International Conference on Hydroinformatics

2014

Levels Of Data Interoperability In The Emerging North American Groundwater Data Network

Boyan Brodaric

Nate Booth

Eric Boisvert

Jessica Lucido

LEVELS OF DATA INTEROPERABILITY IN THE EMERGING NORTH AMERICAN GROUNDWATER DATA NETWORK

BOYAN BRODARIC (1), NATE BOOTH (2), ERIC BOISVERT (3), JESSICA LUCIDO (4)

(1): Geological Survey of Canada, Ottawa, ON, Canada

(2): US Geological Survey, Reston, VA, USA

(3): Geological Survey of Canada, Quebec, QC, Canada

(4): US Geological Survey, Madison, WI, USA

The Canadian Groundwater Information Network (GIN) and the US National Ground-Water Monitoring Network (NGWMN) connect data from a variety of sources including states, provinces and federal agencies. Data heterogeneity is a major challenge faced by these networks, one that must be overcome at five distinct levels: systems, syntax, structure, semantics, and pragmatics. This paper discusses approaches taken at each of the five levels to ensure interoperability between the Canadian and American networks. The result is an emerging North American Groundwater Data Network, which enables users to access data transparently and uniformly on either side of the shared border.

1. INTRODUCTION

Canada and the USA share one of world's longest borders, crossed by many aquifers supplying water to both sides of the border. Cross-border groundwater management is therefore an important issue, one that is greatly facilitated by the efficient sharing of groundwater data. Such data sharing is generally complex, because it typically involves multiple agencies at the watershed, state / provincial, and federal levels, on each side of the border, for any one aquifer. This greatly amplifies the overall fragmentation and heterogeneity of the data, which makes cross-border aquifer data difficult to find and use. To ease overall data access, and overcome fragmentation and heterogeneity, federal agencies in Canada and the US have been developing data networks that provide a single access point to their distributed data sources: the Canadian Groundwater Information Network (GIN) and the US National Groundwater Monitoring Network (NGWMN). Recently, after mutual experimentation [3], the two networks have become interoperable. This has required both networks to implement international standards at five levels of interoperability: systems, syntax, structure, semantics, and pragmatics [2]. This paper describes the challenges and approaches to achieving interoperability at each level. Section 2 provides an example of data heterogeneity; Section 3 describes the interoperability levels; Section 4 briefly illustrates the implementations; and Section 5 concludes with a short summary and a brief indication of future directions.

2. GROUNDWATER DATA HETEROGENITY

Data heterogeneity is a by-product of the multiplicity of groundwater data providers in any region, in that each provider utilizes different database systems, data structures, transfer formats, etc. Figure 1 illustrates such heterogeneity: it shows part of a water well record from each side of the Canada-US border, i.e. from Alberta and Montana. Although Figure 1 is derived from online materials [8, 11], and is not a snapshot of the host systems, it is representative: the underlying database systems are not the same (systems difference); the character sets deployed can vary as Canadian sets often include French letters (syntax difference); the tabular organization is different (structure difference); the lithology vocabulary varies (semantics difference); and the units of measure for rock depths vary, i.e. meters vs feet, due to dissimilar construction practices (pragmatic difference). Alignment of each is required to ensure interoperability, ideally such that local databases remain unchanged and local data management practices remain undisrupted. To achieve this, alignment typically occurs through intermediary brokers, and GIN and NGWMN are thus examples of national-scale brokers.

Depth from ground level (m)	Water Bearing	Lithology Description
3.35		Silt
3.96		Gravel
47.24		Shale & Sandstone Ledges
54.86		Gravel
55.17		Boulders
82.30		Black Shale
82.91		Caprock
108.20		Sandstone

From	To	Description
0	25	BROWN SILT
25	152	GRAY SANDY SILT
152	158	SAND CLAY AND SEEP
158	182	SOFT GRAY SHALE
182	225	VARIEGATED SHALE
225	270	GRAY SHALEROCK
270	281	GRAY SANDY LIMESTONE
281	299	VARIEGATED SHALE WITH LIME STREAMER
299	312	GRAY AND BROWN SANDSTONE
312	320	BLACK SHALE

Figure 1: data heterogeneity between Canadian (AB, top) and US (MT, bottom) water well records.

3. LEVELS OF GROUNDWATER DATA INTEROPERABILITY

The open geospatial standards being developed at ISO and the Open Geospatial Consortium (OGC) are partial solutions to the heterogeneity problem, as they address three of the five interoperability levels, namely systems, syntax, and structure, leaving the remaining semantics and pragmatics levels open to local solutions. Of the three levels addressed by OGC, two are considered cross-disciplinary (systems, syntax) as they are applicable to any domain of interest. The structure level, however, is primarily uni-disciplinary, as it defines domain-specific data transfer schemas and formats. For example, the OGC Hydrology Domain Working Group is developing data transfer structures for water time series (WaterML2 [13]), surface water river networks (HY Features [7]), river channel descriptions (RiverML [9]), and groundwater features (GWML2, inspired by [1]). The remaining semantics and pragmatics levels are also typically uni-disciplinary, being almost always domain-specific. Figure 2 illustrates the five interoperability levels and the role of geospatial standards. These are variously implemented in GIN and NGWMN, as described next—to our knowledge, this is the first linkage of such national groundwater networks.

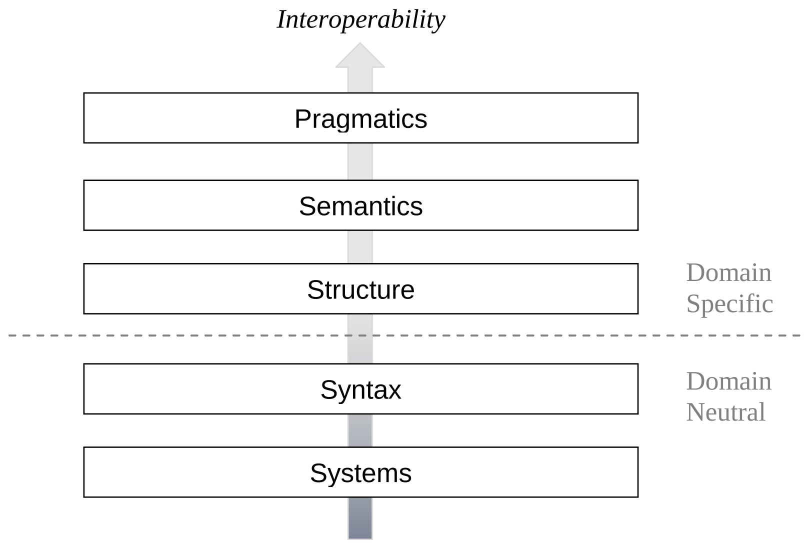


Figure 2: levels of data interoperability and the scope of standards

3.1 Systems Interoperability

System interoperability refers to the overcoming of infrastructure heterogeneities, such as variations in database systems, operating systems, and web protocols. It is addressed by OGC through the development of standard web services, which are a common online interface for accessing geospatial databases, and which hide the specifics of database implementation. The web services deployed by GIN and NGWMN include the Web Mapping Service (WMS) [6] for accessing map images, the Web Feature Service (WFS) [14] for accessing groundwater features such as wells, and the Sensor Observation Service (SOS) [5] for accessing water measurements. Together, these services provide a neutral and common approach to accessing vast amounts of data from either national network.

3.2 Syntax Interoperability

Syntax interoperability refers to the overcoming of syntactic heterogeneities, such as different character sets, e.g. Roman or Asian alphabets, or different data representation languages, such as XML or HTML. Syntax interoperability is achieved in the OGC suite of standards through the deployment of the Geographic Markup Language (GML) [12], which provides a language for the development of various domain-specific data structures. GIN and NGWMN deploy GML-based data structures as outputs from the WFS and SOS web services.

3.3 Structure Interoperability

Structure interoperability refers to the overcoming of heterogeneities related to how data is logically arranged, such as whether some piece of information is represented in one tabular column or many (e.g. well depth in Figure 1). Structure interoperability is manifest in the OGC suite of standards as domain-specific data transfer formats that extend GML. For example, GIN and GWMN deploy WaterML2 [13] for water time series observations, and GWML1 [1] for groundwater features such as aquifers and wells, pending completion of the new international GWML2 standard.

3.4 Semantic Interoperability

Semantic interoperability refers to the overcoming of heterogeneities related to the meaning of the data. It is typified by synonomy and polysemy: synonymy occurs when multiple words have the same meaning, and polysemy occurs when multiple meanings refer to the same word. A typical approach to synonymy and polysemy is the development of a controlled vocabulary with fixed single meanings for each term [4]. The vocabulary becomes an ontology when the meanings are formally represented and inter-related. GIN and NGWMN do not implement shared vocabularies nor ontologies at this time, though each network implements some specific national standards. For example, GIN implements a simple ontology for rock types, to enable interoperability between water well logs. Rock type terms from each data source, including NGWMN, are then mapped to this ontology to provide uniform terms and meanings in GIN.

3.5 Pragmatic Interoperability

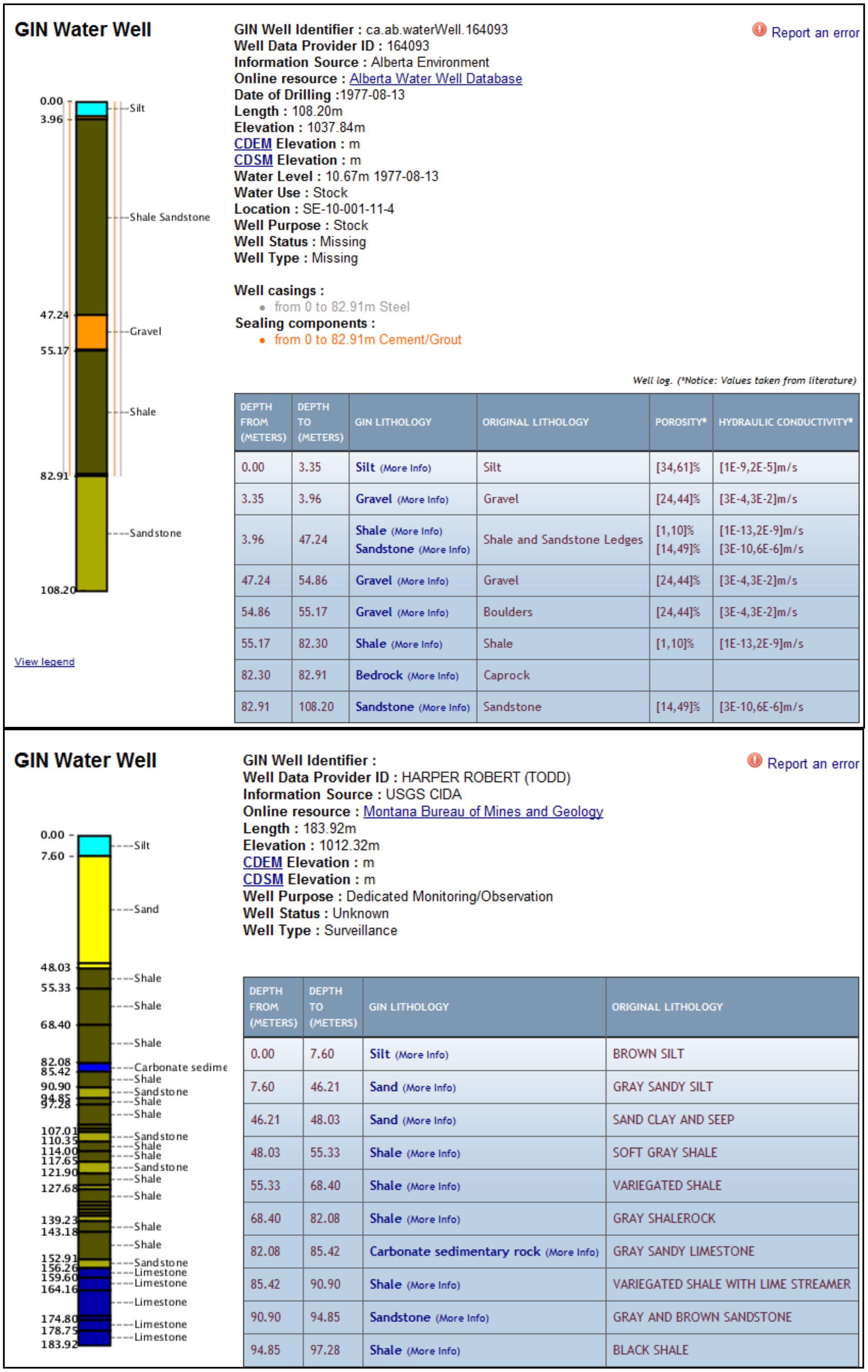
Pragmatic interoperability refers to guidelines about use, including best practices [2]. Such guidelines are important because interoperability can be impeded even if the remaining levels are all aligned. For example, data could be collected using a variety of scientific methods that might be incompatible thus prohibiting the merger of this data in online scenarios. Pragmatic interoperability is exemplified here by a specific profile describing the deployment of the web services by GIN and NGWMN. It is particularly significant for the SOS service where, for example, some optional aspects were designated as mandatory, e.g. a listing of measured parameters, and other general aspects were made more specific, e.g. WaterML 2.0 as the sole output data structure [10]. Indeed, realization of the broader need for such an SOS profile for the water domain has prompted its recent and ongoing development within OGC.

4. IMPLEMENTATION: GROUNDWATER DATA NETWORKS

Both GIN and NGWMN implement hybrid data architectures: data sources are distributed, but a cache of some data is maintained centrally, mainly for performance reasons. The networks differ in their extent of cache, as NGWMN pools all data in its US network, while GIN pools only some Canadian data, obtaining and translating the remainder (including US data) dynamically upon request. Cached data is harvested periodically using a mix of OGC and proprietary standards. Both networks expose OGC web services from a central national access point, and implementation of these services follows agreed profiles. Thus, systems, syntax, structure, and pragmatic interoperability is achieved through the profile-specific implementation of the same OGC standards at the central point of access for each network. However, due to the lack of standards for semantic groundwater interoperability, translation of vocabularies is handled variously within each system, and indeed, is only available from GIN at this time, and is limited to a national norm for rock types. To achieve such interoperability dynamically in a distributed environment, GIN additionally implements middleware known as a mediator [4]. Figure 3 illustrates retrieval of the water wells from Section 2 using the GIN portal—note the transformation to a common structure and translation of rock type terms to a standard vocabulary. Also noteworthy is the hiding of these interoperability mechanics from end-users, who see the Canadian or US data originating from a single virtual source.

5. SUMMARY AND FUTURE DIRECTIONS

Accompanying the rapid growth of online water data networks is an interoperability need: the quicker and larger that such networks grow, the greater is the need to have them function in unison. All comprehensive solutions to this interoperability problem will need to align data at five levels, system, syntax, structure, semantics, and pragmatics, regardless of the adoption of any particular interoperability architecture. The GIN and NGWMN networks exemplify how conformance to international standards, and local agreements about their implementation, can make large water networks interoperable at each level. The outstanding challenge remains at the semantic level, where very few vocabularies / ontologies have been established both within and between respective networks. Nonetheless the current state represents a significant achievement: the first edition of a North American groundwater data network.



GIN Water Well

GIN Well Identifier : ca.ab.waterWell.164093
Well Data Provider ID : 164093
Information Source : Alberta Environment
Online resource : Alberta Water Well Database
Date of Drilling :1977-08-13
Length : 108.20m
Elevation : 1037.84m
CDEM Elevation : m
CDSM Elevation : m
Water Level : 10.67m 1977-08-13
Water Use : Stock
Location : SE-10-001-11-4
Well Purpose : Stock
Well Status : Missing
Well Type : Missing

Report an error

Well casings :

- from 0 to 82.91m Steel

Sealing components :

- from 0 to 82.91m Cement/Grout

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

DEPTH FROM (METERS)	DEPTH TO (METERS)	GIN LITHOLOGY	ORIGINAL LITHOLOGY	POROSITY*	HYDRAULIC CONDUCTIVITY*
0.00	3.35	Silt (More Info)	Silt	[34,61]%	[1E-9,2E-5]m/s
3.35	3.96	Gravel (More Info)	Gravel	[24,44]%	[3E-4,3E-2]m/s
3.96	47.24	Shale (More Info) Sandstone (More Info)	Shale and Sandstone Ledges	[1,10]% [14,49]%	[1E-13,2E-9]m/s [3E-10,6E-6]m/s
47.24	54.86	Gravel (More Info)	Gravel	[24,44]%	[3E-4,3E-2]m/s
54.86	55.17	Gravel (More Info)	Boulders	[24,44]%	[3E-4,3E-2]m/s
55.17	82.30	Shale (More Info)	Shale	[1,10]%	[1E-13,2E-9]m/s
82.30	82.91	Bedrock (More Info)	Caprock		
82.91	108.20	Sandstone (More Info)	Sandstone	[14,49]%	[3E-10,6E-6]m/s

GIN Water Well

GIN Well Identifier :
Well Data Provider ID : HARPER ROBERT (TODD)
Information Source : USGS CIDA
Online resource : Montana Bureau of Mines and Geology
Length : 183.92m
Elevation : 1012.32m
CDEM Elevation : m
CDSM Elevation : m
Well Purpose : Dedicated Monitoring/Observation
Well Status : Unknown
Well Type : Surveillance

Report an error

DEPTH FROM (METERS)	DEPTH TO (METERS)	GIN LITHOLOGY	ORIGINAL LITHOLOGY
0.00	7.60	Silt (More Info)	BROWN SILT
7.60	46.21	Sand (More Info)	GRAY SANDY SILT
46.21	48.03	Sand (More Info)	SAND CLAY AND SEEP
48.03	55.33	Shale (More Info)	SOFT GRAY SHALE
55.33	68.40	Shale (More Info)	VARIEGATED SHALE
68.40	82.08	Shale (More Info)	GRAY SHALEROCK
82.08	85.42	Carbonate sedimentary rock (More Info)	GRAY SANDY LIMESTONE
85.42	90.90	Shale (More Info)	VARIEGATED SHALE WITH LIME STREAMER
90.90	94.85	Sandstone (More Info)	GRAY AND BROWN SANDSTONE
94.85	97.28	Shale (More Info)	BLACK SHALE

Figure 3: water well interoperability across the Alberta-Montana border, using provincial and state data sources.

ACKNOWLEDGEMENTS

The authors are grateful to 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 and the Montana Bureau of Mines and Geology.

REFERENCES

[1] *Boisvert, E., Brodaric, B. (2012) GroundWater Markup Language (GWML) – enabling groundwater data interoperability in spatial data infrastructures. Journal of Hydroinformatics, 14(1):93–107.*

[2] *Brodaric, B. (2007) Geo-pragmatics for the Geo-spatial Semantic Web. Transactions in GIS, 11(3):453-477.*

[3] *Brodaric, B., Booth, N. OGC Groundwater Interoperability Experiment, Final Report. OpenGIS Engineeging Report, 10-194r3, 48pp.*

[4] *Brodaric, B., Gahegan, M. (2006) Representing Geoscientific Knowledge in Cyberinfrastructure: challenges, approaches and implementations. In: Sinha, A.K. (Ed.), Geoinformatics, Data to Knowledge. Geological Society of America Special Paper 397, pp. 1-20.*

[5] *Broring, A., Stasch, C., Echterhoff, J. (2012) OGC Sensor Observation Interface Standard. Open Geospatial Consortium Implementation Standard,12-00,6 v2.0, 163pp.*

[6] *de la Beaujardiere, J. (2006) OpenGIS Web Map Server Implementation Specification. Open Geospatial Consortium Implementation Specification, 06-042, v1.3.0, 85pp.*

[7] *Dornblut, I., Atkinson, R. (2014) OGC HY_Features: a Common Hydrologic Feature Model. Open Geospatial Consortium Technical Report OGC 11-039r3, 55pp.*

[8] *Government of Alberta (1978) Water Well Report, https://environment.extranet.gov.ab.ca/apps/GIC/Report/ViewReport.aspx?wellid=164093&IsMetric=1, accessed 24 Mar 2014.*

[9] *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.*

[10] *Leinenweber, L. (2014) OGC CHISP-1 Summary Engineering Report. Open Geospatial Consortium Engineering Report, 13-046r2, 19pp.*

[11] *Montana Bureau of Mines and Geology (2006) Montana Well Log Report, http://mbmggwic.mtech.edu/sqlserver/v11/reports/SiteSummary.asp?gwicid=2394&agency=mbmg&session=690429, accessed 24 Mar 2014.*

[12] *Portele, C. (2012) OGC Geography Markup Language (GML) – Extended schemas and encoding rules. Open Geospatial Consortium Implementation Standard, 10-129r1, v3.3.0, 91pp.*

[13] *Taylor, P. (2012) OGC WaterML 2.0: Part 1-Timeseries. Open Geospatial Consortium Implementation Standard, OGC 10-126r3, 149pp.*

[14] *Vretanos, P.A. (2010) OpenGIS Web Feature Service 2.0 Interface Standard. Open Geospatial Consortium Implementation Standard,09-025r1, v2.0.0, 253pp.*