interfaces from the web-oriented Representational State Transfer (REST) approach, such as standard web functions (HTTP GET, POST, etc.), identifiers Uniform Resource Identifier (URIs), and transport protocols (HTTP). Obtaining a web resource, such as a data set, then amounts to issuing a web request using a standard web function and identifier. WOA is thus resource-centric in contrast to the function-centric approach of SOA. WOA can be seen as further reducing structure heterogeneity via implementation of simpler web service interfaces, and by advancing pragmatic interoperability through increased ease-of-use of web services.

Geospatial interoperability infrastructures

Interoperability infrastructures refer to particular configurations of levels and architectures, and effectively constitute technical paradigms for data interoperability. Geospatial interoperability infrastructures target geospatial data and are most prominently exemplified by Spatial Data Infrastructures (SDI; [Masser 2010](#)) and Linked Open Data (LOD; [Kuhn et al. 2014](#)).

- **Spatial Data Infrastructures:** an SDI is a network of online geospatial data resources, and affiliated people, policies, agreements, standards, and technological approaches that aim to share geospatial data and executable code over the web ([Masser 2010](#)). The vast majority of SDI implement SOA-based OGC standards at various levels of interoperability, causing SDI to be largely aligned with

SOA-based techniques despite ongoing efforts to incorporate WOA approaches (Harvey et al. 2014). As a result, SDI can be seen to refer to two things, somewhat ambiguously: (1) a suite of technological and social approaches weighted towards SOA, which constitute an infrastructure; and (2) an interlinked collection of specific data sets and codes that implement these approaches, which constitute a network. In this paper we distinguish where required between SDI as an infrastructure vs a network. Functional components in SDI-based infrastructures are focused primarily on data and metadata access, and schemas are the main knowledge artifact. Data components tend to be distributed, while metadata components tend to be centralized. Monistic schemas often exist for various domains, networks can be static or dynamic, and a wide array of data sources are supported, such as those for maps, objects, fields, and sensors.

- **Linked Open Data:** also aim to connect data on the web, however, primarily via WOA. LOD refines WOA principles by distinguishing between (a) web-resolvable identifiers (URIs) that return useful structured information about data, (b) a web page, and (c) its type definition, and also by promoting connectivity within and between these resources (Kuhn et al. 2014). LOD also adopts the Semantic Web stack, and thus supports all interoperability levels, but explicitly omits schemas inasmuch as structure constraints can be implicitly handled at the semantic and pragmatic levels. Semantics is tightly integrated, with intrinsic support for ontologies via RDF, Web Ontology Language (OWL), and related languages. LOD architectures are migrating from centralized to distributed approaches, which are strongly pluralistic and dynamic, particularly for knowledge artifacts such as ontologies. As a result, some aspects of these artifacts can often be discovered on-the-fly without adherence to preset and fixed content standards. Integration of LOD and SDI approaches is an ongoing concern (Schade & Smits 2012; Harvey et al. 2014).

Geospatial data interoperability networks

A geospatial data interoperability network is a group of geospatial data providers that interact using a particular implementation of an infrastructure. For example, major national multi-theme SDI data networks that utilize SDI

infrastructures exist in Canada (e.g., GIN), the United States (e.g., NGWMN), and Europe (INSPIRE 2008), as well as globally (e.g., Global Earth Observation System of Systems (GEOSS); Nativi et al. 2014). For the most part, such SDI data networks are self-contained with limited connectivity between them – a counter-example is GEOSS, which explicitly aims for inter-connectivity. LOD networks are somewhat different, as LOD aims to create one open interconnected data network, the so-called LOD cloud, which is essentially part of the greater world-wide web. However, despite this aim, many data sets within this cloud are relatively isolated, essentially forming a distinct quasi-network, because connectivity is greater within than without; the number of external links, to or from the data set, is significantly less than the number of internal links within the data set, although this situation is highly dynamic and changing (Hogan et al. 2012).

RELATED WORK

Data interoperability has been widely researched theoretically with diverse applications in many disciplines. Of relevance here is the interoperability of geospatial, hydrological, and groundwater data, and the evolution and transference of approaches from the geospatial to hydrological to groundwater domains.

- **Geospatial data interoperability:** interoperability levels for geospatial data (Bishr 1998; Sheth 1999; Brodaric 2007; Manso et al. 2009) are addressed via architectural approaches that have evolved initially from SOA to SDI to LOD (Wache et al. 2001; Manso et al. 2009; Nativi et al. 2014; Kuhn et al. 2014). Associated data networks span the variety of disciplines concerned with geospatial data, including the geosciences such as geology, land cover and use, soils, oceans, vegetation, and others (Brodaric & Gahegan 2006; INSPIRE 2008; Durbha et al. 2009; Lutz et al. 2009), but have not been created to date for the groundwater domain.
- **Surface water data interoperability:** early interoperability data networks in the hydro domain focused on surface water data, rather than groundwater data, initially through the adoption of SOA and proprietary web

services (e.g., Tarboton *et al.* 2011), including approaches to semantics (Beran & Piasecki 2009; Duce & Janowicz 2010), soon followed by SDI architectures and standard web services (Bermudez & Arctur 2011; Yu & Di 2014). Complementing this shift to SDI approaches is ongoing work on standard schemas for various aspects of surface water (Taylor 2012; Dornblut & Atkinson 2014; Jackson *et al.* 2014). Other recent developments, primarily among academic researchers, include LOD approaches to surface water to aid the creation of relations between data, particularly to aid decision-making (Curry *et al.* 2014; Anzaldi & Wu 2014). While many of these architectural aspects are transferable to groundwater data, groundwater features remain largely a secondary concern and are not represented comprehensively, and semantics support is focused primarily on data discovery rather than data delivery.

- **Groundwater data interoperability:** until recently the construction and linkage of groundwater data networks has been largely absent outside the GIN and NGWMN projects. Early SDI efforts in groundwater include the specification of schemas for data exchange in North America and Europe (Boisvert & Brodaric 2012; INSPIRE 2013). These serve as a foundation for the initiation and ongoing development of international standards for groundwater features (Lucido & Booth 2014), as well as the construction of SDI-based groundwater data networks such as the Visualizing Victoria's Groundwater project for the Australian state of Victoria (Dahlhaus *et al.* 2012), and emerging efforts in New Zealand (Klug & Kmoch 2014). However, these groundwater data networks have not (yet) been explicitly linked with other such networks, in contrast to GIN and NGWMN.

RESULTS: GROUNDWATER DATA INTEROPERABILITY

In relation to the GIN and NGWMN data networks, data interoperability refers to the ability of each network to receive requests for data, and return appropriate data unified from multiple distributed, autonomous, and heterogeneous data sources. Both GIN and NGWMN can, in theory, include the other network as a data source – in essence, each can operate as a virtual unified online database that

encompasses the other network's data holdings; although to date only GIN has enabled access to NGWMN. The cumulative holdings are massive, comprising millions of features, thousands of sensors, and billions of data values.

GIN and NGWMN architectures

Both GIN and NGWMN implement an SDI architecture. To enhance data network performance, primarily to optimize responses to data requests, both maintain centralized data caches, but they differ in the extent of cache, the cache contents, and in cache usage and completeness. NGWMN centrally pools all data in its US network, such that all data requests are answered from this cached copy. GIN, on the other hand, pools only some Canadian data, obtaining and translating the remainder dynamically upon request. In terms of cache contents, GIN caches original data, translating it on-the-fly to standard outputs, such as GWML1 (for features) or WaterML2 (for observations), in response to requests, while NGWMN translates and then caches the standard representations directly, thus avoiding on-the-fly translation, but thus limiting itself to specific representations. NGWMN maintains a complete cache of all observational data at all times, intended for all uses, while GIN maintains two caches of partial data for specific uses: a dynamic and temporary cache of recently requested observational data strictly for presentation functions, and a permanent cache of observation and other data for other functions such as download. Cached data are harvested periodically by both data networks via the OGC standard web services.

For external machine access to the data, each data network maintains a central access point consisting of the Web Mapping Service (WMS) (de la Beaujardiere 2006), Web Feature Service (WFS) (Vretanos 2005), and Sensor Observation Service (SOS) (Broring *et al.* 2012) standard OGC interfaces, although some REST-conformant services are also available or planned. This enables third-party web portals and applications to directly incorporate either network's data. For external human access to the data, each data network maintains a central web portal that variously leverages the external access points: the NGWMN portal (<http://cida.usgs.gov/ngwmn/>) uses the web services or direct database connections to reach its data cache, as

shown in Figure 4, while GIN's portal (<http://gw-info.net>) singularly utilizes the external web services for data access. Both data networks utilize a hybrid design in which some data sources and access components are distributed, while the remaining components are centralized. A key centralized functional component is the mediator, used by the data networks as a broker for the translation and distribution of requests and data. However, the role played by each mediator differs: in NGWMN the mediator is primarily used to harvest data, is invoked prior to caching, and solely translates returned data; while GIN's mediator is primarily used to query data, is invoked after caching, and translates both incoming queries and returned data. This affects query strategy: NGWMN queries its cache solely, enabling sophisticated queries to be returned efficiently in predictable times; while GIN queries both its cache and non-cached data sources, with the latter requiring translation to local query languages, often simplified, with varying and unpredictable response times. While these performance differences are mostly negligible for small responses (see demonstration of results below), they can be significant for larger responses: for example, GIN might need to wait for a local data server to respond, as a trade-off for receiving the latest data and ensuring complete

autonomy of data governance for the data provider. Another difference, in addition to the mediation strategy, is GIN's deployment of a centralized repository and web service for the vocabularies used during translation.

Both GIN and NGWMN are primarily monistic, as they support a selected suite of schemas and vocabularies for data requests and responses, although some features can be returned in multiple syntaxes, e.g., GML or JSON. They are also fundamentally static as the data sources are permanently connected to each data network, and while the amount of content might increase, the overall structure, semantics, and pragmatics of the data are relatively stable and do not fluctuate greatly. Another similarity is the adoption by both data networks of a typical three-tier SDI configuration for functional components, as shown in Figure 4: the bottom tier consists of distributed data components wrapped by data access components; the middle tier consists of centralized middleware components such as a mediator, data caches, metadata, and semantic repositories; and the top tier consists of web portals or web applications that use the middle tier to access data from the data networks.

GIN and NGWMN interoperability workflow

Addition of a data source to either data network involves registration of source details in the network's metadata catalog. This involves registration of artifacts at all five levels of interoperability, including web service locations (systems) and profiles for their invocation (pragmatics), as well as a mapping between source data and the network for each language-specific encoding (syntax) of a source schema (structure) and vocabulary (semantics). Once these are registered, the data networks can interoperate with the data sources.

GIN's centralized mediator is the core functional component for such interoperability, as it receives and responds to data requests, and in doing so it carries out distribution and translation functions. Requests for data originate from either human-driven web clients (e.g., the GIN portal) or machine-driven applications (e.g., external applications or its own data harvester). A request is made by calling an external web service, which triggers the mediator. The mediator determines which data sources, or

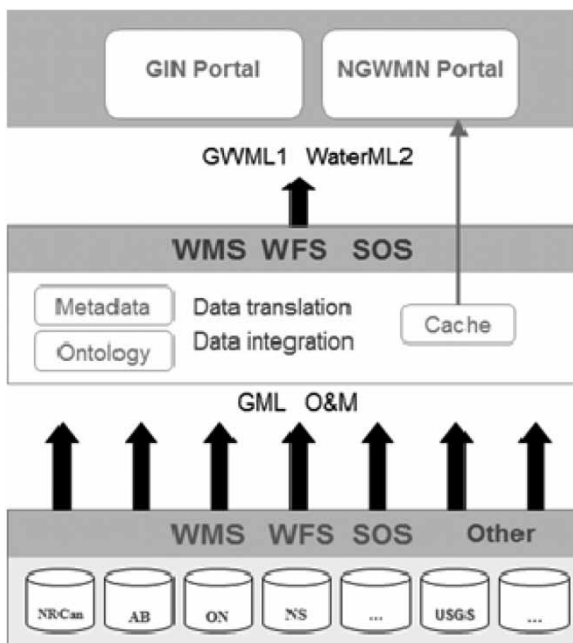


Figure 4 | GIN and NGWMN tiered groundwater data interoperability architecture.

cache, should receive the request by using data discovery components tied to metadata catalogs, and it translates the original request at each level to a request appropriate for each data source, finally sending the suitably translated requests to the data sources. Next, the mediator receives and translates the data responses at each level to a canonical standard (e.g., GWML1 for groundwater features, WaterML2 for observations), using the previously established data mappings and associated technologies, such as RDF repositories for vocabularies. The specific translations at each level are discussed in the next section. Lastly, it integrates the translated responses from each data source into a single result and returns this result to the requester. Where translation is not required at a specific step or level, GIN simply shuffles the request or response to the next portion of the workflow.

NGWMN provides the same external web services as GIN. However, unlike GIN, NGWMN's mediator is not invoked upon data requests to these external services, as the data are simply retrieved from the central pre-translated cache without translation or distribution. Mediation plays a role only in the harvesting process, during population of the cache: the mediator initiates requests for external data, customized to source query needs and using data discovery components, and translates the responses to its internal database structure for caching.

GIN and NGWMN levels of interoperability

The translation component within an interoperability architecture requires messages, primarily data requests and responses, to be transformed at each interoperability level. To compare interoperability strategies at each level, it is then useful to delineate between inter-network and intra-network interoperability: inter-network strategies consider interoperability approaches solely between GIN and NGWMN, while intra-network strategies are focused on approaches between a data network and its source data from Provincial and State agencies.

For inter-network interoperability, heterogeneity is largely eliminated between the data networks due to the adoption of common standards at each level. Consequently, many common knowledge artifacts, such as queries or schema, need not be transformed between the data

networks; transformation is only required for a few disparities, such as some diverse vocabularies. Common aspects are restricted to certain core datatypes and functionality, and these represent a subset of the resources available for each data network. For example, NGWMN includes water quality data and some REST-conformant interfaces, whereas GIN does not. Inter-network interoperability is thus targeted at key overlapping data types and functions, and translation of knowledge artifacts is largely focused on the semantic level.

The same does not hold for intra-network interoperability, as heterogeneity prevails between the data sources and the common standards used by the data networks. Transformations are then required at each level to ensure intra-network interoperability. GIN and NGWMN utilize various strategies at each level to achieve intra-network interoperability.

In greater detail, the inter-network and intra-network interoperability approaches utilized at each level involve the following.

- **GIN and NGWMN system interoperability:** is achieved through the implementation of technologies that enable the deployment of platform-independent OGC web service standards, which eliminate platform-specific system heterogeneity. These web services represent the canonical interfaces for all machine interactions not only between the GIN and NGWMN networks, but also to all other users of the data, and thus denote a common strategy for both inter-network and intra-network interoperability. They include WMS for accessing map images that depict the location of groundwater features such as wells and aquifers, WFS for accessing well-structured descriptions of such features, and SOS for accessing related groundwater measurements. Together, these services and underlying technologies provide a platform neutral and common approach to accessing vast amounts of data.
- **GIN and NGWMN syntax interoperability:** inter-network syntax interoperability is achieved through the deployment of (1) GML for data content and (2) the syntax associated with the WMS, WFS, and SOS web service interfaces for data access and query. Support for GML implies that the transfer of data for groundwater features and measurements, via the WFS and SOS web services,

respectively, is carried out using the Extensible Markup Language (XML) abstract syntax, consisting of a nested graph structure, and its relatively unconstrained concrete syntax. Alternate syntaxes, such as JavaScript Object Notation (JSON), are also variously supported for some feature types, for example, in GIN as a data encoding option for water wells. This diversity further extends to web services where alternate syntaxes such as those implied by REST, or other custom APIs, are deployed for some limited functions.

Intra-network syntax interoperability varies for the data networks: GIN utilizes GML as a canonical syntax between data sources and the data network, hence eliminates syntax heterogeneity for data, whereas NGWMN accepts a variety of non-standard syntaxes, e.g., unconstrained XML for water quality data, requiring source-specific syntax mappings between the data network and its syntactically variant data sources.

- **GIN and NGWMN structure interoperability:** is achieved for both inter- and intra-network interoperability, except where noted, via the deployment of domain-specific data schema that extend GML. These include WaterML2 for water time series accessed via SOS, Observations & Measurements (O & M; Cox 2013) for other types of measurements accessed via SOS, GWML1 for groundwater features accessed via WFS, and GeoSciML (Sen & Duffy 2005) for geological features accessed via WFS.

The common adoption of these schemas enables inter-network interoperability significantly, as specific queries and resulting data can be passed between the GIN and NGWMN networks without further alteration of structure. Minor exceptions include schemas for rock logs from water wells, which are structured heterogeneously and require translation between data networks.

Intra-network structure interoperability is more complex, due to the prevalence of structure heterogeneity between the data network and its sources of data from State or Provincial agencies. Overcoming such heterogeneity requires establishment of unidirectional mappings between each source and a canonical schema. GIN uses GWML1 as a canonical data structure, whereas NGWMN's internal database structure serves as its canonical schema for consuming data from sources. Both GIN and NGWMN use the OGC web service interfaces as their canonical query structures.

GIN's data mappings are expressed using EXtensible Style-sheet Language Transformation (XSLT), an XML transformation language, in combination with a declarative specification that equates XML schema components. XSLT plays a functional role, transforming a data structure from one schema to another using the equivalences between schema components expressed in the declarative specification. GIN query mappings are expressed in custom code, also drawing upon these equivalences. Schema interoperability within NGWMN is two-fold: individual data sources or clusters either support the canonical standards directly without need for translation, or source-specific mappings are constructed for each source schema.

- **GIN and NGWMN semantic interoperability:** is achieved through the translation of data values to terms defined in canonical vocabularies and expressed in RDF or its Simple Knowledge Organization System (SKOS) application. However, in contrast to previous levels, GIN and NGWMN do not adopt the same canonical vocabularies, as each implements their own distinct internal standards. These are used to achieve both inter- and intra-network semantic interoperability, as data values from all sources, including the other data network, are mapped to the internal vocabularies. Both data networks deploy a limited number of such vocabularies for key data types, consequently each network outputs semantically homogenous data for those data mapped to the vocabularies, and heterogeneous data for the rest. For example, GIN implements a rock type vocabulary into which are mapped rock type values from all source data, including NGWMN, while NGWMN implements internal vocabularies for rock types, units-of-measure, water quality constituents, and methods. As with the schema mapping, GIN's semantic mapping consists of two parts: a declarative component that maintains a list of equivalences between source and network vocabulary terms, and a functional component, expressed in XSLT and linked code, that carries out the translation.
- **GIN and NGWMN pragmatic interoperability:** is achieved via usage agreements, or profiles, developed between the data network and its data sources, including the other data network, thus facilitating both inter- and intra-network interoperability. For example, GIN and NGWMN have developed profiles for the

common deployment of WMS, WFS, and SOS web services for inter-network interoperability. This is particularly significant for the SOS service where, for instance, some optional aspects were designated as mandatory, e.g., a listing of measured parameters, and other general aspects were made more specific, e.g., WaterML2 as the sole output data structure. Indeed, realization of the broader need for such an SOS profile

in the water domain has prompted its recent and ongoing development within OGC ([GEOWOW Consortium 2014](#)). Intra-network pragmatic interoperability is similarly enforced by NGWMN through the development of distinct profiles for its data sources, and through the screening for incompatible methods of water quality data collection, to avoid incorrect usage of the data.

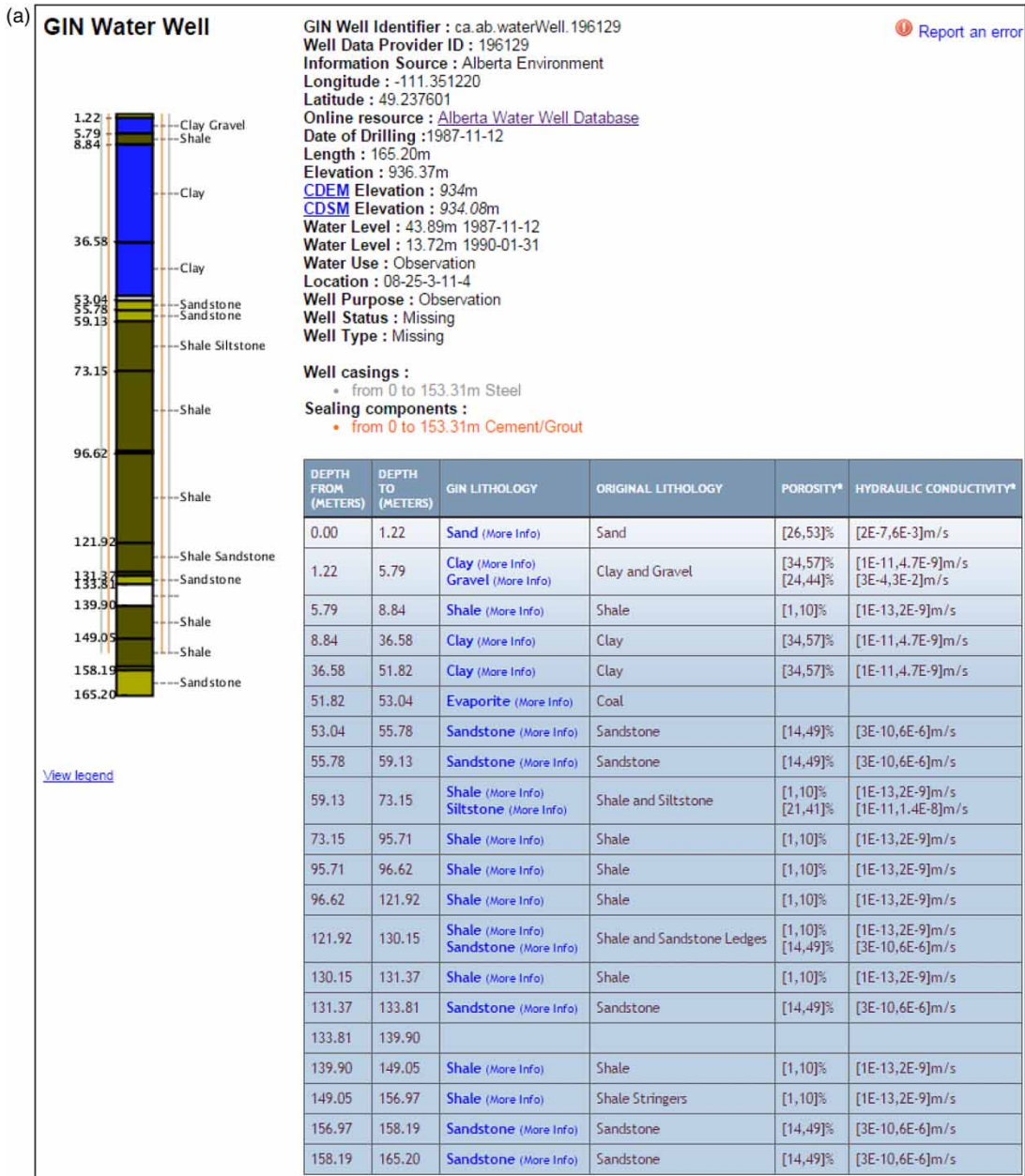


Figure 5 | (a) Water well data interoperability – GIN water well from Alberta. (b) Water well data interoperability – NGWMN water well from Montana. (continued)

DEMONSTRATION OF RESULTS

Data interoperability between and within GIN and NGWMN is illustrated in Figures 5 and 6, and as a consequence these figures also show that the data heterogeneity in Figures 1 and 2 is appropriately resolved. Figures 5(a) and 5(b) show results obtained using the GIN portal to view the water wells from Figure 1, demonstrating both inter- and intra-network interoperability: inter-network interoperability is shown as the Montana data are obtained by using GIN to query NGWMN, and intra-network interoperability is shown by GIN querying the Alberta data source. System and pragmatic interoperability are demonstrated as the results are obtained using standard web services (WFS) with local protocol restrictions. Syntax

interoperability is exemplified in two respects: through translation of both web service responses to the common GWML1 syntax (which is further translated to Hypertext Markup Language (HTML) for visualization), and through translation of rock type names into both English and French (not shown). Structure interoperability is demonstrated via the adoption of the GWML1 schema, including mapping of the original single-value depth interval from Alberta to a canonical two-value range. Semantic interoperability is illustrated via the mapping of rock type names onto terms in the GIN rock type vocabulary. Moreover, despite the differences in mediation between GIN and NGWMN, the response is virtually instantaneous for either well (i.e., <1 second) under a variety of internet connections, although performance can vary greatly for other data

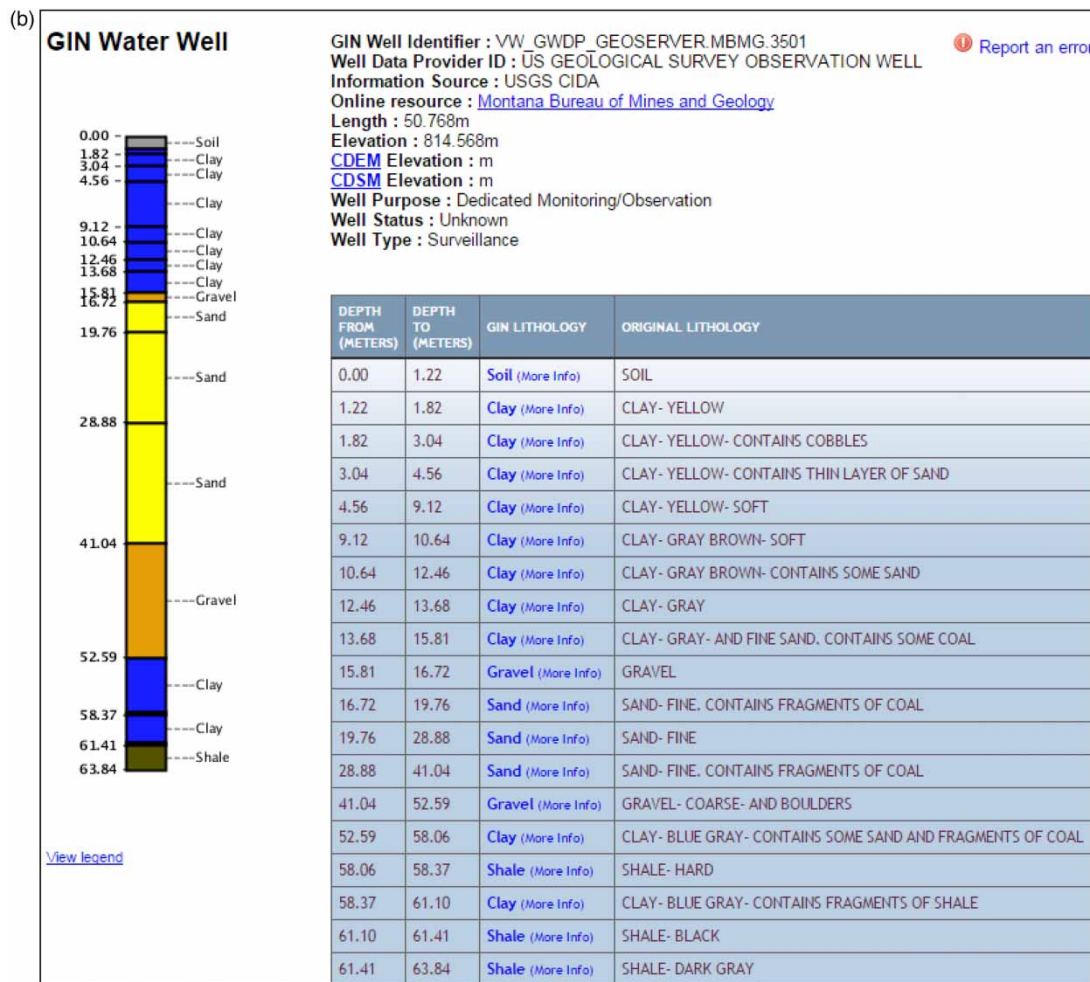


Figure 5 | continued.

sources due to local server conditions. Performance details for downloading multiple wells using GIN are described in Boisvert & Brodaric (2012), and range from ~2 seconds (50 wells) to 142 seconds (5,000 wells): this includes retrieval of data from two sources, translation across the five levels, and integration into a single file result.

Figure 6 illustrates inter- and intra-network interoperability of observational data, through retrieval of groundwater levels from the sites shown in Figure 2, which are located on either side of the international border. Inter-network interoperability is shown via GIN querying NGWMN for the Montana data, while intra-network interoperability is shown via GIN directly querying the Alberta data source. System and pragmatic interoperability is demonstrated via profiled use of SOS, syntax and structure interoperability is achieved via the adoption of WaterML2, and semantic interoperability is achieved through adoption of some common terms for measured properties, most notably for groundwater level. Observed response times, including the time to construct and display the graph, can vary from 1–5 seconds to 20–30 seconds, depending on the number of measurements and their cached status.

CONCLUSIONS AND FUTURE DIRECTIONS

Accompanying the rapid growth of online water data networks is an interoperability need: the quicker and larger such networks grow, the greater is the need to have them function in unison. All comprehensive solutions to this interoperability problem must align data at five levels, system, syntax, structure, semantics, and pragmatics, regardless of the adoption of any particular interoperability architecture. The GIN and NGWMN data networks exemplify how conformance to international standards and SDI architectures, local agreements about their implementation, and the development of shared vocabularies, can make large groundwater data networks interoperable at each level. An outstanding challenge remains at the semantic level, where very few reusable vocabularies have been established both within and between respective data networks, as well as in the international community overall. Another significant challenge involves architectural expansion, to complement the existing SDI architecture with LOD approaches. A specific

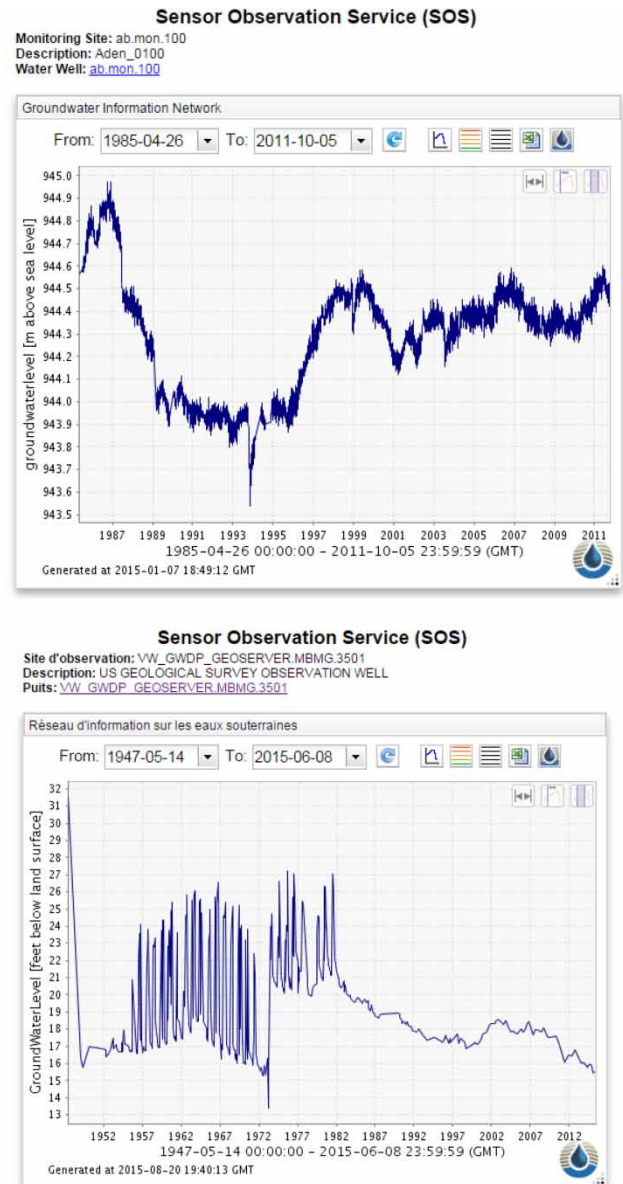


Figure 6 | Groundwater level data interoperability between GIN and NGWMN from Alberta (top) and Montana (bottom) data sources.

concern with LOD in the context of water data interoperability is the granularity problem: which resolution of data should be exposed as a web-indexed entity? For example, should each of the millions of sensors be exposed, or each of the billions of readings taken by those sensors? Another concern is the connectivity and scalability of data: how best to make linkages between data sources without centralizing the data into a common material (vs virtual) RDF repository? This is a problem in the groundwater domain

especially, because relevant information is often distributed among various databases maintained by different agencies: for example, information about a water well itself, as well as about related features such as sensors or aquifers, is rarely available from one database or agency, yet it is often essential to have the full suite of information at hand for key activities such as resource management. Despite these anticipated challenges, the current state of North American groundwater interoperability represents a significant achievement: the first edition of an interoperable USA–Canada groundwater data network.

ACKNOWLEDGEMENTS

The authors gratefully thank the various individuals and agencies that contributed to GIN and NGWMN, particularly to the owners of the data illustrated herein: Alberta Environment and Sustainable Resource Development, as well as the Montana Bureau of Mines and Geology and specifically Luke Buckley. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

REFERENCES

- Anzaldi, G. & Wu, W. 2014 Integration of water supply distribution systems by using interoperable standards to make effective decisions. In: *11th International Conference on Hydroinformatics, HIC 2014* (M. Piasecki, ed.). New York City, USA.
- Beran, B. & Piasecki, M. 2009 *Engineering new paths to hydrologic data*. *Comput. Geosci.* **35** (4), 753–760.
- Bermudez, L. & Arctur, D. 2011 *Water Information Services Concept Development Study*. OGC Engineering Report 11-p013r6, 83 pp.
- Bishr, Y. 1998 *Overcoming the semantic and other barriers to GIS interoperability*. *Int. J. Geogr. Inform. Sci.* **12** (4), 299–314.
- Boisvert, E. & Brodaric, B. 2012 *Groundwater Markup Language (GWML) – enabling groundwater data interoperability in spatial data infrastructures*. *J. Hydroinform.* **14** (1), 93–107.
- Brodaric, B. 2007 *Geo-pragmatics for the Geo-spatial Semantic Web*. *Trans. GIS* **11** (3), 453–477.
- Brodaric, B. & Booth, N. 2010 *OGC Groundwater Interoperability Experiment, Final Report*. OpenGIS Engineering Report 10-194r3, 48 pp.
- Brodaric, B. & Gahegan, M. 2006 *Representing Geoscientific Knowledge in Cyber infrastructure: challenges, approaches and implementations*. In: *Geoinformatics, Data to Knowledge* (A. K. Sinha, ed.). Geological Society of America Special Paper 397, Boulder, CO, pp. 1–20.
- Brodaric, B., Dabolt, T., Booth, N. & Vretanos, P. 2013 CHISP-1 pilot project introduces open architecture for watershed observatories. *Water News* **33** (1), 6–12.
- Brodeur, J., Bedard, Y., Edwards, J. & Moulin, B. 2003 *Revisiting the concept of geospatial data interoperability within the scope of human communication processes*. *Trans. GIS* **7** (2), 234–265.
- Broring, A., Stasch, C. & Echterhoff, J. 2012 *OGC Sensor Observation Interface Standard*. Open Geospatial Consortium Implementation Standard, (on line) 12-00.6 v2.0, 163 pp.
- Butler, H., Daly, M., Doyle, A., Gillies, S., Schaub, T. & Schmidt, C. 2008 *The GeoJSON Format Specification*. <http://geojson.org/geojson-spec.html> (accessed 15 January 2015).
- Cox, S. 2013 *Geographic information – Observations and measurements*. OGC Standard: Abstract Specification 10-004r3 v2.0, 54 pp.
- Curry, E., Degeler, V., Clifford, E., Coakley, D., Costa, A., Van Andel, S., Van De Giesen, N., Kouroupetroglou, C., Messervey, T., Mink, J. & Smit, S. 2014 *Linked water data for water information management*. In: *11th International Conference on Hydroinformatics, HIC 2014* (M. Piasecki, ed.). CUNY Academic Works, New York City, USA.
- Dahlhaus, P. G., MacLeod, A. D. & Thompson, H. C. 2012 *Federating hydrogeological data to visualise Victoria's groundwater*. In: *34th International Geological Congress: Proceedings* (I. Lambert & A. C. Gordon, eds). Australian Geoscience Council, Brisbane, Australia, p. 592.
- de la Beaujardiere, J. 2006 *OpenGIS Web Map Server Implementation Specification*. Open Geospatial Consortium Implementation Specification 06-042, v1.3.0, 85 pp.
- Dornblut, I. & Atkinson, R. 2014 *OGC HY_Features: A Common Hydrologic Feature Model*. Open Geospatial Consortium Technical Report 11-039r3, 55 pp.
- Duce, S. & Janowicz, J. 2010 *Microtheories for spatial data infrastructures – accounting for diversity of local conceptualizations at a global level*. In: *6th International Conference on Geographic Information Science (GIScience 2010)* (S. I. Fabrikant, T. Reichenbacher, M. J. van Kreveld & C. Schlieder, eds). Lecture Notes in Computer Science 6292, Springer, Berlin Heidelberg, pp. 27–41.
- Durbha, S. S., King, R. L., Shah, V. P. & Younan, N. H. 2009 *A framework for reconciliation of disparate earth observation data*. *Comput. Geosci.* **35** (4), 761–773.
- GEOOW Consortium 2014 *OGC Sensor Observation Service 2.0 Hydrology Profile*. In: *OGC Discussion Paper 14-004*, 36 pp.
- Government of Alberta. 1978 *Water Well Report*. <https://environment.extranet.gov.ab.ca/apps/GIC/Report/ViewReport.aspx?wellreportid=196129> (accessed 24 March 2014).
- Harvey, F., Jones, J., Scheider, S., Iwaniak, A., Kaczmarek, I., Lukowicz, J. & Strzelecki, M. 2014 *Little steps towards big*

- goals. Using linked data to develop next generation spatial data infrastructures (aka SDI 3.0). In: *17th AGILE Conference on Geographic Information Science, 2014*, Castellón, Spain.
- Heitmann, B., Cyganiak, R., Hayes, C. & Decker, S. 2014 Architectures of linked data applications. In: *Linked Data Management: Principles and Techniques* (A. Harth, K. Hose & R. Schenkel, eds). Chapman-Hall/CRC Press, Boca Raton, FL, pp. 4–26.
- Hogan, A., Umbrich, J., Harth, A., Cyganiak, R., Polleres, A. & Decker, S. 2012 [An empirical survey of linked data conformance](#). *J. Web Sem.* **14**, 14–44.
- Horrocks, I., Parsia, B., Patel-Schneider, P. & Hendler, J. 2005 Semantic web architecture: stack or two towers? In: *Principles and Practice of Semantic Web Reasoning* (F. Fages & S. Soliman, eds). LNCS 3703, Springer, Berlin, Heidelberg, Germany, pp. 37–41.
- INSPIRE 2008 *Drafting Team 'Data Specifications' – Deliverable D2.3: Definition of Annex Themes and Scope*. Technical Report D2.3_v3.0, INSPIRE, 132 pp.
- INSPIRE 2013 *D2.8.II.4 INSPIRE Data Specification on Geology – Draft Guidelines*. Technical Report D2.8.II.4_v3.p0 rc3, INSPIRE, 369 pp.
- ISO/TS 19150-1. 2012 *Geographic information – Ontology – Part 1: Framework*, Technical Specification, ISO/TS 19150-1:2012 (E), 30 pp.
- Jackson, S. R., Maidment, D. R. & Arctur, D. K. 2014 Towards a standardized river geometry format. In: *American Water Resources Association 2014 Spring Speciality Conference, GIS & Water Resources VIII: Data to Decisions*. 12–14 May 2014, Salt Lake City, UT, USA.
- Klug, H. & Kmoch, A. 2014 [A SMART groundwater portal: an OGC web services orchestration framework for hydrology to improve data access and visualization in New Zealand](#). *Comput. Geosci.* **69**, 78–86.
- Kuhn, W., Kauppinen, T. & Janowicz, K. 2014 Linked data – a paradigm shift for geographic information science. In: *Eighth International Conference on Geographic Information Science, GIScience 2014* (M. Duckham, E. Pebesma, K. Stewart & A. U. Frank, eds). LNCS 8728, Springer International, Switzerland, pp. 173–186.
- Lucido, J. & Booth, N. L. 2014 Improving groundwater data interoperability: results of the second groundwater interoperability experiment. In: *Proceedings, American Geophysical Union Annual Meeting*, 15–19 September 2014, San Francisco, CA, USA, IN31D-3752.
- Lutz, M., Sprado, J., Klein, E., Schubert, C. & Christ, I. 2009 [Overcoming semantic heterogeneity in spatial data infrastructures](#). *Comput. Geosci.* **35** (4), 739–752.
- Manso, M.-A., Wachowicz, M. & Bernabé, M.-Á. 2009 [Towards an integrated model of interoperability for spatial data infrastructures](#). *Trans. GIS* **13** (1), 43–67.
- Masser, I. 2010 *Building European Spatial Data Infrastructures*, 2nd edn. ESRI Press, Redlands, CA, USA, 108 pp.
- Montana Bureau of Mines, Geology. 1999 *Montana Well Log Report*. <http://mbmgwic.mtech.edu/sqlserver/v11/reports/SiteSummary.asp?gwid=3501&agency=mbmg&session=69042> (accessed 18 August 2015).
- Nativi, S., Mazzetti, P., Craglia, M. & Pirrone, N. 2014 [The GEOSS solution for enabling data interoperability and integrative research](#). *Environ. Sci. Pollut. Res.* **21** (6), 4177–4192.
- Perry, M. & Herring, J. 2012 *OGC GeoSPARQL – A Geographic Query Language for RDF Data*. OGC Standard 11-052r4, 75 pp.
- Portele, C. 2012 *OGC Geography Markup Language (GML) – Extended schemas and encoding rules*. Open Geospatial Consortium Implementation Standard 10-129r1, v3.3.0, 91 pp.
- Schade, S. & Smits, P. 2012 Why linked data should not lead to next generation SDI. In: *Geoscience and Remote Sensing Symposium 2012 Proceedings, 22–27 July 2012*, IEEE International, Munich, Germany, pp. 2894–2897.
- Sen, M. & Duffy, T. 2005 [GeoSciML: development of a generic Geoscience Markup Language](#). *Comput. Geosci.* **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 Academic Publishers, Boston, MA, USA, pp. 5–29.
- Stamper, R., Liu, K., Hafkamp, M. & Ades, Y. 2000 [Understanding the roles of signs and norms in organisations](#). *J. Behav. Inform. Technol.* **1** (1), 15–27.
- Tarboton, D. G., Maidment, D., Zaslavsky, I., Ames, D., Goodall, J., Hooper, R. P., Horsburgh, J., Valentine, D., Whiteaker, T. & Schreuders, K. 2011 [Data interoperability in the hydrologic sciences, the CUAHSI Hydrologic Information System](#). In: *Proceedings of the Environmental Information Management Conference 2011*, pp. 132–137.
- Taylor, P. 2012 *OGC WaterML 2.0: Part 1-Timeseries*. Open Geospatial Consortium Implementation Standard 10-126r3, 149 pp.
- Tolk, A., Diallo, S. & Turnitsa, C. D. 2007 Applying the levels of conceptual interoperability model in support of integratability, interoperability, and composability for system-of-systems engineering. *J. System. Cyber. Inform.* **5** (5), 65–74.
- Vretanos, P. A. 2005 *Web Feature Service Implementation Specification*. Open Geospatial Consortium Implementation Standard, 04-094, v1.1.0, 131 pp.
- Wache, H., Vogele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H. & Hubner, S. 2001 [Ontology-based integration of information – a survey of existing approaches](#). In: *Proceedings of the IJCAI'01: 17th International Joint Conferences on Artificial Intelligence*. Seattle, WA, USA, pp. 108–117.
- Yu, G. & Di, L. 2014 *OGC OWS-10 CCI Hydro Model Interoperability Engineering Report*. OGC Engineering Report 14-048, 64 pp.

First received 27 February 2015; accepted in revised form 16 September 2015. Available online 28 October 2015