OGC Environmental Linked Features Interoperability Experiment Engineering Report

Table of Contents

Publication Date: YYYY-MM-DD

Approval Date: YYYY-MM-DD

Posted Date: 2018-11-21

Reference number of this document: OGC 18-097

Reference URL for this document: http://www.opengis.net/doc/PER/elfie-ID

Category: Public Engineering Report

Editor: David Blodgett, Byron Cochrane, Rob Atkinson, Silvain Grellet, Abdel Feliachi, Alistair Ritchie

Title: OGC Environmental Linked Features Interoperability Experiment Engineering Report

---

OGC Engineering Report

COPYRIGHT

WARNING

LICENSE AGREEMENT

Permission is hereby granted by the Open Geospatial Consortium, ("Licensor"), free of charge and subject to the terms set forth below, to any person obtaining a copy of this Intellectual Property and any associated documentation, to deal in the Intellectual Property without restriction (except as set forth below), including without limitation the rights to implement, use, copy, modify, merge, publish, distribute, and/or sublicense copies of the Intellectual Property, and to permit persons to whom the Intellectual Property is furnished to do so, provided that all copyright notices on the intellectual property are retained intact and that each person to whom the Intellectual Property is furnished agrees to the terms of this Agreement.

If you modify the Intellectual Property, all copies of the modified Intellectual Property must include, in addition to the above copyright notice, a notice that the Intellectual Property includes modifications that have not been approved or adopted by LICENSOR.

THIS LICENSE IS A COPYRIGHT LICENSE ONLY, AND DOES NOT CONVEY ANY RIGHTS UNDER ANY PATENTS THAT MAY BE IN FORCE ANYWHERE IN THE WORLD. THE INTELLECTUAL PROPERTY IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. THE COPYRIGHT HOLDER OR HOLDERS INCLUDED IN THIS NOTICE DO NOT WARRANT THAT THE FUNCTIONS CONTAINED IN THE INTELLECTUAL PROPERTY WILL MEET YOUR REQUIREMENTS OR THAT THE OPERATION OF THE INTELLECTUAL PROPERTY WILL BE UNINTERRUPTED OR ERROR FREE. ANY USE OF THE INTELLECTUAL PROPERTY SHALL BE MADE ENTIRELY AT THE USER'S OWN RISK. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR ANY CONTRIBUTOR OF INTELLECTUAL PROPERTY RIGHTS TO THE INTELLECTUAL PROPERTY BE LIABLE FOR ANY CLAIM, OR ANY DIRECT, SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES, OR ANY DAMAGES WHATSOEVER RESULTING FROM ANY ALLEGED INFRINGEMENT OR ANY LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR UNDER ANY OTHER LEGAL THEORY, ARISING OUT OF OR IN CONNECTION WITH THE IMPLEMENTATION, USE, COMMERCIALIZATION OR PERFORMANCE OF THIS INTELLECTUAL PROPERTY.

This license is effective until terminated. You may terminate it at any time by destroying the Intellectual Property together with all copies in any form. The license will also terminate if you fail to comply with any term or condition of this Agreement. Except as provided in the following sentence, no such termination of this license shall require the termination of any third party end-user sublicense to the Intellectual Property which is in force as of the date of notice of such termination. In addition, should the Intellectual Property, or the operation of the Intellectual Property, infringe, or in LICENSOR's sole opinion be likely to infringe, any patent, copyright, trademark or other right of a third party, you agree that LICENSOR, in its sole discretion, may terminate this license without any compensation or liability to you, your licensees or any other party. You agree upon termination of any kind to destroy or cause to be destroyed the Intellectual Property together with all copies in any form, whether held by you or by any third party.

Except as contained in this notice, the name of LICENSOR or of any other holder of a copyright in all or part of the Intellectual Property shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Intellectual Property without prior written authorization of LICENSOR or such copyright holder. LICENSOR is and shall at all times be the sole entity that may authorize you or any third party to use certification marks, trademarks or other special designations to

indicate compliance with any LICENSOR standards or specifications.

This Agreement is governed by the laws of the Commonwealth of Massachusetts. The application to this Agreement of the United Nations Convention on Contracts for the International Sale of Goods is hereby expressly excluded. In the event any provision of this Agreement shall be deemed unenforceable, void or invalid, such provision shall be modified so as to make it valid and enforceable, and as so modified the entire Agreement shall remain in full force and effect. No decision, action or inaction by LICENSOR shall be construed to be a waiver of any rights or remedies available to it.

None of the Intellectual Property or underlying information or technology may be downloaded or otherwise exported or reexported in violation of U.S. export laws and regulations. In addition, you are responsible for complying with any local laws in your jurisdiction which may impact your right to import, export or use the Intellectual Property, and you represent that you have complied with any regulations or registration procedures required by applicable law to make this license enforceable.

Chapter 1. Summary

Systems that maintain and disseminate information representing and/or related to spatial features often lack mechanisms to describe or discover how features relate to each other, to other kinds of features, and to a wide variety of related information that may be relevant. The Environmental Linked Features Interoperability Experiment (ELFIE) explored Open Geospatial Consortium (OGC) and World Wide Web Consortium (W3C) standards with the goal of establishing a best practice for exposing cross-domain links between environmental domain and sampling features. The Interoperability Experiment (IE) focused on encoding relationships between cross-domain features and linking available observations data to sampled domain features. An approach that leverages the OGC service baseline, W3C data on the web best practices, and JavaScript Object Notation for Linked Data (JSON-LD) contexts was developed and evaluated. Outcomes of the experiment demonstrate that broadly accepted web technologies for linked data can be applied using OGC services and domain data models to fill important gaps in existing environmental data systems' capabilities. While solutions were found to be capable and promising, OGC services and domain model implementations have limited utility for use in linked data applications in their current state and the universe of persistent URIs that form the foundation of a linked data infrastructure is still small. In addition to improvement of the standards baseline and publication of linked data URIs, establishing conventions for URI dereferencing behavior and default content given multiple options for a resource remain for future work.

1.1. Requirements & Research Motivation

Services and linked data for distributed spatio-temporal information on the internet hold the promise to increase interoperability and expressivity while decreasing duplication and data maintenance overhead. In order to achieve this potential, web service APIs and payloads need to work seamlessly with linked open data ontologies and hypermedia. Current OGC services, while flexible and capable, do not directly allow exposure of features in a REST-ful way or provide traversable hypermedia describing available methods, data, or interfaces to related (linked) content. Linked open data has the potential to turn the normal OGC service pattern (GetCapabilities request followed by introspection to understand contents) inside out by encoding associations between features as linked data predicates but specific implementation conventions and best practices are yet to be established.

Two core functional goals motivated the engineering described in this report: 1) The need to encode relationships between and among environmental features, and 2) the need to link observational data to the features they describe. Given that relationships in the absence of context are of little value, a third goal was the need for an ultimately general preview of a feature to provide just enough context in a ubiquitous form. These three functional goals coupled with a non-functional goal of using highly-adoptable technologies provided more than enough problem space for the interoperability experiment to take on.

1.2. Prior-After Comparison

Prior to initiating ELFIE, significant work had been completed on data models for numerous domains and types of data including hydrology, hydrogeology, soils, observations, and timeseries but almost never used together. The OGC services baseline was well implemented across the

community but practices for use of OGC services, primarily those based on the Web Feature Service (WFS) standard, were extremely varied and fully interoperable. ELFIE sought to provide implementation conventions as a target for 1) future data model encodings to integrate data across domains, 2) design and use of web services in a linked data context, and 3) design of future services and software to fill an important gap in open distributed environmental data systems.

At the conclusion of the project, technical best practices for encoding links between and among features and observations are in sight. A general approach to using the power of OGC web service APIs to expose features in the context of rich domain-feature-model-based linked data while following W3C best practices has been described. The solution requires additional work to establish and publish ontologies and points to a number of opportunities to evolve data and service standards to meet the needs of environmental features and observations. This work has begun or will be proposed under the OGC Naming Authority, relevant OGC and W3C Standards and Domain Working Groups, and a follow-on Second ELFIE (SELFIE) focused on URI dereferencing and default content is being planned.

1.3. Future Recommendations

The ELFIE provides recommendations and conventions for work in working groups across OGC and W3C. The WFS Standards Working Group (SWG)'s work on WFS 3.0 appears to be well-suited to ELFIE goals, but consideration for HTTP URIs to identify and query for features should be clearly supported by the future incarnation of WFS. Various OGC Working Groups need to establish ontologies and work through the OGC Naming Authority to publish their feature types and associations for broad use in linked data. Similarly, all data providers should pursue enterprise or community capabilities to mint and publish URIs and available representations of features and observations they own. Without this critical foundation, which can and should be recognized as a rudimentary starting point, linked data applications cannot move forward. Finally, the scope of ELFIE was purposefully limited to encoding of links in linked data documents. A future IE should look at an expanded scope to include URL dereferencing and default response content for URIs that identify real world features and content that links to various information resources that represent or are linked to the real-world non-information resource.

1.4. Document contributor contact points

All questions regarding this document should be directed to the editor or the contributors:

Table 1. Contacts

Name	Organization
David Blodgett	U.S. Geological Survey
Byron Cochrane	Land Information New Zealand
Ingo Simonis	Open Geospatial Consortium
Rob Atkinson	Metalinkage
Sylvain Grellet	Bureau de Recherches Géologiques et Minières

Name	Organization
Abdelfettah Feliachi	Bureau de Recherches Géologiques et Minières

1.5. Foreword

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

Chapter 2. References

The following normative documents are referenced in this document.

- OGC: OGC 06-121r9, OGC® Web Services Common Standard [https://portal.opengeospatial.org/files/?artifact_id=38867&version=2]

- OGC: OGC 11-052r4, OGC® GeoSPARQL - A Geographic Query Language for RDF Data [http://www.opengis.net/doc/IS/geosparql/1.0]

- OGC: OGC 14-111r6, OGC® WaterML 2: Part 3 - Surface Hydrology Features (HY_Features) - Conceptual Model [http://docs.opengeospatial.org/is/14-111r6/14-111r6.html]

- OGC: OGC 16-032r2, OGC® WaterML 2: Part 4 – GroundWaterML 2 (GWML2) [http://docs.opengeospatial.org/is/16-032r2/16-032r2.html]

- OGC: OGC 16-008, OGC® Geoscience Markup Language 4.1 (GeoSciML) [https://docs.opengeospatial.org/is/16-008/16-008.html]

- OGC: OGC 15-043r3, Timeseries Profile of Observations and Measurements [http://docs.opengeospatial.org/is/15-043r3/15-043r3.html]

- OGC: OGC 09-025r2, OGC® Web Feature Service 2.0 Interface Standard – With Corrigendum [http://docs.opengeospatial.org/is/09-025r2/09-025r2.html]

- OGC: OGC 12-006, OGC® Sensor Observation Service Interface Standard [https://portal.opengeospatial.org/files/?artifact_id=47599]

- ISO: ISO 19101-1:2014, Geographic information — Reference model — Part 1: Fundamentals [https://www.iso.org/standard/59164.html]

- ISO: ISO 19143:2010, Geographic information — Filter encoding [https://www.iso.org/standard/42137.html]

- ISO: ISO 19150-2:2015, Geographic information — Ontology — Part 2: Rules for developing ontologies in the Web Ontology Language (OWL) [https://www.iso.org/standard/57466.html]

- ISO: ISO 19156:2011, Geographic information — Observations and measurements [https://www.iso.org/standard/32574.html]

Chapter 3. Terms and definitions

For the purposes of this report, the definitions specified in Clause 4 of the OWS Common Implementation Standard OGC 06-121r9 [https://portal.opengeospatial.org/files/?artifact_id=38867&version=2] shall apply. In addition, the following terms and definitions apply.

3.1. feature

Abstraction of real-world phenomena. [ISO 19101:2002, definition 4.11].

3.2. property

Facet or attribute of an object referenced by a name [ISO 19143:2010, definition 4.21]

3.3. subject

The subject is the first part of an RDF statement. A subject in the context of a triple <?s ?p ?o> refers to who or what the RDF statement is about. (See https://www.w3.org/TR/ld-glossary/#subject [1])

3.4. predicate

The middle term (the linkage, or "verb") in an RDF statement. For example, in the statement "Alice knows Bob" then "knows" is the predicate which connects "Alice" (the subject of the statement) to "Bob" (the object of the statement). (See https://www.w3.org/TR/ld-glossary [1])

3.5. Object

In the context of RDF, the object is the final part of an RDF statement. (See https://www.w3.org/TR/ld-glossary/#object [1])

3.6. resource

A resource represents anything that can be identified, including physical things, documents, abstract concepts, numbers and strings (see https://www.w3.org/TR/rdf11-concepts/ [2])

3.7. observation

Act of measuring or otherwise determining the value of a property [ISO19156:2011, definition 4.11]

3.8. (graph) view

specific (simplified and useful) subset of a rich (complete but complicated) linked data graph

3.9. non-information URI

A URI for a non-information resource, which is a web resource that cannot be transmitted electronically. A URI is defined as a non-information URI when a HTTP Get Request does not return a 2xx "Success" response when dereferenced. No data are returned because the URI represents something other than a web information resource. i. e. a person or object.

3.10. watershed | catchment

A physiographic unit defined by a common hydrologic outlet and potentially defined input(s). May have several representations per OGC 14-111r6 HY_Features [http://docs.opengeospatial.org/is/14-111r6/14-111r6.html]. Note that "watershed" translations may be problematic (e.g. translating the term into German leads to two valid but different terms, i.e. "Wasserscheide", which is the actual drainage divide, and "Einzugsgebiet", which is the area. The definition here encompases both.).

Chapter 4. Abbreviated terms

- API - Application Programming Interface

- ELFIE - Environmental Linked Features Interoperability Experiment

- GeoSciML - Geoscience Markup Language

- GeoSPARQL - Geospatial SPARQL Protocol and RDF Query Language (recursive acronym)

- GWML2 - WaterML2 Part 4: Groundwater Markup Language

- HTTP - Hypertext Transfer Protocol

- HY_Features - WaterML2 Part 3: Surface Water Hydrologic Features

- IE - Interoperability Experiment

- JSON-LD - JavaScript Object Notation for Linked Data

- O&M - Observations and Measurements

- OWL - Web Ontology Language

- OWS - OGC Web Services Common

- RDF - Resource Description Framework

- REST - Representational State Transfer

- SHACL - Shapes Constraint Language

- SOS - Sensor Observation Service

- SOSA - Sensor, Observation, Sample, and Actuator

- URI - Uniform Resource Identifier

- URL - Uniform Resource Locator

- WaterML2 - Suite of Data Standards for the Hydrology Domain

- WFS - Web Feature Service

- UML - Unified Modeling Language

- W3C - World Wide Web Consortium

- WKT - Well Known Text

- XML - eXtensible Markup Language

Chapter 5. Overview

The Environmental Linked Feature Interoperability Experiment (ELFIE) has explored existing OGC and W3C standards with the goal of establishing a best practice for exposing cross-domain links between environmental domain and sampling features. The IE focused on two high-level functional goals:

1. The need to encode topological and domain specific relationships between cross-domain features

2. The need to link available observations data to sampled domain features.

> ### Box 1: Linked Data Graph Views
>
> While considering specific use cases, the IE has addressed issues of encoding specific (simplified and useful) views of a rich (arbitrarily complex) linked data graph that could potentially be known, discovered through links, and passed between systems. These graph views, which are italicized throughout for clarity of usage, support linked data architectures including catalogs and registries. For example, data providers could use these views as a best practice to expose their monitoring features and/or domain features to systems that traverse, harvest, and index available data. Similarly, integrated catalogs that index and construct links between features can use the views as a linked data response to queries. Views may correspond to concepts such as shapes (as per the W3C Shapes Constraint Language – SHACL), schema and profiles, however at this stage the formal correspondence has yet to be fully tested and articulated, and the relationships between these concepts themselves is not fully articulated. The W3C Data Exchange Working Group is currently developing a guidance document around the concept of profiles which may provide a useful baseline for formalizing these concepts in future.

Current OGC services, while flexible and capable, do not directly allow exposure of features in a REST-ful way or provide traversable hypermedia describing available methods, data, or interfaces to related (linked) content. The encodings demonstrated in the ELFIE turn the normal OGC service pattern (GetCapabilities request followed by introspection to understand contents) inside out by encoding associations between features as linked data predicates. This approach uses Resource Description Framework (RDF) triples (subject-predicate-object) to describe resource-relation-resource associations where the resources are unambiguous feature identifiers, but may resolve (dereference) to OGC web service requests as illustrated in the W3C/OGC Spatial Data On the Web Best Practices (Best Practice 1 [https://www.w3.org/TR/sdw-bp/#globally-unique-ids] [3]). This leverages the power of services such as WFS to provide feature instances, but expresses the structure of complex features or cross-service associated features explicitly.

The ELFIE has chosen to use JSON-LD as its encoding of RDF data. This was done in alignment and collaboration with the concurrent OGC work on JSON Best Practices [https://github.com/opengeospatial/architecture-dwg/tree/master/json-best-practice]. The ubiquity of JSON in Web development also contributed to the choice so as to ease implementing the RDF model by this community. The IE has specified a number of JSON-LD contexts [https://www.w3.org/TR/json-ld/#the-context] [4] that describe useful RDF views. JSON-LD documents that conform to these contexts can be delivered as the

payload returned from a REST API endpoint. In this case, they are self-describing through inclusion of the context document reference. They may also be negotiated, for instance using the view pattern described by the Linked Data API [https://github.com/UKGovLD/linked-data-api/blob/wiki/API_Query_Parameters.md]. (The DXWG is currently developing a formal standard for "content negotiation by profile" which confirms the relevance of this pattern.). Using this approach, the recommendation takes advantage of the power and flexibility of Linked Data architecture while providing just enough guidance to foster interoperability within a simplified implementation regime based on JSON. By describing a partially resolved useful view of a linked data graph, it also allows resources that are important but not yet available using Semantic Web formalisms to be incorporated into a linked data system. This mechanism should provide a gateway to inclusion in and understanding the value of the Linked Open Data Cloud [https://lod-cloud.net/].

The scope of this work could have easily gotten out of control, since the Linked Data and Semantic Web itself has never fully come to grips with the issue of pragmatic limits to graph shapes, non-unique naming and content-negotiation, nor has the Web Services world addressed linking objects and granularity of responses in a distributed environment. To focus the work, the IE attempted to avoid issues not directly related to RDF encodings as would be passed from server to client via http to support the two high level abstract use cases above.

5.1. IE Summary

The ELFIE tested the assumption that broadly adopted web technologies for linked data could be applied in the environmental domain to satisfy the needs of a number of compelling use cases. The Objectives section describes engineering design goals and details the assumptions and scope that were used to constrain the magnitude of the IE. Much of the IE involved design discussions amongst participants where candidate use cases were considered and potential technologies and implementations were explored. The Use Case Summary section describes these use cases and the select set that were more thoroughly tested to evolve the technologies that were agreed upon during design discussions.

The ELFIE technology selection criteria were that they be adoptable, common, and suitable to the needs of the ELFIE use cases. In some cases, new and thus lightly tested domain feature models were applied with the chosen, well established, technologies. The Applicable Best Practices and Standards section is a brief discussion of the technologies and standards tested in the ELFIE. The technology chosen to provide a precise but not overly prescriptive description of the RDF views is the JSON-LD context. Details of the contexts defined in the ELFIE are discussed in the Relations and JSON-LD Contexts section.

The outcomes of the ELFIE provide guidance for creating (graph) views of environmental features and observations. A simple view intended to be a preview, suitable for basic indexing is specified. Views intended to provide topological as well as observational and domain feature model relationships were also explored. The Preview view was largely successful, but there was no one satisfactory solution for expression of a preview geometry because GeoJSON is incompatible with JSON-LD, WKT is thought to be difficult to work with in typical web environments, and the schema.org geometries were found to have limited utility. The experimentation with domain feature models led to the conclusion that the underlying linked data technologies are mature enough, but progress is still needed on domain feature ontologies and encoding geometry. These issues are discussed in more detail in the Summary of Experiment and Analysis of Outcomes and

Issues and Recommendations for Future Work sections.

The ELFIE explored a number of interesting but tangential topics. Annex A: Strategies for Publishing Domain Ontologies as Linked Data from OGC domain standards is a summary of one important topic discussed; how to generate a feature ontology from a UML model. Annex B: Default Response for Multiple Representions and Relations is a brief discussion of issues related to what the default response should be when resolving a URI for a non-information resource that may have multiple representations.

Chapter 6. Objectives

The ELFIE objectives were to create and demonstrate a reusable framework that met the following functional and operational goals. The IE participants took a positive view (assumption)—that the below goals are possible and sought to test this assumption through design and implementation of compelling test cases.

6.1. Goals

6.1.1. Functional Goals

Functional goals of the ELFIE focused on providing

1. the ability to describe and use links between features that adhere to domain-specific feature models and

2. the ability to link the above features to observations and measurements data from samples collected on these features.

6.1.2. Operational Goals

Within this, the ELFIE aimed to leverage existing standards and best practices while maintaining a high degree of simplicity. This led to IE participants to seek operational solutions that:

1. are adoptable by developers and users with a wide array of technical skill levels,

2. are ubiquitous and easily adoptable in any programming language,

3. support mass market search techniques, (SDW Best Practice 2 [https://www.w3.org/TR/sdw-bp/#bp-indexable]: Make your spatial data indexable by search engines [3])

4. apply best practices for linked data on the web, (SDW Best Practice 3 [https://www.w3.org/TR/sdw-bp/#bp-linking]: Link resources together to create the Web of data [3])

5. and adhere, to the extent possible, to standard taxonomies and ontologies.

6.2. Assumptions

1. It is possible and desirable to implement a common approach to encoding links between environmental domain features which allow cross-domain and cross-system sharing and interoperability of such linked information

2. Existing and pending OGC standards for the encoding of environmental observation data in an integrated dataset of features can be linked according to RESTful and Linked Data principles

3. The ability to encode documents containing links between monitoring sites and environmental domain features in a common way will enable automation and lines of inquiry that are not possible without manual intervention today.

4. A reusable approach to encode documents that use these information models in cross-disciplinary applications is doable even while currently not apparent given existing encodings, implementation standards, and best practices,

6.3. Scope

1. To focus this IE, only environmental domain models concerning landscape interactions with the hydrologic cycle were considered. This included surface water, groundwater, well/borehole and soil moisture. (WaterML2, GroundwaterML 2, GeosciML 4 and SoilML)

2. Focus was on:

   a. linked data documents that contain collections of linked features and related observations and

   b. feature or observation representation documents that contain links to related features or observations.

3. The IE constrained itself to using existing ontologies and vocabularies wherever possible.

4. Problems regarding network architecture for resolving links and systems design and governance for applications that store and retrieve links or concept relationships were deemed out of scope.

5. Issues related to default behavior when dereferencing an identifier, while of great importance to the ELFIE goals, was also declared out of scope of this IE.

6. Discovery of available views and negotiation of choice between client and server were deemed out of scope, although the ability to define a view is seen as a key stepping stone to such options in the future.

6.4. Experiment Summary

To test the assumptions described above, a number of example test cases (described below) were used to guide the design and testing of potential technical solutions. The use cases were hypothetical, clearly of value to the domain, and technically difficult using existing technologies and practices. Discussion of use cases was predicated on the assumption that these would be implemented as test cases under the IE. While time and resources did not allow all these use cases to be implemented for the ELFIE, the use cases provided invaluable context and example systems to guide discussion and vetting of design alternatives.

The specifics of technologies and practices the IE tested are described below. With a clear direction, participants implemented test data and code to exercise and demonstrate the IE outcomes. Application of the use cases considered in the design phase led to some refinement to the design, but the design selected largely satisfied the goals of the IE with some clear next steps and further work needed to formalize data models, encodings, and best practices. The experiment succeeded in identifying shortcomings of some existing practices and gaps in the standards base line that need to be filled. Putting further work and needed new work aside, the IE failed to find significant or fundamental problems with the best practices and standards tested.

Chapter 7. Use Case Summary

7.1. Introduction

The design phase of the ELFIE focused on identification of a number of use cases with extensive need for linked environmental data. These use cases focused on known difficulties of the systems the ELFIE participants maintain or depend on as users. Many of these use cases were used for the discovery and vetting of ideas only. Others were further developed with implementation of example encoded linked data and client code to produce desired results of the use case.

Test cases that were implemented as demonstrations of the design outcomes are described below in full detail to include notional diagrams of how datasets involved link to each other, a thorough description of the use case, and an example application that speaks to and/or satisfies the needs of the use case.

The full suite of use cases considered during the design phase of the project are described in the Use Cases Considered in Design Process appendix. Each use case takes the form of a simple description of who or what system has an interest in what related datasets. Specific datasets are listed to clarify the scope and provide a starting point for future technical design work.

7.2. Test Cases Implemented as Demonstrations of Outcomes

Water budget summary: This use case provides a person interested in a basic summary of the water budget for a given watershed information about a collection of watersheds and their water budget data. It links together various hydrographic representations of each watershed as well as observational water budget data and related web resources.

As shown in the Figure 1, the watershed feature (black) is linked to various representations of it or information characterizing it (white). This follows the HY_Features HY_Catchment concept (Watershed in figure 1) as an unrealized feature that is related to various realization features, there is no canonical representation of the watershed itself.
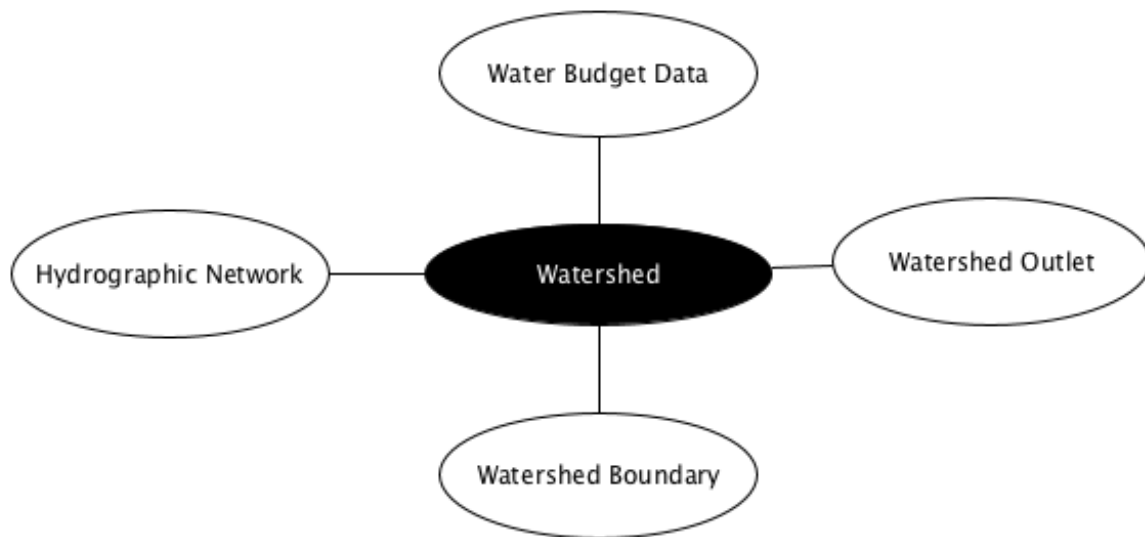
Figure 1. Notional diagram of relationships between watershed feature (black) and linked representations and water budget data (white) in the water budget summary use case.

Additional details about the water budget summary use case implementation is available on the demonstration ELFIE web page. [https://opengeospatial.github.io/ELFIE/demo/uswb]

Flood risks and impacts: This use case provides a decision maker needing to respond to flooded transportation infrastructure the information they need to understand the impacted assets and flooded roads for a forecast flood. Under this use case, a flood forecasting system would be able to discover vulnerable infrastructure and assets published by local jurisdictions as linked data and publish flood forecasts that include potentially impacted features in forecast information products.

Figure 2. Notional diagram of features (black) and linked data (white) in the floodcast use case. Note that many more relationships could be expressed if desired.

Additional details about the flood risk use case implementation is available on the demonstration ELFIE web page. [https://opengeospatial.github.io/ELFIE/demo/floodcast]

Ground water level monitoring: This use case is meant to demonstrate how from a given well URI any user (domain expert, machine) can then traverse to the monitoring strategy deployed (piezometer information etc.) and then access ground water level time series and/or information about the monitored aquifer.
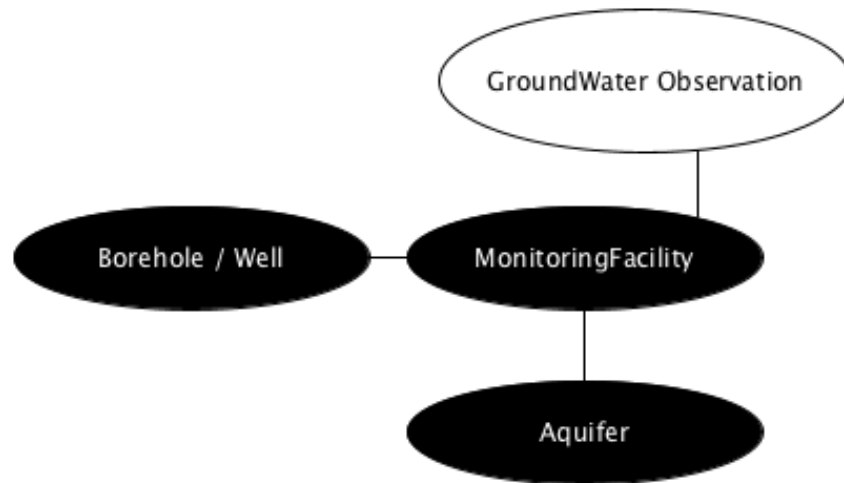
Figure 3. Notional diagram of features (black) and data (white) in the ground water level monitoring use case.

Additional details about the ground water level monitoring use case implementation is available on the demonstration ELFIE web page. [https://opengeospatial.github.io/ELFIE/demo/groundwater_monitoring]

Surface-ground water networks interaction: This use case is meant to demonstrate how, from a given Piezometer URI, any user (domain expert, machine) can traverse to the ground water monitoring strategy (see Ground water level monitoring Use Case) but also to the associated surface water monitoring one. Provided each surface/groundwater feature is properly linked together (River network, Aquifer system), it is then feasible to discover information about the full, comprehensive water system. This use case can be seen as a flagship demonstration of the usefulness of linked data in the environmental/cross-domain context.

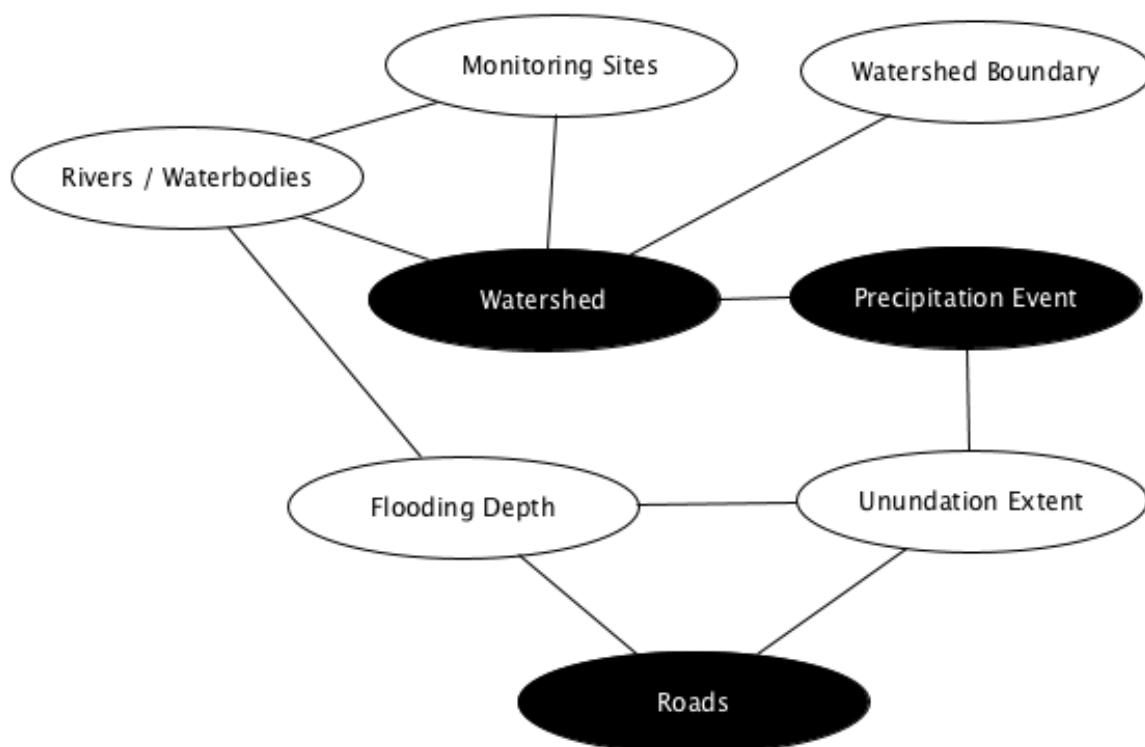*Figure 4. Notional diagram of features (black) and linked observational data (white) in the surface-ground water networks interaction use case.*

Additional details about the surface-ground water networks interaction use case implementation is available on the demonstration ELFIE web page. [https://opengeospatial.github.io/ELFIE/demo/surface_groundwater_network_interaction]

Watershed data index: This use case is meant to demonstrate the use of HY_Features to link a catchment to the data representing it as well as the monitoring network associated with it. It serves as a general demonstration that could be used for a wide array of linked watershed information use cases.
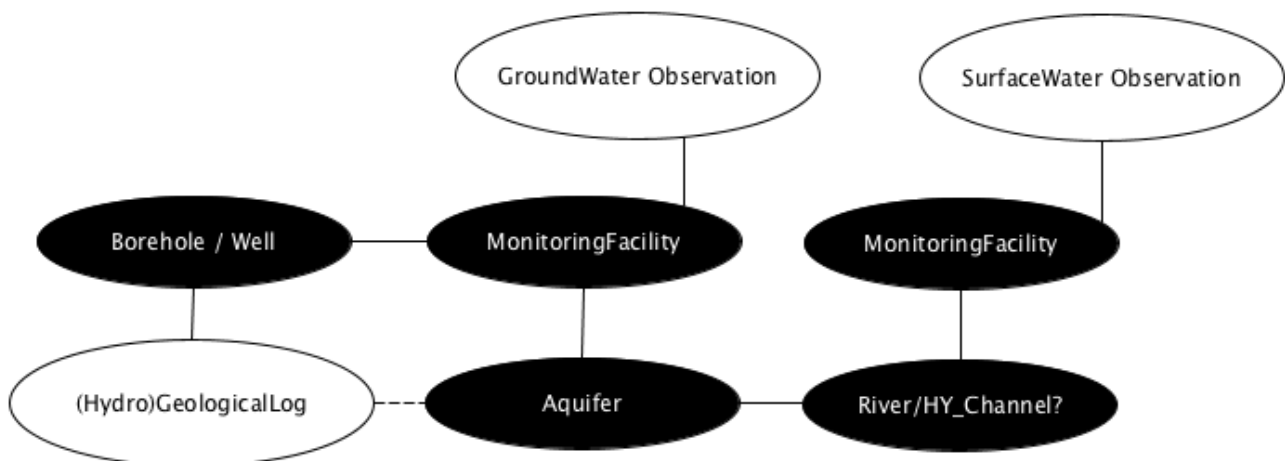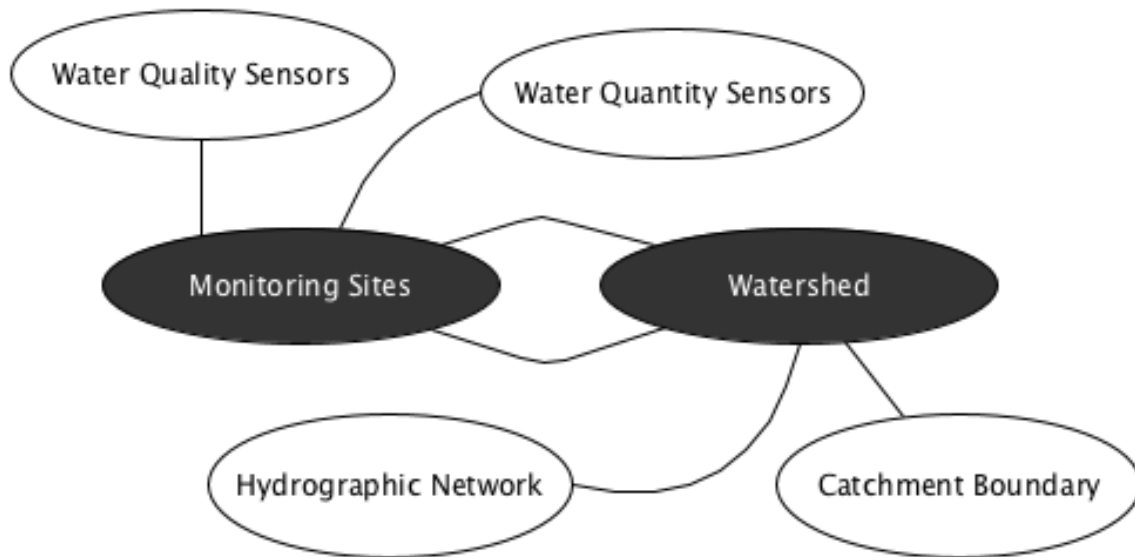

Figure 5. Notional diagram of relationships between the features (black) and linked data (white) in the watershed data index use case.

Additional details about the watershed data index use case implementation is available on the demonstration ELFIE web page. [https://opengeospatial.github.io/ELFIE/demo/huc12obs]

7.3. Watershed Data Index Use Case in Depth

This use case is introduced in more detail than those above and its technical details are presented below. Technical details of other use cases can be found at the ELFIE demonstration web page. [https://opengeospatial.github.io/ELFIE/]

The watershed data index use case is focused on a single HY_Catchment feature with an identifier of "070900020601" from the U.S. watershed boundary dataset. Given that HY_Catchment is an unrealized feature type, the document describing "070900020601" [https://opengeospatial.github.io/ELFIE/usgs/huc/huc12obs/070900020601] links to realizations of "070900020601". Three catchment realizations are included:

1. the boundary [https://opengeospatial.github.io/ELFIE/usgs/hucboundary/huc12obs/070900020601] polygon from the watershed boundary dataset,

2. the hydrographic [https://opengeospatial.github.io/ELFIE/usgs/nhdplusflowline/huc12obs/070900020601] network from the National Hydrography Dataset,

3. and the hydrometric [https://opengeospatial.github.io/ELFIE/usgs/hydrometricnetwork/huc12obs/070900020601] network, a collection of monitoring sites in the catchment.

A more complete implementation could include multiple versions of any of these feature types as well as additional realization types such as a network of channels, or a network of sub-catchments. The boundary polygon and hydrographic network are geospatial features that only link back to the catchment they realize. The hydrometric network is a more complex feature that aggregates a set of network stations, each of type HY_HydrometricFeature.
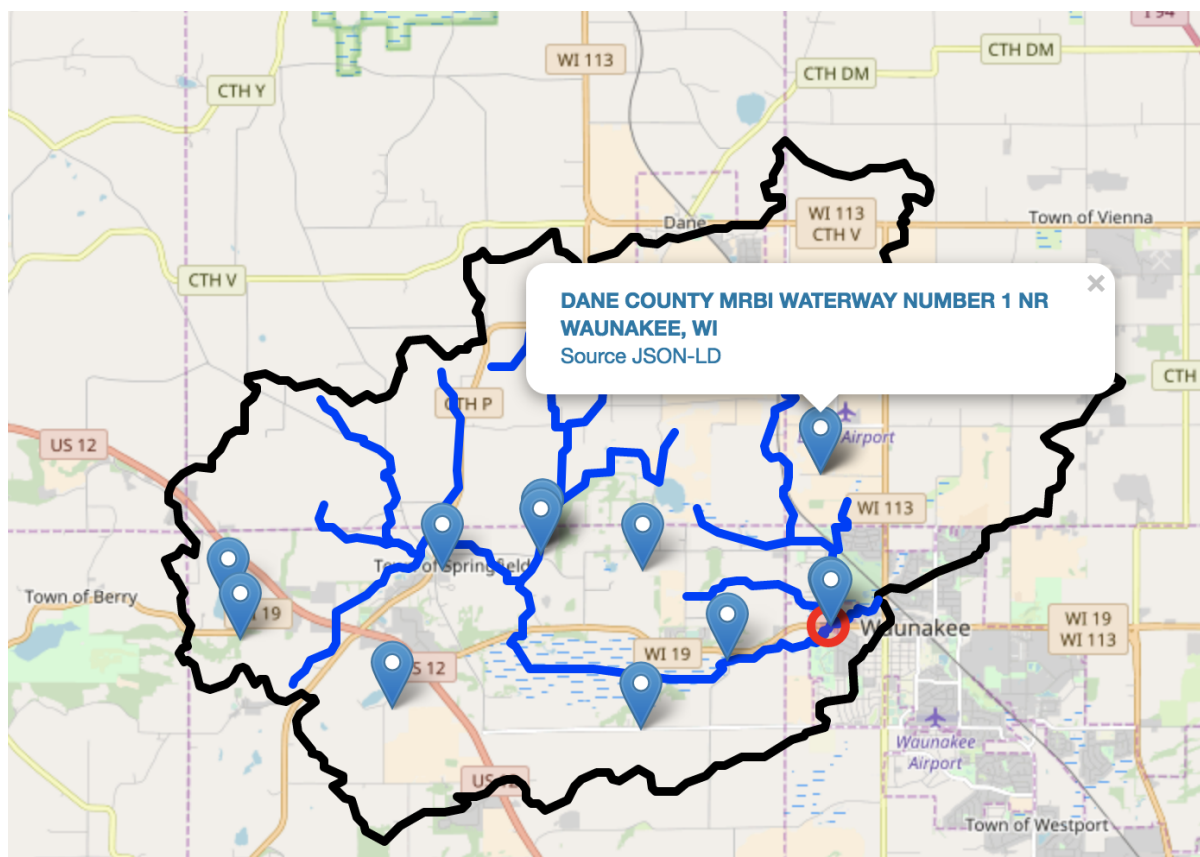
Figure 6. Screenshot of watershed data index use case.

7.4. Surface-ground water networks interaction Use Case in Depth

This use case is introduced in more detail than those above and its technical details are presented below as it makes intensive use of linked data technologies.

The Surface-ground water networks interaction use case is focused on a single Piezometer with an identifier from the French Ground Water Information Network (00463X0036-H1.2). From its JSON-LD description it is possible to dereference URIs of observations, the monitored Hydrogeounit, and also the associated Stream gage. In hydrological contexts where surface water - groundwater interactions are properly described, groundwater and surface water monitoring stations can be 'associated' with correlation coefficients as behavior of their monitored featured has an impact on the other. The associated Stream gage description is also linked to the river network.

The implementation use case tested during ELFIE allowed traversal of the data graph depicted below. Using a dedicated application, it was possible to interact with it: either displaying on maps geographical features or triggering observation display widgets (timeseries).

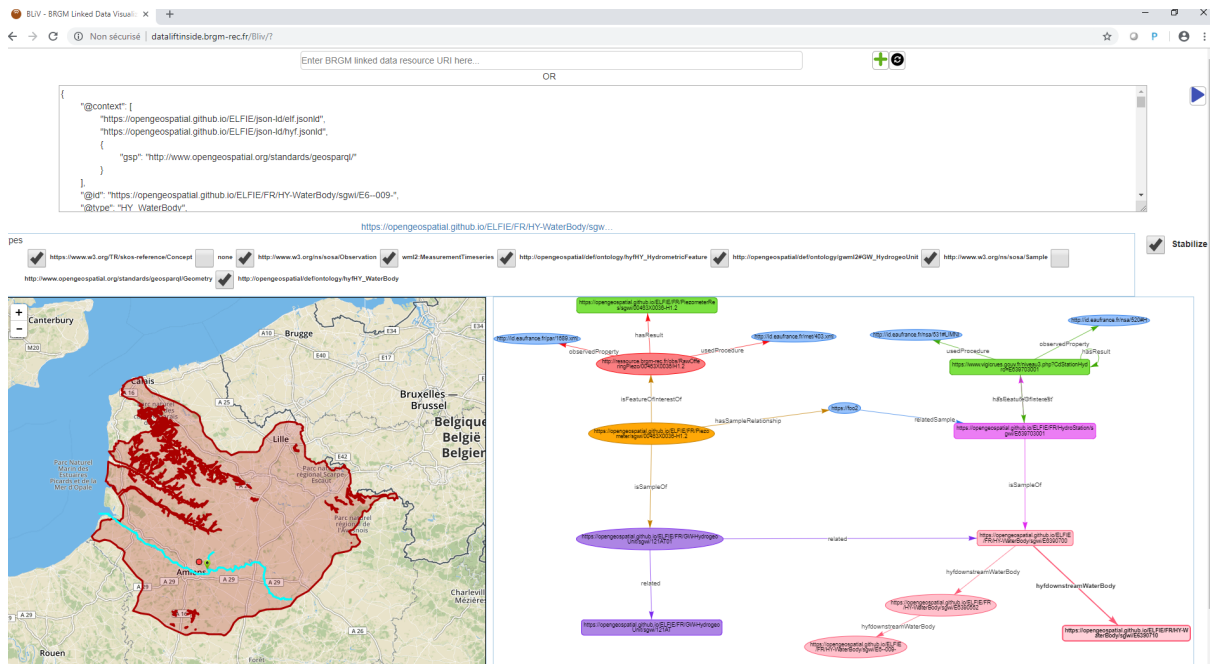*Figure 7. Screenshot of the surface-ground water networks interaction use case (map visualization data graph traversed).*
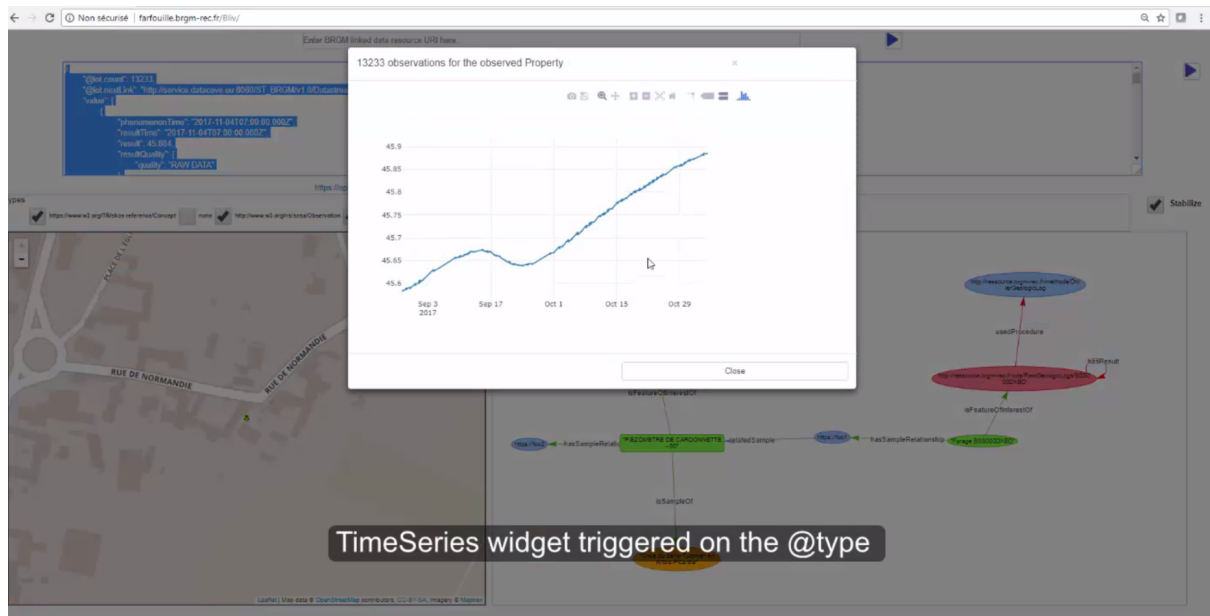

Figure 8. Screenshot of the surface-ground water networks interaction use case (widget visualizations of dereferenced observations).

Chapter 8. Applicable Best Practices and Standards

8.1. Introduction

The approach taken in the ELFIE was limited in contrast to recent IEs concerning environmental information models. The ELFIE focused on linked data documents that contain collections of linked features and related observations. With regard to the OGC Reference Model, this IE seeks to understand how features (geospatial information, observations) can be linked to each other in specific resolved views (as in Box 1: Linked Data Graph Views) of the linked data graph or by reference through use of HTTP URIs. The ELFIE did not seek to solve problems regarding network architecture for resolving links or systems design and governance for applications.

Given this, the ELFIE sought to define a common approach to encoding links between environmental domain features that would allow cross-domain and cross-system sharing and interoperability of such linked information. To do this, the ELFIE worked with existing and pending OGC standards for encoding of environmental observation data in an integrated dataset of features. Such OGC data models were then linked together according to W3C and other Linked Data best practices.

8.2. OGC Web Standards

OGC Web Service standards for features and observations (WFS, Sensor Observation Service (SOS), and SensorThings API) were used to construct HTTP URLs for use in the ELFIE linked data documents as much as possible. While the WFS and SOS APIs present some challenges in creating URLs for use in linked data, the APIs do offer sufficient HTTP GET functionality to embed useful content in linked data with a great degree of flexibility. The situation is different with the SensorThings API, which has HTTP GET functionality, but only supports 'internal ids' to request desired features. Workarounds have been tried but they involve URL encoding features URIs which is not seen as an ideal solution. Issues related to URL structure and a given API-call's relation to a canonical non-information resource URI were encountered but deferred for future investigations.

8.3. OGC Domain Models

OGC data models were used extensively in the ELFIE. The Observations and Measurements (O&M) conceptual model provided a high-level organizing framework for most ELFIE documents. The OGC-W3C Spatial Data On the Web Working Group implementation of O&M (Sensor, Observation, Sample, and Actuator (SOSA)) [https://www.w3.org/TR/vocab-ssn/] was used directly because of its applicability to the linked data technology pursued in the ELFIE. The GeoSPARQL [https://www.opengeospatial.org/standards/geosparql] ontology was also used directly for representing geometries and spatial relations between features.

Domain specific data models such as HY_Features [http://docs.opengeospatial.org/is/14-111r6/14-111r6.html], GWML2 [http://docs.opengeospatial.org/is/16-032r2/16-032r2.html] and GeoSciML [https://docs.opengeospatial.org/is/16-008/16-008.html] were also used for feature types and relations in the

ELFIE linked data documents.

<div style="border:1px solid #ccc; padding:1em;">

Box 2: Use of Unformalized Identifiers

A major limitation for use of these data models in the ELFIE was a lack of published taxonomies and ontologies using stable namespaces. Placeholder URIs were minted for their feature types and relations for the ELFIE project and formalization of such taxonomies and ontologies was left for future work.

Not all URIs used in examples in this project are intended to be stable. Formal publication of these ontologies will supersede interim choices.

</div>

8.4. W3C

W3C best practices for data on the web, spatial data on the web, and linked data were used to guide decisions related to the ELFIE linked data documents. Several members of the ELFIE project team were involved in the development of these best practices and represented them throughout the project. Given this, the ELFIE was an opportunity to apply these best practices in a practical engineering process, using them together and attempting to apply them to OGC web service and information models.

This was largely successful and no major issues were encountered related to the best practices themselves. However, they were not achieved fully due to realities of the internet, organizational operational constraints, and limitations of current APIs and data models. In other words, a project seeking to apply current OGC standards on their infrastructure within the suggested W3C data on the web best practices would have to implement ad hoc system behavior (proxy negotiation by 'profiles', feature request by its URI, rewriting to JSON-LD, etc.). This situation is mainly due to the fact that several standardization efforts needed to support these practices are in progress (W3C DXWG, OGC WFS 3, etc.) and that enterprise systems are slow to evolve.

Experiences from the ELFIE project are being taken forward into influencing future W3C standards through the Data Exchange Working Group, where the concept of views is being addressed within the scope of the profiles concept. Description, discovery, metadata and negotiation according to formally identified profiles mirrors the ELFIE architectural choice to publish and reuse JSON-LD contexts.

8.5. JSON-LD

Given the ELFIE goals related to adoptability of solutions and support of mass market search techniques, JSON-LD was used as the encoding for the ELFIE documents. While there were other potential options that could be used, the ease of use (parsing and serialization) of JSON-LD across practically all programming languages made it the clear choice for use in ELFIE. The decision to use JSON-LD was further supported by ongoing work within the OGC Architecture Domain Working Group to modernize OGC standards to supported JSON. As described below, some technical limitations inherent to RDF and thus JSON-LD were encountered while attempting to encode preview geometries, but all basic requirements of the ELFIE were satisfied by the JSON-LD encoding.

The incompatibility of JSON-LD and GeoJSON is a finding that will need further consideration by the OGC community.

Chapter 9. Relations and JSON-LD Contexts

9.1. Introduction

An important goal of the ELFIE was "adoptability" by web developers and compatibility with technologies that are ubiquitous on the internet. Given these goals, the IE team started to think in terms of *"views"* (concept introduced in Box 1: Linked Data Graph Views) of the linked data graph describing a given environmental feature or observation. Two conceptual *views*, "preview" and "network", were the primary focus of the ELFIE. "Preview" is meant to be a very simple summary of a feature including type, name, and geometry. "Network" is meant to be a summary of relevant first-neighbor features in a topological network. These two form the core interoperable set of ELFIE *views* but lack domain specific relations such as observational relations and hydrologic topology.

Initial ELFIE discussions of the graph view concept used the Shapes Constraint Language SHACL [https://www.w3.org/TR/shacl/] [5] technology as the logical implementation. Further evaluation and consideration of the ELFIE goals regarding adoptability and ubiquity, the JSON-LD context [https://www.w3.org/TR/json-ld/#the-context] was chosen as the most appropriate technical solution to express the graph *views*. The ELFIE contexts are meant to be precise descriptions of what the IE found to be the most appropriate linked data properties (attributes and relations) and types. They are not intended to be used as a schema to specify or validate the contents of a given document. However, they could provide a useful set of aliases for the subset of types and properties that could be expected to be parsed and interpreted by the client.

From a practical perspective, the contexts provide clear guidelines for what properties and types to use while not being overly prescriptive, limiting, or difficult to use in a typical web development workflow. Unlike schemas, such contexts support the "Open World Assumption" where missing elements simply indicate that this information is not currently available to the data provider, without the burden (and potential for misuse) of stating a placeholder value. Furthermore, to simplify the use of the contexts and reduce possible errors while parsing the *views*, Type Coercion [https://www.w3.org/TR/json-ld11/#type-coercion] [6] was not used while aliasing properties in the contexts files.

More information about the design and engineering of ELFIE contexts can be found on the ELFIE JOSN-LD web page [https://opengeospatial.github.io/ELFIE/json-ld/].

9.2. General ELFIE Contexts

The ELFIE "preview" graph *view*, is described by the elf.jsonld [https://opengeospatial.github.io/ELFIE/json-ld/elf.jsonld] context. It uses schema.org [https://schema.org/], skos [https://www.w3.org/TR/skos-reference/] [7], and geosparql [http://www.opengeospatial.org/standards/geosparql] properties and types. Schema.org is used whenever possible to maintain compatibility with search engines, uses SKOS for a lose "related" relationship, and GeoSPARQL for precise specification of a geometric preview. The geometric preview of a feature was an area of contention for the IE.

GeoJSON-LD [http://geojson.org/geojson-ld/] was seen as a logical solution, but as described in the outstanding issues for GeoJSON-LD, nested GeoJSON coordinate arrays are not supported by JSON-LD parsers. Additionally, the schema.org geometry schema was seen as under specified in that

among other reasons, it does not provide a default coordinate reference system or a mechanism to declare one. For this reason, the GeoSPARQL well known text format was included providing a precise and JSON-LD compatible geometry format. However, provided the technical limitations listed below, if some logical assumptions are made for the use of schema.org geometry (schema:GeoShape [https://schema.org/GeoShape]), it can be used to satisfy the basic geometry preview use case with significantly lower technical overhead than support for the full well known text standard. Further work should look more closely at this issue in an attempt to reconcile and provide guidance.

The ELFIE "network" *view*, is described by the elf-network.jsonld [https://opengeospatial.github.io/ELFIE/json-ld/elf-network.jsonld] context. It uses GeoSPARQL [http://www.opengeospatial.org/standards/geosparql] and OWL-time [https://www.w3.org/TR/owl-time/] relations. As shown below, its emphasis is on simple topological relationships such as touches, intersects, before, etc. A "network" context would logically include additional domain specific topological relations derived from other contexts, but they are not included in the general "network" context.

Box 3: *elf-network.json-ld* [https://opengeospatial.github.io/ELFIE/json-ld/elf-network.jsonld]

```
{"@context": {
        "gsp": "http://www.opengeospatial.org/standards/geosparql/",
        "time": "https://www.w3.org/TR/owl-time/",
        "intersects": "gsp:sfIntersects",
        "touches": "gsp:sfTouches",
        "within": "gsp:sfWithin",
        "after": "time:after",
        "before": "time:before",
        "intervalAfter": "time:intervalAfter",
        "intervalBefore": "time:intervalBefore",
        "intervalDuring": "time:intervalDuring"
    }}
```

9.3. Domain Specific Contexts

Two extended general contexts and five domain specific contexts were created for testing in ELFIE. The extended general contexts have content from the timeseriesML [https://opengeospatial.github.io/ELFIE/json-ld/tsml.jsonld] and SOSA [https://opengeospatial.github.io/ELFIE/json-ld/sosa.jsonld] data models. The domain specific contexts have content from HY_Features [https://opengeospatial.github.io/ELFIE/json-ld/hyf.jsonld], GWML2 [https://opengeospatial.github.io/ELFIE/json-ld/gw.jsonld], soilIE [https://opengeospatial.github.io/ELFIE/json-ld/soilie.jsonld], GeoSciML [https://opengeospatial.github.io/ELFIE/json-ld/gsml.jsonld], and the floodcast [https://opengeospatial.github.io/ELFIE/json-ld/floodcast.jsonld] use case. All of these contexts, except SOSA, were created as part of the ELFIE project without formal ontologies to back them. Work on these ontologies should follow ELFIE.

9.4. Contexts in context: exemple on the Watershed Data Index Use Case

As can be seen in the main watershed entry point to the index here [https://opengeospatial.github.io/

ELFIE/usgs/huc/huc12obs/070900020601], the watershed data index use case uses two contexts. elf.jsonld and and hyf.jsonld shown below in Box 4: elf.jsonld [https://opengeospatial.github.io/ELFIE/json-ld/elf.jsonld] and Box 5: hyf.jsonld [https://opengeospatial.github.io/ELFIE/json-ld/hyf.jsonld] respectively.

Box 4: elf.jsonld [https://opengeospatial.github.io/ELFIE/json-ld/elf.jsonld]

```
{"@context": {
  "schema": "http://schema.org/",
  "skos": "https://www.w3.org/TR/skos-reference/",
  "gsp": "http://www.opengeospatial.org/standards/geosparql",
  "description": "schema:description",
  "geo": "schema:geo",
  "hasGeometry": "gsp:hasGeometry",
  "asWKT": "gsp:asWKT",
  "image": {
    "@id": "schema:image",
    "@type": "@id"
  },
  "name": "schema:name",
  "sameAs": "schema:sameAs",
  "related": "skos:related"
}}
```

```
{"@context": {
  "hyf": "http://opengeospatial/def/ontology/hyf",
  "HY_Catchment": "hyf:HY_Catchment",
  "HY_CatchmentDivide": "hyf:HY_CatchmentDivide",
  "HY_HydrographicNetwork": "hyf:HY_HydrographicNetwork",
  "HY_HydrometricNetwork": "hyf:HY_HydrometricNetwork",
  "HY_HydroNexus": "hyf:HY_HydroNexus",
  "HY_HydroLocation": "hyf:HY_HydroLocation",
  "HY_HydrometricFeature": "hyf:HY_HydrometricFeature",
  "HY_WaterBody": "hyf:HY_WaterBody",
  "HY_Lake": "hyf:HY_Lake",
  "HY_Impoundment": "hyf:HY_Impoundment",
  "HY_River": "hyf:HY_River",
  "HY_Channel": "hyf:HY_Channel",
  "HY_FlowPath": "hyf:HY_FlowPath",
  "HY_DistanceDescription": "hyf:HY_DistanceDescription",
  "HY_HydroLocationType": "hyf:HY_HydroLocationType",
  "HY_IndirectPosition": "hyf:HY_IndirectPosition",
  "HY_DistanceFromReferent": "hyf:HY_DistanceFromReferent",
  "upstreamWaterBody": "hyf:upstreamWaterBody",
  "downstreamWaterBody": "hyf:downstreamWaterBody",
  "lowerCatchment": "hyf:lowerCatchment",
  "upperCatchment": "hyf:upperCatchment",
  "realizedCatchment": "hyf:realizedCatchment",
  "catchmentRealization": "hyf:catchmentRealization",
  "contributingCatchment": "hyf:contributingCatchment",
  "outflow": "hyf:outflow",
  "realizedNexus": "hyf:realizedNexus",
  "networkStation": "hyf:networkStation",
  "referencedPosition": "hyf:referencedPosition",
  "distanceExpression": "hyf:distanceExpression",
  "distanceDescription": "hyf:distanceDescription",
  "linearElement": "hyf:linearElement",
  "hydrometricNetwork": "hyf:hydrometricNetwork",
  "nexusRealization": "hyf:nexusRealization"
}}
```

Note that the JSON-LD document that describes the watershed using these two contexts does not implement all the relations in either context. Rather, they are limited to relations from them and only implement the relations from them that make sense or have content for the feature that is being described. Box 6: Watershed Data Index Use Case Entry JSON-LD Document [https://opengeospatial.github.io/ELFIE/usgs/huc/huc12obs/070900020601] shows this entry point JSON-LD document. Note that it uses elf.jsonld relations like schema:name, and schema:description, but not gsp:hasGeometry or schema:image. This is because a feature of type HY_Catchment is not expected to have a particular geometry and there is no image available for the feature. The flexibility of the JSON-LD context approach is a strength in that it can be applied to many cases easily, but this may be seen as a weakness if specific requirements and validations need to be implemented.

```
{"@context": [
    "https://opengeospatial.github.io/ELFIE/json-ld/elf.jsonld",
    "https://opengeospatial.github.io/ELFIE/json-ld/hyf.jsonld"
],
"@id": "https://opengeospatial.github.io/ELFIE/usgs/huc/huc12obs/070900020601",
"@type": "http://www.opengeospatial.org/standards/waterml2/hy_features/HY_Catchment",
"name": "Waunakee Marsh-Sixmile Creek",
"description": "USGS Watershed Boundary Dataset Twelve Digit Hydrologic Unit Code
Watershed",
"catchmentRealization": [
    {
        "@id":
"https://opengeospatial.github.io/ELFIE/usgs/nhdplusflowline/huc12obs/070900020601",
        "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_HydrographicNetwork"
    },
    {
        "@id":
"https://opengeospatial.github.io/ELFIE/usgs/hucboundary/huc12obs/070900020601",
        "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_CatchmentDivide"
    },
    {
        "@id":
"https://opengeospatial.github.io/ELFIE/usgs/hydrometricnetwork/huc12obs/070900020601"
,
        "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_HydrometricNetwork"
    }
]}
```

Chapter 10. Summary of Experiment and Analysis of Outcomes

As described previously, the ELFIE was intended to test existing OGC and W3C standards with the goal of establishing a best practice for expressing and exposing links between and among environmental domain and sampling features. The project sought to test existing solutions to linking data on the web with a "just enough" semantic annotation approach that would be approachable by developers and be adoptable using existing systems. Many use cases were discussed and a few of them were selected for experimental implementation.

All experimental implementations are summarized on the ELFIE demo pages. [https://opengeospatial.github.io/ELFIE/] The collection of experiments show that existing practices for encoding linked data can be used to satisfy the environmental linked features use cases considered by the ELFIE. In cases where features and related information were already published as linked data or had well established web services, these data and/or service-generated representations were easy to incorporate and use. Many of the core linked data relations used for search engine optimization were found to be suited to typical environmental linked feature use cases.

Note that in the JSON-LD files provided on the ELFIE web pages, URIs are relative to the file locations on opengeospatial.github.io for linked data demo purposes. This choice helped save time for the IE as it avoided each participant needing to set up the corresponding mechanics within its own data infrastructure. In production situations, each data provider will end-up exposing those files using their own URI policy.

10.1. Watershed Data Index Use Case Example

While additional content is available on the ELFIE demo pages [https://opengeospatial.github.io/ELFIE/], the watershed data index use case, that was described in greater detail in previous sections, is described in greater detail here. As shown in the JSON-LD example here, this use case has a primary entry point JSON-LD document for a single HY_Catchment feature that uses the elf.jsonld and hyf.jsonld contexts. As described in this section, the HY_Catchment is thought of as an unrealized feature that has multiple features that "realize" one of the ways we conceptualize catchments. The JSON-LD documents for these are shown in the source code box below.

Items to note below:

- Only two contexts are shown. While other contexts were used to some extent in ELFIE, this use case was kept purposefully simple and only demonstrates elf.jsonld and hyf.jsonld. See Box 4: elf.jsonld [https://opengeospatial.github.io/ELFIE/json-ld/elf.jsonld] and Box 5: hyf.jsonld [https://opengeospatial.github.io/ELFIE/json-ld/hyf.jsonld] for these context documents.

- Several geospatial preview geometry options are shown. As described elsewhere, use of the schema:geoshape, schema:geoCoordinates, gsp:hasGeometry, and/or a link to a geojson document are all possible and have potential value.

- HY_Features relations are draft URIs and will not resolve. Work to publish HY_Features feature types and associations is underway at the time of writing.

- Much more information could be included in the hydrometricNetwork document or at the

feature page of the individual hydrometricFeatures. Those details were not implemented for this demo.

- The catchment example shows three different representations of the abstract Catchment, each showing a different realization of the catchment. First, it is described as a HY_HydrographicNetwork, which is a network of waterbodies. Second, as a HY_CatchmentDivide, which is defined as the topographic divide around a catchment; and third as a HY_HydrometricNetwork, which lists all the monitoring stations that are found in this catchment.

Box 7: Catchment realizations of: [*https://opengeospatial.github.io/ELFIE/usgs/huc/huc12obs/070900020601*](https://opengeospatial.github.io/ELFIE/usgs/huc/huc12obs/070900020601) *geometry and networkStation lists truncated for clarity.*

```
{
  "@context": [ ①
    "https://opengeospatial.github.io/ELFIE/json-ld/elf.jsonld",
    "https://opengeospatial.github.io/ELFIE/json-ld/hyf.jsonld"
  ],
  "@id":
"https://opengeospatial.github.io/ELFIE/usgs/nhdplusflowline/huc12obs/070900020601",
②
  "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_HydrographicNetwork",
③
  "name": "NHDPlus V2.0 Network for Waunakee Marsh-Sixmile Creek",
  "description": "Collection of Flowlines for HUC12 watershed.",
  "realizedCatchment": [ ④
    {
      "@id": "https://opengeospatial.github.io/ELFIE/usgs/huc/huc12obs/070900020601",
      "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_Catchment"
    }
  ],
  "geo": [
    {
      "@type": "schema:GeoCoordinates",
      "schema:latitude": 43.2103,
      "schema:longitude": -89.5171
    },
    {
      "@type": "schema:GeoShape",
      "schema:url":
"https://opengeospatial.github.io/ELFIE/usgs/nhdplusflowline/huc12obs/070900020601.geo
json",
      "schema:polygon": "-89.5793623030186 43.1836688965559 ... -89.5793623030186
43.1836688965559"
    }
  ],
  "gsp:hasGeometry": {
    "@type": "gsp:Geometry",
    "gsp:asWKT": "MULTILINESTRING ((-89.57936 43.18367, -89.5793 43.18399, ···
```

```
      -89.46324 43.19541, -89.46321 43.19576))"
    }
}
⑤
{
  "@context": [
    "https://opengeospatial.github.io/ELFIE/json-ld/elf.jsonld",
    "https://opengeospatial.github.io/ELFIE/json-ld/hyf.jsonld"
  ],
  "@id":
"https://opengeospatial.github.io/ELFIE/usgs/hucboundary/huc12obs/070900020601",
  "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_CatchmentDivide",
  "name": "Waunakee Marsh-Sixmile Creek Boundary",
  "description": "USGS Watershed Boundary Dataset Twelve Digit Hydrologic Unit Code
Watershed Boundary",
  "realizedCatchment": [
    {
      "@id": "https://opengeospatial.github.io/ELFIE/usgs/huc/huc12obs/070900020601",
      "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_Catchment"
    }
  ],
  "geo": [
    {
      "@type": "schema:GeoCoordinates",
      "schema:latitude": 43.2114,
      "schema:longitude": -89.521
    },
    {
      "@type": "schema:GeoShape",
      "schema:url":
"https://opengeospatial.github.io/ELFIE/usgs/hucboundary/huc12obs/070900020601.geojson
",
      "schema:polygon": "-89.42488760601 43.216595418436 ⋯ -89.42488760601
43.216595418436"
    }
  ],
  "gsp:hasGeometry": {
    "@type": "gsp:Geometry",
    "gsp:asWKT": "MULTIPOLYGON (((-89.42489 43.2166, -89.42525 43.21684, ... -89.42532
43.21637, -89.42489 43.2166)))"
  }
}
⑥
{
  "@context": [
    "https://opengeospatial.github.io/ELFIE/json-ld/elf.jsonld",
    "https://opengeospatial.github.io/ELFIE/json-ld/hyf.jsonld"
  ],
  "@id":
```

```
"https://opengeospatial.github.io/ELFIE/usgs/hydrometricnetwork/huc12obs/070900020601"
,
  "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_HydrometricNetwork",
  "name": "Waunakee Marsh-Sixmile Creek Monitoring Network",
  "description": "Monitoring locations in the Waunakee Marsh-Sixmile Creek watershed.
",
  "realizedCatchment": [
    {
      "@id": "https://opengeospatial.github.io/ELFIE/usgs/huc/huc12obs/070900020601",
      "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_Catchment"
    }
  ],
  "networkStation": [
    {
      "@id": "https://opengeospatial.github.io/ELFIE/usgs/nwissite/huc12obs/USGS-
05427880",
      "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_HydrometricFeature"
    },
    {
      "@id": "https://opengeospatial.github.io/ELFIE/usgs/wqp/huc12obs/WIDNR_WQX-
10001227",
      "@type":
"http://www.opengeospatial.org/standards/waterml2/hy_features/HY_HydrometricFeature"
    }
  ]
}
```

① The two contexts used are here

② The identifier for this hydrograhic network. If you follow the link, you will see the full example including geospatial property details.

③ This is "HY_HydrographicNetwork". The link does not resolve yet because HY_Features have not been made available yet on the OGC server. A HY_HydrographicNetwork is a collection of waterbody features that drain to a catchment outlet.

④ The HY_HydrographicNetwork realizes a catchment describing it with as a network of lines representing waterbodies. Each realization of HY_Catchment uses both "schema:geo" and "gsp:hasGeometry" to provide both simplified representations that describe the location of the HY_Catchment as point or surrounding polygon and a more detailed representation respectively.

⑤ The second JSON document describes the catchment as a HY_CatchmentDivide. Notice that the geometry types for the "schema:GeoShape" and the "gsp:Geometry" are polygonal. However "schema:GeoShape" only allows an envelope while "gsp:Geometry" can handle multipolygons and other more complex geometries.

⑥ The third JSON document describes the catchment as a Hydrometric_Network. Here, the Network_Stations are of primary interest. The list of stations is truncated here to improve readability, in reality, there are many more stations located in this catchment.

10.2. Surface-ground water networks interaction Use Case Example

While additional content is available on the ELFIE demo pages, experience gained from the Surface-ground water networks interaction use case, that was described in greater detail in previous sections, is provided here.
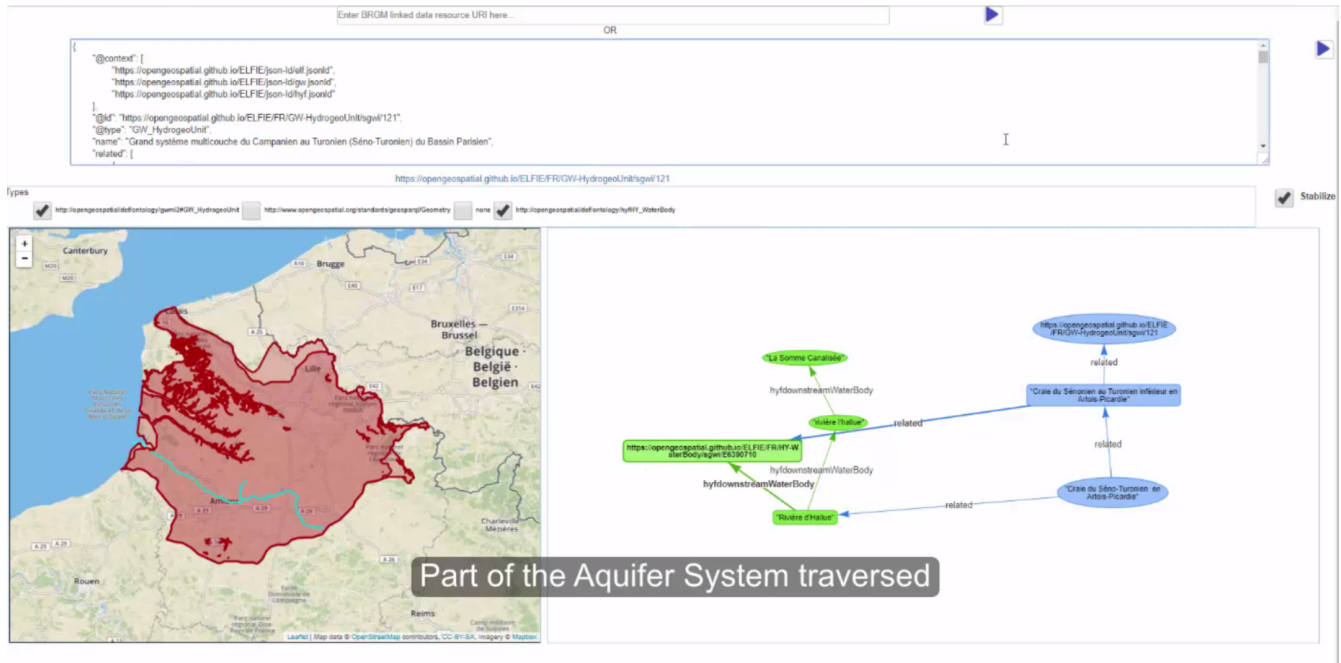
Figure 9. Screenshot of Surface-ground water networks interaction Demo.

Developing a generic client (here BLiV. [https://github.com/BRGM/BLiv]) in parallel to populating JSON-LD files helped in various ways:

- It triggered the geometry discussion mentioned several times in the ER

- It led the group to be more cautious and, in addition to validate JSON-LD payload, also validate in JSON-LD playground the contexts. The client's use of JSON-LD documents triggered various remarks on the contexts themselves.

- Using the @id, @type, name pattern when linking to another domain object proves useful to chose an interaction type (map, timeseries graph, etc…) in a GUI.

- Keeping the client as generic as possible, allows visualization of not only Surface-ground water networks interaction Use Case related JSON-LD files but also, those from the other Use Cases.

- Such an approach should prove valuable for demonstration of what linked-data is using domain-scientists own datasets as well as how it can help relate to other domains without having to modify source data.

10.3. Future Work

The experiments exposed several issues with existing and new technologies that need to be addressed to realize the full potential of the core linked data encoding technologies tested. The issue of representation of a preview geometry for a feature should be addressed and seems to be a tractable problem that could lead to significant benefits. Another tractable and important issue is

creation of ontologies for domain features. As of the completion of the ELFIE, this work was already being undertaken for existing domain models where a UML to OWL conversion is possible. This was the topic of an ad hoc meeting during the OGC TC meeting in March 2018 in Orléans, France. The main outcomes are summarized in Annex A: Strategies for Publishing Domain Ontologies as Linked Data from OGC domain standards. Other domain models, such as flood impact features, need to be created and or vetted by a community.

There were also several issues highlighted by the ELFIE that were deemed out of scope and purposefully left for future efforts. The foundational issue in this category is the "landing page" or "default representation" problem. That is, how should we handle the case where a dereferenceable URI is meant to identify a real-world entity with multiple digital representations. There are many related issues related to the network architecture and expected behavior when dereferencing URIs. These were generally out of scope for the ELFIE but need to be addressed to achieve interoperability and include: URI structure, use of non-information URIs with WFS and other services, and strategies for managing collections of potentially temporary linked data among many data providers. Use of the domain feature ontologies declared in the JSON-LD contexts was also out of scope. That is, JSON-LD files produced by ELFIE were not ingested by reasoning software. The extra vocabulary source was not yet taken into account by major search engines crawlers that focus mainly on the schema.org ontology.

Chapter 11. Issues and Recommendations for Future Work

The results of ELFIE are quite promising in that all major issues encountered are well known and subject of ongoing work. Technologies tested, where mature and well used, worked well for the purpose and it seems that the future for an easily adoptable resource-oriented linked web of data resources is bright. The following sections outline issues encountered and describe recommended prospective path forward to overcome them.

11.1. OGC Service Standards and Implementation issues

OGC service interfaces were not the primary focus of ELFIE. However, along the way, both WFS and SOS services were discussed and tested. The major issues encountered were around querying these services by feature identifier. There is something of a chicken and egg situation in that, in most cases, http feature identifiers have not been published so are not available to use to query a web service, but the OGC services also do not lend themselves to publishing feature identifiers.

If queries to WFS or SOS services are used as un-structured links that return representations of features or results of observations, they can be used readily in linked data systems. But to be used with embedded feature identifiers or with more sophisticated content-negotiated retrievals, changes to the web service APIs will be needed.

The work ongoing with WFS appears promising with regard to the issues outlined above. The ability to query by a specific id that can be an http url is supported and is a clear solution that would solve the problem illustrated here. e.g. something like: http://server.com/wfs/getfeature?id=http://feature.x.y.x

A similar discussion has taken place in sensor-related services. SOS V2.0 specification (OGC 12-006), discusses identifier handling in Annex B and promotes reuse of a global identifier (here gml:identifier) as opposed to local identifier (gml:id) when querying services. The SensorThings API Part 1 implementations use their own internal identifier for REST API operations (ex : http://sensorthings.brgm-rec.fr/SensorThingsGroundWater/v1.0/Observations(1752377)). Both services' endpoints (SOS, SensorThingsAPI) were tested, passing them URIs for featuresOfInterest, observed properties, process, observation identifier, etc. to use them in a linked data context. Resulting queries work but further implementation is required to improve usability of the approach.

A JSON-LD response from these services is not available. While response rewriters (to convert existing encodings to a linked data form) have been tested on top of the SensorThings API (JSON to JSON-LD). It may be useful to include JSON-LD capabilities in the standards themselves.

Apart from the semantic association and serialization of features and observations, assignment of URIs to observational resources remains a system by system and use-case by use-case problem that is not necessarily trivial. Identification of unique observations and maintaining identifiers for them can become very complicated considering that time series subsetting is a complex subject in and of itself. The Research Data Alliance 'Data Citation WG Recommendations' [https://rd-alliance.org/group/

data-citation-wg/outcomes/data-citation-recommendation.html]   [<https://doi.org/10.15497/rda00016,8>]
might help guide the community to a useful solution on this front.

11.2. Domain Feature Models

Domain feature models, such as GWML2 and SoilML [http://www.opengeospatial.org/projects/initiatives/soildataie], are defined for encoding and exchange of representations of environmental features. Other models, like HY_Features, are only available in conceptual (UML) form. For use in linked data, best practices and publication workflows for domain feature models as linked data ontologies need to be established. As described in Box 8: HY_Features Ontology Note and Strategies for Publishing Domain Ontologies as Linked Data from OGC domain standards this work has begun and needs to be brought to conclusion for use across the community.

W3C and OGC work on Observations and Measurements and Sensor Ontologies (See https://www.w3.org/TR/vocab-ssn/ [https://www.w3.org/TR/vocab-ssn/] [9]) is among the most advanced with regard to support for linked data approaches as pursued by ELFIE. While attempting to use the Semantic Sensor Network ontology, the concept of a monitoring network was found to be missing or not handled directly. Ad-hoc collections of monitoring features are a common need. Either as a collection that is meant to characterize a specific regional environmental feature, such as a watershed, or as a collection of monitoring features with common ownership, data offerings, or quality. In any case, further analysis of the nature of collections of monitoring features for use within and among environmental monitoring domains is warranted.

11.3. Publication of URIs, Content, Types, and Associations

A major issue (which was known and was out of scope for ELFIE) was the lack of published feature identifiers and a linked-data baseline to build on. It is clear that, to grow the linked-data baseline, features and data linked to them need to be published using existing basic practices as a starting point. This starting point could be as trivial as a URL that returns a JSON-LD document with ID, Type, Name, and representative point location. The addition of a URI broker, content negotiation to provide HTML and RDF representations, and additional linked content would be possible given current techniques. However, it should not be seen as a necessary prerequisite to publishing linked data. Without a rudimentary starting point, the community has nothing to build on and could continue to debate the nuances of eventual solutions for ever.

Similar to publication of features and other linked data, URLs that define feature types and relations are not available in many cases. Without these referenceable, resolvable definitions, the linked data practices described here are very difficult to use across multiple systems that need to use common types and associations. Further, even when URIs have been established, the network behavior and content received when dereferencing them is not consistent across the community. For example, https://schema.org/GeoCoordinates returns an html page by default and implements both HTTP "accept" headers (e.g. "Accept: application/ld+json") and URL suffixes (e.g. https://schema.org/GeoCoordinates.jsonld) to request specific content types and http://www.w3.org/2004/02/skos/core#related defaults to an html page and implements accept headers but does not support the same content types as schema.org and does not support URL suffixes. Other linked data URIs, those from GeoSPARQL for example (http://www.opengeospatial.org/standards/geosparql/

asWKT), do not resolve anything. If the community is going to start building cross-organization solutions based on linked data, these baseline feature types and associations need to be available and well described in a common way. It should be no surprise that schema.org's linked data is used very broadly given that they provide a rich and approachable suite of content for their types and associations.

Box 8: HY_Features Ontology Note

Since the ELFIE project has completed the OGC Naming Authority has embarked upon provision of a consistent, extensive and extensible Linked Data compatible "Definitions Server" to address some of these concerns. Publication of HY_Features has been prioritized and is available at **https://www.opengis.net/def/appschema/hy_features/hyf** Further experimental work is required to identify the range of different representations that can be used, such as automatic provision of JSON-LD context resources.

11.4. Default Response When a URI has Multiple Representations

Non-information resource identifiers, URIs that are an identifier for a thing that may or may not have information representations, are useful to provide persistent and reusable identifiers for entities such as real-world features. These non-information resource identifiers provide a means for global identification across a distributed system such that multiple members of the system can contain information about and/or in reference to shared entities. There is a basic conundrum with non-information resource identifiers in that there can only be one default response (for a given encoding) when dereferencing a URI and only one member of a distributed system can be the source of that default response. Further, in order to achieve linked system interoperability, dereferencing behavior of non-information resource URIs needs to be in accord with the expectations of a member node of a system. The httpRange-14 [https://www.w3.org/2001/tag/group/track/issues/14] and the Cool URIs Technical Report [https://www.w3.org/TR/cooluris/] [10] provide a significant basis for solving this conundrum and W3C working groups continue to focus on it, but many details remain to be agreed upon for use cases such as are being pursued by the ELFIE.

While details of the default dereferencing behavior of non-information resource identifiers and subsequent redirects to information resources was discussed in the ELFIE, the IE has attempted to focus on encoding of linked data information resources rather than the network behavior for discovery and retrieval of non-information resources. Future work and best practice development will be needed to address these issues:

- HTTP vs HTTPS

- embedded hints and headers to canonical identifiers vs information resources

- permanence and bookmarkability of information resource URLs

- detecting information resource URLs and injecting canonical URIs into systems, so that other resources can be found

- declaration of sameAs relationships

- where sameAs relationships are mandatory and optional - where can an agent find these?

- stricter definition of version vs schema vs content model vs profile

- use of expected W3C formalisms for describing profiles

- optimal registration processes and registry provision for identifying, cataloguing, discovering and serving profiles/contexts

- use of expected W3C/IETF mechanisms for profile negotiation in URI dereferencing

- Adding additional dimensions of representation choice when dereferencing for geometry type, resolution, CRS etc

11.4.1. Preview Geometry

As demonstrated in the Watershed Data Index use case JSON-LD document, how to encode a preview geometry is not clear. The only widely used preview is a point geometry defined by schema.org GeoCoordinates. The schema.org polygon and line features have not been implemented widely and, like GeoSPARQL WKT geometries, require interpretation above and beyond what would be required for GeoJSON. Linking to a GeoJSON file works, but is not common in practice and requires additional web-requests to retrieve content, which is undesirable for a basic preview geometry. With the advent of javascript libraries that can handle WKT, it seems likely that a GeoSPARQL WKT geometry could be used effectively, but further experimentation may be needed to confirm this assumption.

Appendix A: Strategies for Publishing Domain Ontologies as Linked Data from OGC domain standards

Table 2. Authors

Name	Affiliation
Sylvain Grellet	BRGM
Abdelfettah Feliachi	ATOS
Rob Atkinson	OGC
Eric Boisvert	GSC
Marcus Sen	BGS
Mickaël Beaufils	BRGM
Alistair Ritchie	Manaaki Whenua – Landcare Research
James Passmore	BGS
Boyan Brodaric	GSC
Katharina Schleidt	Data Cove
Chuck Heazel	WiSC Enterprises
Steve Richard	AZGS

A.1. Introduction

OGC Environmental Linked Features Interoperability Experiment (ELFIE) has explored publishing environmental observations as linked data, with data interpretation supported by having unambiguous references to key domain concepts in the data represented by additional links (URIs). Such concepts cover the nature of the properties of a data object, but also the key Use Case of linking an observation to the Feature of Interest.

Parallel to this activity, members from several OGC SWGs (GeoSciML, Groundwater, Hydrologic Features) are attempting to 'port' the initial work done using ISO 191xx series UML paradigm to OWL.

Taking the opportunity of having all the involved parties available during the 2018 Orléans TC an ad-hoc meeting was organized on March 23rd, 2018.

This document builds on those joint discussions in order to share best practices. It is shared under the GeoSciML SWG GitHub repository (https://github.com/opengeospatial/GeoSciML/tree/master/documents).

A.2. Problem statement

Common semantics for the description of key domain concepts and the type of observations' the

Feature of Interest are identified by using published models – such as HY_Features, GWML, GeoSciML. These models (currently) are available in normative forms as UML notation, and in some cases as XML schema, neither of which is conducive to easily dereference a URI for a concept to get more detail about that specific concept, or even to confidently match the identity of any references to concepts.

The requirement therefore is to be able to publish such models, as stable URIs, dereferencing as fine-grained Linked Data. This means that a predictable and human readable URI naming scheme for elements in a model is required. (this may not be strictly necessary for a single model, but when a body of such models are managed by a single authority – such as OGC and its delegated Standards Working Groups – then commonality is necessary to avoid revisiting the naming strategy for each case, and to allow users to become familiar and comfortable with a consistent product.) Within the ELFIE context the need has been identified for at least two models, HY_Features and GWML2, and their multiple component Application Schemas. The GeoSciML community is also looking at the same issues. There is a need for both efficiency and consistency to harmonize and document a common approach.

ISO 19150-2 defines a set of rules for encoding UML as an OWL ontology referencing the suite of OWL artifacts created to model the ISO Harmonized Model. This form of OWL has an unknown utility – by tying data types and concepts into the extensive ISO model this adds a high burden on clients to reason over the entire model (quite a large set of content) in order to identify fairly simple, but highly important, semantic baseline information – such as that a property may be treated as an xsd:integer for purposes of a calculation). The feasibility of such reasoning is unknown, as there is no way yet to test the ISO models behave as expected under reasoning conditions. Furthermore, for the purposes of Linked Data, much of the modeling of behavior of low-level data types is not relevant, when the main requirement is to identify concepts, access explanations, or potentially access information about implementations of these concepts that may be available. Finally, from the perspective of infrastructure support, OGC provides a "definitions server" that is designed around the SKOS meta-model and will provide dereferencing services and Linked Data representations. The OGC Definitions Server is controlled by the OGC Naming Authority Subcommittee (OGC-NA).

A.3. Background

BRGM and GSC have performed an initial analysis of the issues and questions that arise in the context of:

- Generating ontologies for a subset of GeoSciML: GeoSciML Basic, Borehole and Lite,

- Testing populating instances for Boreholes, Geologic Units pointing to vocabularies when available (FR, EU, CGI GTWG) and testing some representations (maps, graph)."

The output of that exercise has been shared under the BRGM repository (https://github.com/BRGM/GeoSciMLontology/blob/master/documents/ISO191xx_2_OWL_NoteBRGM.docx)

Identified issues and questions are relevant to the equivalent challenges for the ELFIE, and collaboration with the GeoSciML, GWML, HY_Features and the ELFIE activities has been initiated to develop a common solution. Those elements have then been further characterized from the perspective of definition publication process and governance.

Building on this during the 2018 Orléans TC ad-hoc meeting on March 23rd, identified issues were discussed.

A.4. Proposed strategy

Resources, time and testing methodology for "hand building" optimal equivalent ontologies are not available (notwithstanding comments that these may lead to iterative refinement and improvement of the model). The only automated pathway available to all participants currently is using ShapeChange (http://shapechange.net), which means developing a set of encoding rules from a wide range of choices and options from previous experiments in encoding different styles of UML model. The ISO 19150-2 artefacts will be seen as an intermediate artefact available by an annotation property reference in the metadata describing the provenance of the ontology. For such perspective the use of PROV-O should be used. "Manual interventions" to artefact to correct bugs or apply further rules will be encoded as RDF transformations or additional statements that can be applied in sequence.

The process can be summarized like this:

1. ShapeChange starting point: conceptual/logical model

2. manual adjustments according to what is written in this document

If necessary, rules to extract basic OWL and xsd: datatype equivalence, without embedded reference to ISO datatypes will be developed and applied as part of OGC-NA infrastructure. Rules to extract SKOS equivalent glossary/taxonomy from OWL class models will be developed and applied as part of OGC-NA infrastructure. URIs under www.opengis.net/def will be assigned, and content will be marked as draft, as an exercise in publication governance by OGC-NA, to make content and process available for wider discussion.

A.5. Issues identified

A.5.1. Naming Policy

OGC published ontologies will have URIs and Linked Data access based on identifiers under a pattern yet to be chosen from the following options.

Option 1

> @prefix s1: http(s)://www.opengis.net/def/{ontology}/<authority>/<schemaAcronym> ex: http(s)://www.opengis.net/def/ontology/HydroDWG/hyf or http(s)://www.opengis.net/def/ontology/HydrologicFeaturesSWG/hyf

Option 2

> http(s)://www.opengis.net/def/<authority>/{ontology}/<schemaAcronym> ex: http(s)://www.opengis.net/def/HydroDWG/ontology/hyf or http(s)://www.opengis.net/def/HydrologicFeaturesSWG/ontology/hyf

Option 3

> http(s)://www.opengis.net/def/{ontology}/<schemaAcronym> ex:

http(s)://www.opengis.net/def/ontology/hyf

http(s)://www.opengis.net/def/ontology/hyf/HY_Waterbody → will be the identifier of the class in the ontology

Option 4

http(s)://www.opengis.net/def/<schemaAcronym>   ex: http(s)://www.opengis.net/def/hyf (http(s)://www.opengis.net/def/hyf/HY_Waterbody)   http(s)://www.opengis.net/def/gwml2 http(s)://www.opengis.net/def/gsml Under that option, the reserved word {ontology} being removed, the client has to specify which representation is desired. Thus owl model, rdf, xsd, json-ld context will be returned based on content negotiation (Accept:header) or an explicit file extension (ex: https://www.opengis.net/def/gwml2.xsd , https://www.opengis.net/def/gwml2.ttl).

Note:

- Words between curly brackets (ex:{ontology}) are 'reserved words' thus will remain as is when applied in URIs

- Words between angle brackets will be replaced by the corresponding values when applied in URIs (ex : 'hyf', 'gwml2', 'gsml' for ex:<schemaAcronym>)

- as http://www.opengis.net/def/ and http://www.opengis.net/def/auth/ resolve to a wide variety of different notions

- writing convention

  ◦ Class names will be UpperCamelCase names e.g. **s1:Class1**

  ◦ Properties will be lowerCamelCase e.g. **s1:prop1**. Except for class scoped properties which names are ambiguous (ex: 2 classes having homonym properties but with different semantics) where the applied formalism will be **s1:Class1.prop1**

  ◦ For more details: see options below

- General semantic web BP

  ◦ base/document/ for identifying informative resources

  ◦ base/id/ for identifying real world entities

  ◦ base/def/ for identifying ontologies and their components

A.5.2. Weaknesses or issues with ISO 19150-2 rules

- The rules of ISO 19150-2 restrict the resulting ontologies to the way that the UML metamodel works. Respecting all of the 19150-2 means we do not take into account the Open World Assumption when working with ontologies (missing a piece of information doesn't mean that piece of information is false). For instance, placeholder properties or classes in UML are transformed to OWL properties and classes where there is no need for them.

- The transformation rules are consistent but limits the resulting ontologies to the UML paradigm. Some additional work may be done on the resulting ontologies to add semantics between classes (disjunctions, subsumption, equivalence, etc) and within or between properties (functional properties, transitive properties, symmetric properties, inverse of, etc).

- No specific indications about association classes are mentioned in the norm. It is obvious that an association class is translated as an OWL class. No rule for linking this class to the related

class(es) appear.

- Union: ISO 19150-2 recommends to use owl:UnionOf, the implementation in ShapeChange seems rather to stick to ISO 19118 approach (disjoint union) but does it in a very complex way, as explained in the OGC Testbed-12 [http://docs.opengeospatial.org/per/16-020.html#clause_solutions_rdf_unions]. Instead, this ER suggests that this solution can be simplified using the OWL2 property owl:disjointUnionOf. This should generate a ChangeRequest to ISO 19150-2.

A.5.3. Property names and definitions

- Properties naming when translating attributes: dots in properties identifiers could be interpreted somehow that they are still scoped to classes, while in ontologies, properties are scoped to a namespace instead. Properties are independent entities that may or may not have a specific class as a domain. This is one major structural difference between UML and OWL.

  - Use general (non-scoped to class) property names when the name of the attribute or association is unique. Thus, leave the domain of the properties open (or typed as owl:Thing). The restrictions on the properties values in the class definition can be used for this purpose.

  - When there is an ambiguity, allow scoped names for properties (class.Property) then verify whether

    - automatically created properties can be merged into one (eg. GeologicFeature.purpose and EarthMaterial.purpose).

    - or if automatically created properties can be subPropertyOf a higher one. It was brought to our attention after the ad-hoc meeting that the Application Schema-based Ontology Development Engineering Report (OGC Testbed-14) [https://portal.opengeospatial.org/files/?artifact_id=81893&version=1] provides in its chapter 7 an analysis for "OWL Property Generalization" that should be implemented in ShapeChange 2.7.0.

- Domains and ranges of properties

  - Domains and ranges properties should not be defined in the reference ontology to favor reuse. They could be specified in application ontologies that reuse the properties (if needed). Instead, restriction on the values of the properties should be defined for every class.

A.5.4. Alignment documents (UML → OWL)

- These are the place to put subPropertyOf relationships (roleA and roleB are flavours of role) - also equivalences across application schema

- Skos:notation (datatype to be determined) to preserve original property name token - for display and reference to xpath elements

- Both are not automated yet in ShapeChange

- There is no direct Sensor Web Enablement (SWE) ontology but several concepts from SWE can be found elsewhere (e.g. https://www.w3.org/TR/vocab-ssn/)

- Reference to basic SWE types must be modified if needed by specialized Classes from other ontologies or by defining new ones.

- Use GSML_QuantityRange instead of swe:QuantityRange as recommended in GeoSciML definition.

- Rename swe:Category to skos:Concept or mdl:Lineage (depending on the case) and swe:Quantity to the relevant class in the context (ts:TimePosition, mdq:PositionalAccuracy, etc.).

- Preparing for application ontologies: To enable GeoSciML Basic and Borehole properties to be reused in application ontologies like GeoSciML Lite, we activate the ShapeChange rule "rule-owl-prop-globalScopeByUniquePropertyName" that scopes unique name property to global use, and thus not specify the domain of these properties. The scoping of the properties to their classes in Basic and Borehole is done using restrictions on the values that these properties can take for their corresponding classes. This can be done thanks to the ShapeChange rule "rule-owl-prop-range-local-withUniversalQuantification". In the considered standards as in many OGC domain models, UML constraints are expressed in non-canonical forms in the UML classes definitions or in OCL. Nevertheless, it was mentioned attention after the ad-hoc meeting that the Application Schema-based Ontology Development Engineering Report (OGC Testbed-14) [https://portal.opengeospatial.org/files/?artifact_id=81893&version=1] provides in its chapter 5 an analysis for "Conversion of OCL Constraints" that could be useful for future implementations.

- The requirements of the model cannot be all respected in the ontology representation (eg. "QuantityRange properties that must report a single value SHALL assign both lower and upper value as equal to that single value."). This should be checked and translated manually as restrictions (owl:Restriction, other classes axioms, properties relations, ...) when possible afterwards. Implementation choices for specific communities

- ShapeChange "Map entries" provide a flexible way to choose recommended names for properties and classes. This would enable one to reuse existing specialized classes and properties from external ontologies.

- GeologicUnitView contains mixed information from both GeologicUnit and MappedFeature. A decision must be made to which entity the view must be associated (using the same URI as the GeologicUnit or MappedFeature)

A.5.5. Meta-model issues (expressivity mismatches between OWL and UML)

- The placeholder attribute "any" (in GeoSciML Lite) becomes useless property in owl delete it. Choice made to replace the "character string" data properties by object properties from GeoSciML Basic, borehole and other ontologies when possible (using the XPath mapping detailed in GeoSciML specification).

- «typ» and «FeatureType» serialise to owl:Class - we need to have further annotation or axiomitisation (e.g. «datatype»)

- Abstract class: According to ISO 19150-2, abstract classes in UML are transformed to annotated owl class. But in GeoSciML, some abstract classes were created to provide an extension point for GeoSciML extension (ex: FoliationAbstractDescription); they provided a bag to list properties. Some might then be revisited/deleted (the only reason to keep them would be for schema mapping purposes but we considered it a low priority use case compared to LinkedOpenData, Websem reasoning)

- The expressiveness of ontology languages should be used to enrich the reasoning: axioms on classes (equivalence, disjointness), and properties relations (inverse, equivalence) and characteristics (transitivity, symmetry, functionality and inverse functionality). In this scope, ShapeChange provides for example a general rule [https://shapechange.net/targets/ontology/uml-rdfowl-based-isois-19150-2/#rule-owl-cls-disjoint-classes] for defining disjointness of the direct subClasses of a

Class. In addition, it was brought to our attention after the ad-hoc meeting that the Application Schema-based Ontology Development Engineering Report (OGC Testbed-14) [https://portal.opengeospatial.org/files/?artifact_id=81893&version=1] provides in its chapter 6 an analysis for "OWL Property Enrichment" that should be implemented in ShapeChange 2.7.0.

- UML class union should be transformed using owl:disjointUnionOf

- The key meta-model issue is the use of a character string (UML option) to hold an Internationalized Resource Identifier (IRI) in a particular implementation profile - and the trickiness of modeling this as an objectProperty or not. Option could be to model it as an rdfs:Property, and allow implementation profiles to constrain it to an owl:ObjectProperty.

A.5.6. Bugs and limitations in software (or things too hard to configure)

- Association classes must be handled differently: ShapeChange transforms an association class into separate class and properties. Thus, no link is created between the association class and the classes that are initially related by it in the UML. No direct rule is found in ShapeChange to handle that. As it was brought to our attention after the ad-hoc meeting, a workaround solution is to use ShapeChange Transformer [https://shapechange.net/transformations/association-class-mapper/] in order to transform association classes into a semantically equivalent structure as explained in the OGC Testbed-12 Engineering Report [http://docs.opengeospatial.org/per/16-020.html# rdf_cr_associationclass]. This solution wasn't tested during the experiment.

This must be defined afterwards with two properties: associationSource and associationTarget (exactly as in passing from conceptual model to a logical schema). As a solution, this could be locally defined as "Source" and [association name]"Target". These two properties must have the right domain and range. The direct property between the source and the target automatically created by ShapeChange must be deleted.

A.5.7. Annotation practices

- Version the ontology: use owl:priorVersion and owl:versionInfo properties to describe the ontology, and owl:deprecatedClass and owl:deprecatedProperty annotation properties to specify the version status of a class or a property when deprecated.

- Use PROV-O to describe the provenance of the ontology with reference to 19150-2, ShapeChange configuration, …

A.5.8. Proposed behavior when external classes is specified as properties values

When a UML class from another schema is referenced (Observation class for example), it should be replaced by the specialized classes from the ontology of that schema (could be automated in ShapeChange). If such ontology is not defined (SWE types for example) use (equivalent) classes from other ontologies or define new ones.

A.5.9. Standing issues

- Usage of SKOS VS dedicated classes when transforming «codeList» from the UML: The pattern proposed by ISO-19150-2 is to create a class for each property designed to hold a "term". This

class shall be a subtype of skos:Concept according to the spec. This is seen as a problem for some as SKOS is not the only possible way to encode vocabularies, as some might prefer to encode vocabularies as formal ontologies. Both solutions for implementing codeList (as skos:Concepts or as a dedicated class) can be done using ShapeChange as explained in Testbed-12 (ShapeChange Engineering Report) [http://docs.opengeospatial.org/per/16-020.html#_code_lists]. However, encoding vocabularies as formal ontologies (ex. owl:Class hierarchy) requires a different tool or must be handled manually.

- Version URI: Do we need to specify where version numbers go in the URI schemes discussed above ?

A.6. Support material

A.6.1. Configuration references

Shapechange configuration: https://github.com/opengeospatial/GeoSciML/blob/master/tools/shapeChange/gsml4_bh.xml Example of transforming GeoSciML Borehole UML Model into OWL. Should be re-used for other models (just need to change the source EAP file, appSchemaName, URIbase).

A.6.2. GeoSciML encoding example

Example of transforming GeoSciML Borehole UML Model into OWL Resulating raw ontology from Shapechange: https://github.com/opengeospatial/GeoSciML/blob/master/ontology/1_raw_from_Shapechange/gsmlbh.ttl Ontology after manual edition: https://github.com/opengeospatial/GeoSciML/blob/master/ontology/2_after_manual_edit/gsmlbh.ttl

Appendix B: Default Response for Multiple Representions and Relations

Non-information resource identifiers, URIs that are an identifier for a thing that may or may not have information representations, are useful to provide persistent and reusable identifiers for entities such as real-world features. These non-information resource identifiers provide a means for global identification across a system such that multiple members of the system can contain information about and/or in reference to shared entities. There is a basic conundrum with non-information resource identifiers in that there can only be one default response when derefencing a URI and only one member of a distributed system can be the source of that default response. Further, in order to achieve linked system interoperability, dereferencing behavior of non-information resource URIs needs to be in accord with the expectations of a member node of a system. The httpRange-14 [https://en.wikipedia.org/wiki/HTTPRange-14] and the Cool URIs Technical Report [https://www.w3.org/TR/cooluris/] provide a significant basis for solving this conundrum and W3C working groups continue to focus on it [https://www.w3.org/2017/dxwg], but many details remain to be agreed upon for use cases such as are being pursued by the ELFIE.

While details of the default dereferencing behavior of non-information resource identifiers and subsequent redirects to information resources was discussed in the ELFIE, the IE has attempted to focus on encoding of linked data information resources rather than the network behavior for discovery and retrieval of non-information resources. Future work and best practice development will be needed to address these issues:

1. HTTP vs HTTPS

2. embedded hints and headers to canonical identifiers vs information resources

3. permanence and bookmarkability of information resource URLs

4. detecting information resource URLs and injecting canonical URIs into systems, so that other resources can be found

5. declaration of sameAs relationships

6. where sameAs relationships are mandatory and optional - where can an agent find these?

7. stricter definition of version vs schema vs content model vs profile

Appendix C: Use Cases Considered in Design Process

River network linked data to address hydrologic drought and ecology

Water consumers and decision makers are interested in potential ecological and economic impacts of additional water use.

Relevant datasets

- River network
- Incremental catchments
- Low flow and temperature sensitive species distributions
- Monitoring locations
- Low flow and temperature forecasts
- Ecological stream classification
- Historical streamflow characteristics
- Flow and temperature regulations

Stream water quality and quantity monitoring

Water consumers are interested in ambient stream characteristics and water quality impacts of water allocation decisions.

Relevant datasets

- River network
- Monitoring locations
- Water quality data
- Water quantity data
- Water allocation information
- Diversions and discharges

Flood risks and impacts

A person in a flood-impacted area needs information about the risks to them for a forecast or ongoing flood event.

Relevant datasets

- River network
- Stream flow and stage forecasts
- Topography
- Flood depth and extent data for event
- Flood plain map(s)
- Flood control assets

- Impacted transportation assets

- At risk critical infrastructure

Water quality impacts monitoring of mining activities

A water quality analyst needs to know what data is available to assess baseline and impacted conditions of a waterbody impacted by mine-related runoff.

Relevant datasets

- River network

- On network waterbodies

- Realtime water quality monitoring locations and data

- Water quality sampling locations and data

- Mine locations

- Snowpack and runoff

- River flow monitoring locations and data

Cross border monitoring index

An analyst working in a cross border watershed or aquifer needs information characterizing their study region from another country of jurisdiction.

Relevant datasets

- River network data

- Catchment data

- Aquifer data

- Streamflow monitoring locations and data

- Aquifer monitoring wells and data

Water budget summary

A person interested in a basic summary of the water budget for a given watershed needs information about the watershed and its water budget data.

Relevant datasets

- Watershed boundary

- Watershed outlet location

- Water budget component data

- Links to additional information

Watershed data index

A person with a stake in the conditions of a given watershed needs an index of all water-quality related information about the watershed.

Relevant datasets

- Impaired (e.g. polluted or degraded) waters

- Drinking water quality violation reports

- Facilities with discharge permits

- Water quality infrastructure investments

- Watershed characteristics

- Modeled data such as nutrient concentrations

Flood monitoring at bridges

A jurisdiction with significant infrastructure near or over waterways needs a system to link waterways, bridges, and various observations for situational awareness and model calibration/validation.

Relevant datasets

- River network

- Bridge inventory

- Monitoring stations and data

- Locations where model results are available

Flood Impact Study

A hydrologist conducting a flood impact study needs to collect relevant information from various agencies.

Relevant datasets

- River network

- Watershed boundaries

- Hydrologic locations of vulnerable infrastructure

- Hydrologic locations of monitoring stations and data

- Meteorological monitoring stations and data

- Elevation data

- Water table / aquifer data

- Rainfall forecast

Ground water level monitoring

Groundwater level monitoring networks are deployed on the field to monitor Aquifer status. A groundwater specialist needs to be able to traverse links from the Well to the piezometer deployed and, furthermore to the acquired observation (timeseries).

Relevant datasets

- Groundwater wells description

- Piezometer description

- Groundwater levels at piezometer level

Surface-ground water networks interaction

Management of the pressures on natural resources requires properly linking domain features not only using surface/ground water interaction at a local level but also taking benefit of the broader river networks and aquifer systems connectivity.

Relevant datasets

- Borehole/Well

- Piezometer (+ associated observations)

- Aquifer

- AquiferSystem

- Aquifer/River interaction

- River Network

- River

- gage (+ associated observations)

Surface water impacted groundwater level forecasting

A groundwater extraction modeler needs information about surface water that replenishes groundwater so they can understand and forecast groundwater availability.

Relevant datasets

- Groundwater wells and level data

- Surface water monitoring sites and flow data

- Meteorological monitoring sites and observations

- Output predictions at groundwater prediction wells

Appendix D: Revision History

Table 3. Revision History

Date	Editor	Release	Primary clauses modified	Descriptions
July 28, 2018	D. Blodgett	0.1	all	initial version
August 28, 2018	B. Cochrane	0.2	all	reorganize and edit
September 19, 2018	R. Atkinson	0.3	all	Updated status of HY_Features ontology and W3C directions
November 1, 2018	S. Grellet, A. Feliachi	0.4	all	Use case additions and technical edits
November 12, 2018	D. Blodgett	0.5	all	Final formatting and summary sections

Appendix E: Bibliography

1. B. Hyland, G. A. Atemezing, M. Pendleton, and B. Srivastava, "Linked Data Glossary," W3C, W3C Note, Jun. 2013.

2. R. Cyganiak, M. Lanthaler, and D. Wood, "RDF 1.1 Concepts and Abstract Syntax," W3C, W3C Recommendation, Feb. 2014.

3. P. Barnaghi, J. Tandy, and L. van den Brink, "Spatial Data on the Web Best Practices," W3C, Sep. 2017.

4. G. Kellogg, M. Sporny, and M. Lanthaler, "JSON-LD 1.0," W3C, W3C Recommendation, Jan. 2014.

5. D. Kontokostas and H. Knublauch, "Shapes Constraint Language (SHACL)," W3C, W3C Recommendation, Jul. 2017.

6. G. Kellogg, "JSON-LD 1.1," W3C, W3C Working Draft, Oct. 2018.

7. A. Miles and S. Bechhofer, "SKOS Simple Knowledge Organization System Reference," W3C, W3C Recommendation, Aug. 2009.

8. [[https://doi.org/10.15497/rda00016]]A. Rauber, A. Asmi, D. V. Uytvanck, and S. Proell, "Data Citation of Evolving Data: Recommendations of the RDA Working Group on Data Citation (WGDC)," 2016.

9. A. Haller, S. Cox, M. Lefrançois, D. L. Phuoc, K. Taylor, and K. Janowicz, "Semantic Sensor Network Ontology," W3C, W3C Recommendation, Oct. 2017.

10. L. Sauermann and R. Cyganiak, "Cool URIs for the Semantic Web," W3C, W3C Note, Dec. 2008.