

Second Environmental Linked Features Experiment:

Publication Date: 2020-10-22

Approval Date: 2020-09-23

Submission Date: 2020-08-24

Reference number of this document: OGC 20-067

Reference URL for this document: <http://www.opengis.net/doc/PER/SELFIE-ER>

Category: OGC Public Engineering Report

Editor: David Blodgett

Title: Second Environmental Linked Features Experiment:

OGC Public Engineering Report

COPYRIGHT

Copyright © 2020 Open Geospatial Consortium. To obtain additional rights of use, visit <http://www.opengeospatial.org/>

WARNING

This document is not an OGC Standard. This document is an OGC Public Engineering Report created as a deliverable in an OGC Interoperability Initiative and is not an official position of the OGC membership. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an OGC Standard. Further, any OGC Public Engineering Report should not be referenced as required or mandatory technology in procurements. However, the discussions in this document could very well lead to the definition of an OGC Standard.

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.

Table of Contents

| | |
|--|----|
| 1. Subject | 6 |
| 2. Executive Summary | 7 |
| 2.1. Document contributor contact points | 9 |
| 2.2. Foreword | 9 |
| 3. References | 10 |
| 4. Terms and definitions | 11 |
| 4.1. Abbreviated terms | 13 |
| 5. Overview | 15 |
| 5.1. Objectives | 15 |
| 5.2. Domain Use Cases | 16 |
| 5.3. General SELFIE Use Cases | 17 |
| 5.3.1. Organizational architectures | 17 |
| 5.3.2. SELFIE Functionalities | 17 |
| 5.3.3. General Use Case Descriptions | 18 |
| 6. Resource / Content Model | 21 |
| 6.1. W3C Resources Summary | 21 |
| 6.2. ELFIE Resources | 22 |
| 6.3. "In band" and "out of band" resources | 24 |
| 6.4. Resource Resolution Alternatives | 26 |
| 6.5. Contexts | 28 |
| 6.6. Non-information resources | 29 |
| 6.7. Landing content | 30 |
| 6.7.1. Properties for documenting data resources | 31 |
| 6.7.2. Properties for documenting related data | 31 |
| 6.7.3. Properties for relating non-information resources | 32 |
| 6.8. Data content | 33 |
| 6.9. Relationship to OGC-API | 34 |
| 6.10. Properties / Relations | 34 |
| 6.10.1. Schema.org | 35 |
| 6.10.2. Dublin Core | 35 |
| 6.10.3. SKOS | 35 |
| 6.10.4. RDFS | 35 |
| 6.10.5. Geosparql | 36 |
| 6.10.6. OWL-time | 36 |
| 6.11. Domain models Classes / Properties / Relations | 36 |
| 6.12. Summary and outcomes | 37 |
| 6.13. Issues and recommendations | 37 |
| 6.13.1. Landing content type and the URL-14 resource | 38 |

| | |
|--|----|
| 6.13.2. In band and out of band resources | 38 |
| 6.13.3. Domain Ontologies and JSON-LD Contexts..... | 39 |
| Appendix A: Domain Use Cases..... | 40 |
| A.1. US Internet of Water Distributed Data and Observations | 40 |
| A.1.1. Use Case Description | 40 |
| A.1.2. Demo Description and Links | 41 |
| A.1.3. Demo findings and potential next steps | 42 |
| A.2. Groundwater Surface Water Interoperability Pilot - SELFIE Demonstration..... | 42 |
| A.2.1. Use Case Description | 42 |
| A.2.2. Links to Demo Resources | 45 |
| A.2.3. Demo findings and potential next steps | 49 |
| A.2.4. Annex | 50 |
| A.3. BRGM - INSIDE research center - Surface / Ground water linked data gazetteer..... | 55 |
| A.3.1. Use Case Description | 55 |
| A.3.2. Demo Description and Links | 56 |
| A.3.3. Demo findings and potential next steps | 57 |
| A.4. CeRDI VVG - SELFIE Demonstration | 58 |
| A.4.1. Use Case Description | 58 |
| A.4.2. Demo Description and Links | 60 |
| A.4.3. Demo findings and potential next steps | 62 |
| A.5. SELFIE Demonstration Write Up: Loc-I | 63 |
| A.5.1. Use Case Description | 63 |
| A.5.2. Demo Description and Links | 64 |
| A.5.3. Demo findings and potential next steps | 65 |
| Appendix B: Revision History | 66 |
| Appendix C: Bibliography | 67 |

Chapter 1. Subject

This report documents the Second Environmental Linked Features Interoperability Experiment (SELFIE). SELFIE evaluated a proposed Web resource model and HTTP behavior for linked data about and among environmental features. The outcomes are building blocks to establish a system of real-world feature identifiers and landing pages that document them. OGC API - Features was found to be a useful component for systems implementing both landing content and representations of linked-features. More work is needed to establish best practices related to negotiation between varied representations of a feature, observations related to a feature, and for expressing and mediating between varied content from a given resource. These technical / meta-model details were found to be difficult to evaluate given the small number of example implementations and limited number of domain-feature models available for use with linked data.

Chapter 2. Executive Summary

At the outset of the SELFIE project, the team stated:

SELFIE aims to answer the question, what is the Web architecture that will allow us to use linked data for environmental features and observations in a way that is easily adoptable and compatible with World Wide Web Consortium (W3C) best practices and leverages OGC standards? The experiment aims for focused simplicity, representing resources built from potentially complex data for easy use on the Web. While the IE was focused on testing a specific resource model and followed W3C best practices and OGC standards, a wide range of participant-provided domain use cases will be used for testing. Ultimately, this work is intended to satisfy the needs of many use cases and many kinds of features, from disaster response and resilience to environmental health and the built environment.

The business case for the SELFIE can be illustrated considering two use cases:

1. indexing and discovering models and research from public sector, private sector, or academic projects about a particular place or environmental feature.
2. building a federated multi-organization monitoring network in which all member-systems reference common monitored features and are discoverable through a community index.

These use cases imply needs along several dimensions:

1. a shared reference network of environmental features,
2. the ability to use the reference network to index and provide access to information resources from many organizations,
3. support for multiple disciplines' information models, conceptual models, research topics, and monitoring practices.

While the IE did not come to conclusion on all these fronts, it did show that the core Web architecture to support identification of real-world features and retrieving information about them exists and should be pursued in earnest. The architecture has three basic components; referred to here as URI-14, URL-14, and URL-200 resources.

1. A URI-14 resource is one that has an identifier and is itself a real-world entity.
2. A URL-14 resource is one that is the target of a redirect from a URI-14 and provides information about a URI-14 resource.
3. a URL-200 resource is any other resource that would be linked to by a URL-14's content.

These three resource types can be hosted in a wide variety of organizational architectures and/or governance schemes. No one right or wrong solution was found on this front, and the technical solutions explored proved flexible and capable of adapting to many architectural patterns.

These resources were applied in the context of five functionalities:

1. Publication of identified non-information resources
2. Describing a network of linked features
3. Providing landing content about non-information resources
4. Providing structured-data to support search indexing
5. Providing links to representations and related data

These functionalities were seen to be satisfied by four technical use cases that are loosely aligned with the functionalities:

1. Real-world feature identification
2. Landing content and links to other features
3. Structured data for search indexing
4. Links to representations and other data

The details of the URL-14 resource's content were the main subject of debate in the IE. Some important outcomes include:

1. A URL-14's HTTP URL should almost never be the subject or object of a linked data triple. It is a convenience resource about the URI-14. The URI-14 should be referenced rather than the URL-14.
2. The content of a URL-14 should, at the top level, be a set of statements about a single URI-14. While nested, or complex information about the URI-14 could be included, the document should be centered on one real-world feature.
3. Spatial topology, monitoring relationships, and domain-specific associations between real world features should be expressed as relationships between URI-14 identifiers.
4. Associations between URI-14 resources and representations of the feature should be expressed with a <https://schema.org/subjectOf> relationship. Additional nuances of URI-14 to URL-200 resources should be the subject of future work.
5. URL-200 resources with a semantic representation (JSON-LD) can be the object of a <https://schema.org/subjectOf> relation. URL-200 resources that do not have a semantic representation should be represented as a "blank node" [https://en.wikipedia.org/wiki/Blank_node] with a <https://schema.org/url> association to the URL-200 resource.

OGC API - Features was found to be compatible with all of the above and can be used as a core enabling Web API as networks of linked environmental features are established.

At the outset of SELFIE, the team hoped to experiment with use cases related to variation of available content and multiple data providers for URL-200 resources about a single URI-14 resource. Gaining an appreciation for the nuances of the functionalities and technical use cases required in the context of the broadly varied organizational architectures considered was a large task. Further, some basic characteristics of URL-14 resources and the landing content to include in URL-14 landing content needed to be established before further investigation could continue. Given that, future work should investigate issues such as variation of content for a single URL-200

resource, multiple URL-200 representations of the same feature with variation of content across the providers, and content negotiation of URL-14 resources to either directly access URL-200 resources or access differing profiles of URL-14 landing content.

The IE also aimed to make progress on technical solutions with observational data models and domain feature models. This work was largely deferred for the same reasons as discussed above and because publication of domain feature models and domain features themselves is a prerequisite to meaningfully testing how to work with them in the context of observations data models. The technical baseline provided by the first and second ELFIE now sets the stage for this work to move forward.

2.1. Document contributor contact points

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

Contacts

| Name | Organization | Role |
|----------------------|--|-------------|
| David Blodgett | U.S. Geological Survey | Editor |
| Alistair Ritchie | Manaaki Whenua | Contributor |
| Bruce Simons | Federation University Australia | Contributor |
| Eric Boisvert | Natural Resources Canada | Contributor |
| Abdelfettah Feliachi | BRGM - INSIDE environmental information systems research center | Contributor |
| Sylvain Grellet | BRGM - INSIDE environmental information systems research center | Contributor |

2.2. 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 3. References

The following normative documents are referenced in this document.

- [OGC: OGC 17-069r3, OGC API - Features - Part 1: Core](http://docs.openeospatial.org/is/17-069r3/17-069r3.html) [http://docs.openeospatial.org/is/17-069r3/17-069r3.html]
- [OGC: OGC 16-008 GeoSciML - version 4.1 0](https://docs.openeospatial.org/is/16-008/16-008.html) [https://docs.openeospatial.org/is/16-008/16-008.html]
- [OGC: OGC 16-032r2 WaterML 2: Part 4 – GroundWaterML 2 \(GWML2\) – Version 2.2](http://docs.openeospatial.org/is/16-032r2/16-032r2.html) [http://docs.openeospatial.org/is/16-032r2/16-032r2.html]
- [OGC: OGC 14-111r6 WaterML 2: Part 3 - Surface Hydrology Features \(HY_Features\) - Conceptual Model](https://docs.openeospatial.org/is/14-111r6/14-111r6.html) [https://docs.openeospatial.org/is/14-111r6/14-111r6.html]

Chapter 4. Terms and definitions

- **Resource**

an item of interest in the distributed network of environmental data.

- **Non-Information Resource**

A real-world or conceptual object of interest that is identified by a Uniform Resource Identifier bound to the HTTP protocol (HTTP URI).

NOTE

In the context of this document, Non-Information Resources are strictly identified by HTTP URIs. In other contexts, a Non-Information Resource may be identified using other protocols such as Uniform Resource Names.

- **Information Resource**

A digital resource that can be sent as a message over the internet using a protocol such as HTTP. Located using a Uniform Resource Locator (URL).

- **Information Index Resource**

An information resource that provides an index of annotated (metadata) links to information and non-information resources that describe or are related to the non-information resource of interest.

- **Indirect Identifier**

As defined in the [W3C Architecture of the World Wide Web, Volume 1 Section 2.2.3](https://www.w3.org/TR/2004/REC-webarch-20041215/#indirect-identification) [https://www.w3.org/TR/2004/REC-webarch-20041215/#indirect-identification]. In the context of SELFIE, a URI that would usually be used to identify a digital resource is sometimes used as an indirect identifier of a real-world feature.

- **In-band resource**

An in-band resource is one that can provide information according to a given technical architecture. In the context of linked data, an in-band resource can provide Hypertext Markup Language (HTML) and Resource Description Framework (RDF) content serialized as JavaScript Object Notation for Linked Data (JSON-LD) representations. Other representations may also be considered in-band if the specified architecture expects them (GeoJSON for example). In-band resources can extend the linked-data graph.

- **Out-of-band resource**

An out-of-band resource is one that does not adhere to the technical architecture from which it is found. A resource that can provide observation result graphs, XML, CSV, PNG, PDF and JSON representations but no linked-data representation would be considered out-of-band if linked from a linked-data document. Out-of-band resources cannot extend the linked data graph.

- **HyperText Transfer Protocol Uniform Resource Identifier (HTTP URI)**

An identifier with the potential to be used with the HTTP protocol to dereference (look up) the identified resource.

- **HyperText Transfer Protocol Uniform Resource Locator (HTTP URL)**

A type of URI that can be used to locate an information resource.

- **Data Resource**

An information resource providing a representation of a non-information resource.

- **Registry**

Per [ISO 19135, Geographic information, Procedures for item registration](https://www.fgdc.gov/standards/organization/FGDC-SWG/meetings/2005/2005-12-20/ISO%2019135.ppt) [https://www.fgdc.gov/standards/organization/FGDC-SWG/meetings/2005/2005-12-20/ISO%2019135.ppt]: An information system that manages a set of files containing identifiers assigned to items with descriptions of the associated items.

- **Resource Model**

A taxonomy and functional description of the system of non-information, index and data resources.

- **Node**

A source of information about real world features. May be specific to a geospatial or scientific application domain. Acts as a node in a system of linked data providers.

- **Hub**

An aggregator or indexer of information about real world features. Provides integrated information as landing content derived from a community of nodes.

- **Provider**

An originating source of data.

- **Resolver**

A registry system that provides 303 redirection from URI-14s to URL-14.

NOTE

In general, outside the scope of this document and depending on the intent of a given system, a resolver can be more than a 303 registry system. Broadly It could be any intelligent function that adapts the response of a server to the context of request

- **Landing resource**

The ‘default’ information resource provided, through a range-14 303 redirect, when a non-information resource’s URI is dereferenced. An abstract thing, the actual information resource returned is based on content negotiation. HTML → landing page; JSON-LD → landing data. Assumes constant content in the concrete landing page and data.

- **Landing page**

Presentation-oriented HTML representation of the landing resource. Resource description data are included as values of HTML tags and as structured data: JSON-LD in an HTML script tag.

- **Landing data**

Machine-oriented representation of the landing resource. Is the structured data object in the HTML page presented on its own. SELFIE expects the landing data media type to be JSON-LD but

others are allowed, encouraged even, (RDF/XML; TTL; GML; GeoJSON etc).

- **Structured data**

As per Google: <https://developers.google.com/search/docs/guides/intro-structured-data>

- **URI-14**

The HTTP URI identifying a non-information resource. When dereferenced the host will respond with 303 redirect to a URL for an information resource. Content can be negotiated.

- **URL-14**

A URL, provided by the 303 redirect from a non-information resource's HTTP URI, that locates an information resource. Ideally these are kept hidden (not provided as values for in data) as they shouldn't be confused with the non-information resources HTTP URI.

- **URL-200**

A common or garden URL. So called because the most likely HTTP response code is a '200 OK' with content. It could be the URL for a service request or a file on a file server. Content can be negotiated.

4.1. Abbreviated terms

- API - Application Programming Interface
- CSV - Comma Separated Values
- CURIE - Compact Uniform Resource Identifier
- ELFIE - Environmental Linked Features Interoperability Experiment
- GeoJSON - Geographic JavaScript Object Notation
- GML - Geography Markup Language
- GWML2 - Groundwater Markup Language 2
- HTML - HyperText Markup Language
- HTTP URI - HyperText Transfer Protocol
- HY_Features - Surface Hydrologic Features Conceptual Model
- IE - Interoperability Experiment
- JSON - JavaScript Object Notation
- JSON-LD - JavaScript Object Notation for Linked Data
- OWL - Web Ontology Language
- RDF - Resource Description Format
- SELFIE - Second Environmental Linked Features Interoperability Experiment
- TTL - Terse RDF Triple Language
- URI - Uniform Resource Identifier
- URL - Uniform Resource Locator

- XML - eXtensible Markup Language

Chapter 5. Overview

[Objectives](#) provides a high-level overview of how the first ELFIE's outcomes provide context for the objectives of the SELFIE.

[Domain Use Cases](#) describes how domain use cases and more general technical use cases were used in SELFIE.

[Resource / Content Model](#) is the core discussion of the SELFIE experiment. It includes four subsections:

1. [W3C Resources Summary](#)
2. [ELFIE Resources](#)
3. ["In band" and "out of band" resources](#)
4. [Resource Resolution Alternatives](#)

Detailed descriptions of the technical outcomes of the experiment are provided in:

- [Contexts](#)
- [Non-information resources](#)
- [Landing content](#)
- [Data content](#)
- [Relationship to OGC-API](#)
- [Properties / Relations](#)

The report wraps up with [Summary and outcomes](#) and [Issues and recommendations](#) that wrap up and illustrate issues for future work respectively.

Finally, [Domain Use Cases](#) provides summaries of selected domain use cases contributed by SELFIE participants.

5.1. Objectives

The first Environmental Linked Features Interoperability Experiment (ELFIE) sought to answer the question, *"what linked data content should be included in a landing page describing an environmental feature and its relationship to other features and data?"*. Limiting scope in this way allowed the team to avoid the complex issues related to network behavior and the semantics of requesting default or alternate representations of a feature, representations of it, or other features and data in some way related to it. These issues were discussed in the first ELFIE — but often only briefly or in the context of defining what was expressly out of scope for the project [1]. The Second Environmental Linked Features Interoperability Experiment (SELFIE) took these issues on in earnest.

Objectives of SELFIE, from the project charter were:

1. Evaluate a proposed resource model for multi-provider environmental feature and observation registries.

2. Evaluate proposed HTTP behavior for non-information resources and their representations.
3. Design and evaluate linked data feature information index resources with media-type, language, and profile content negotiation as an extension of the building blocks provided by OGC API – Features (formerly called WFS3). Within the context of these objectives, the [functional and operational goals of the first ELFIE](https://docs.opengeospatial.org/per/18-097.html#Objectives) [https://docs.opengeospatial.org/per/18-097.html#Objectives] were upheld.
4. linked-data content for describing and linking features and associated data and
5. maintaining the rigor of OGC and W3C standards and best practices while providing easily-adopted approaches.

These can be summarized with the question, *"what is the expected network behavior and resource model when resolving a Web identifier for a non-information resource?"*. While fairly simple on its face, this question proved to be challenging on a number of levels.

Objectives added during the IE:

6. At a high level, we found that the architectural resource model that seemed to fit our understanding of the problem — a three-tiered resource model of Non-information resources, Meta-resources, and Data-resources — broke down when implemented in web-resources. The distinction between metadata and data is ultimately defined by the use of information and not the information itself.
7. Semantic web technology and the rigor required for systems that support reasoning over a graph presented great opportunity and potential while introducing a level of complexity and technical specificity that was challenging to navigate as a group. The diverse backgrounds and levels of understanding of technologies made communication break-downs all too frequent.
8. There are very few example systems that have implemented solutions to the problem pursued in SELFIE — Web-friendly landing pages for spatial features and related data. Where systems do approach the problem, they have used wide ranging technical and architectural approaches that proved difficult to compare and harmonize. The general lack of common language, standard web-resource models, or common implementation patterns meant the team often felt they were forging their own path through a thicket.

Due to these challenges, many issues discussed in SELFIE were tabled for later once more example implementations have had a chance to experiment and understand what works and why we might choose some approaches over others.

5.2. Domain Use Cases

The SELFIE relied on participants' domain-specific use cases to provide context and drive decision making in the context of the IE. The use cases included hydrogeology, soils, hydrology, and land-survey information. Common to these use cases was the need to work with identifiers for environmental features for which multiple representations are available. Each of the use cases was implemented to one extent or another. Full details of the use cases are included in [Domain Use Cases](#), domain use cases. Taken together and harmonized, these domain-specific use cases provided sufficient scope to determine a useful set of general use cases that are summarized in the following sections.

5.3. General SELFIE Use Cases

General, as opposed to domain-specific, SELFIE use cases are described in the section that follows. To provide greater insight into their purpose, the organizational architecture and functionalities they entail are first described in some detail. The use cases aim to maintain technical rigor while being practical and approachable. This can be seen as a balance or tension, but rather, ease of implementation was used as a filter on technically rigorous solutions—leaving complexity out where the team was not ready to recommend an easy-to-implement approach. While technical, and in some cases very specific, these use cases do not imply complete technical approaches.

5.3.1. Organizational architectures

The SELFIE included applications with a variety of organizational architectures. This diversity resulted from the social, political, and technical setting the applications were situated in. Aspects that were potentially diverse included:

- Single to multiple non-information identifier (URI-14) registry and redirect systems.
- Single to multiple interlinked providers of landing content (URL-14).
- Single to multiple providers of feature representations and other data (URL-200). This diversity required some careful consideration and handling and the solutions explored in SELFIE proved to hold up well across the range of organizational architectures encountered.

5.3.2. SELFIE Functionalities

The linked data architecture that resulted from SELFIE is based on the five functionalities that are described below. These are described as functional use cases that loosely align with the general use cases. These functions were common across practically all the examples considered by the experiment and are presented here as a general set of functions for linked environmental features and related data.

Publication of identified non-information resources is a prerequisite for establishing links between features and related data. Persistence and long-term uniqueness of URIs used to identify non-information resources is helpful but cannot be guaranteed. A robust system of linked data must be able to deal with changes to identifiers through re-indexing or similarity relationships. Similarly, use of common identifiers across organizations is helpful but cannot be guaranteed. Systems of linked data must be able to handle when organizations use different identifiers to refer to the same real-world feature.

A **network of linked features** is formed when considering topological and domain-specific linked data associations between identified features. From an indexing perspective, this network can be "crawled" and indexed by both domain-specific and general web search crawlers. While a rich graph of linked features can be resolved or may exist within a linked data system, the functionality required here is exposure of direct links from one feature to adjacent neighbors such that the linked-feature network can be traversed by a human user or Web crawler.

Landing content is common metadata about a feature and data associated with it. In addition to this common-core metadata, landing content might also include:

- A multi-organization index of information about the feature
- Links to multiple or alternative representations of the feature
- Pre-fetched information (e.g. labels and media-types) about resources.

Structured-data to support search indexing is the representation of landing content that is presented to a web-crawler. The lexicon of this most-default representation must be common to the Web (e.g. schema.org) and the breadth of content focused such that only specific pertinent details for general search, discovery, and general preview (such as a knowledge panel) are included.

Providing links to **representations and related data** is the ultimate purpose of the system of linked data explored in SELFIE. Such resources are generally not natively defined in linked-data formats and cannot be incorporated into the linked data graph directly. As described in detail later, such out-of-band resources must be referred to with associations like schema.org/url [https://schema.org/url] rather than as in-band linked data resources.

5.3.3. General Use Case Descriptions

The content model is best described with the use cases described in the following paragraphs:

1. real-world feature identification
2. landing pages and other default content
3. structured data for search indexing
4. links to representations and other data

The **feature identification** use case involves association of an HTTP URI with a recognized real-world feature. In the most sophisticated implementations, this would be a "URI-14" URI which only ever returns a HTTP-303 see-other directing a client to a "URL-14" which would return landing content. However, a less sophisticated implementation may conflate the URI-14 and URL-14 resources such that the feature identification use case is satisfied with a URL that returns landing content. This was found to be valid and a practical approach. While practical, it must be noted that this should be an exception to the norm and that conflating identifiers for both a non-information resource and an information resource (URI-14 and URL-14), introduces ambiguity with wanting to refer to the actual real-world entity or the digital resource.

The **landing content and network of linked features** use case focuses on the default content and encoding that a search-engine crawler expects. It involves the HTML media-type content returned by default when resolving a feature identification resource whether via 303-redirect or not. Structured data in landing content must be designed in the lexicon of the web, focused on schema.org and other common ontologies and encoded in JSON-LD. HTML content provides useful natural language descriptors and uses appropriate link relations wherever possible. The URL that is used to retrieve landing content could have a number of sophisticated alternative behaviors accessed via HTTP content negotiation and/or appended API patterns, but the default response when the accept header indicates HTML, would typically be designed to satisfy the needs of the landing content use case.

Structured-data for search indexing, what we might call in-band resources, could involve various lexicons and graph-views of linked data that adheres to the RDF data model and are part of a

consistent structured data web architecture. Logically, such content should be returned from the URL used to retrieve landing content if the HTTP-accept header indicates a linked data type such as JSON-LD, html, and other hypermedia and media types included in a defined architecture. With regard to linked data types, use of API patterns such as those introduced in the Linked Data API or HTTP content negotiation by profile may be relevant here, a system may return all known associations to the identified feature, or a custom view of an extended linked data graph that meets the needs of an implementation. Given that multiple profiles of linked data may be available, the linked data rendered in the `<script>` header of the HTML representation of a landing resource may provide different content than other linked-data representations. This follows from the fact that many linked data use cases don't necessarily focus on search-engine indexing.

Data that represent or are otherwise related to features are what we refer to here as out-of-band. Such data are not part of the system of linked data and related content. These might be a complex GML representation of a feature, an image or map, a report, or a JSON representation of a timeseries. The distinction is that a given representation of a resource is either compatible with a technical architecture (i.e. can be parsed and handled by software that works with it) or is not (i.e. is opaque to software that works with the architecture). In-band content can directly extend the linked-data graph and out-of-band cannot.

[Figure 1](#) provides a summary of the SELFIE functionalities in the context of the range of potential implementation sophistication.

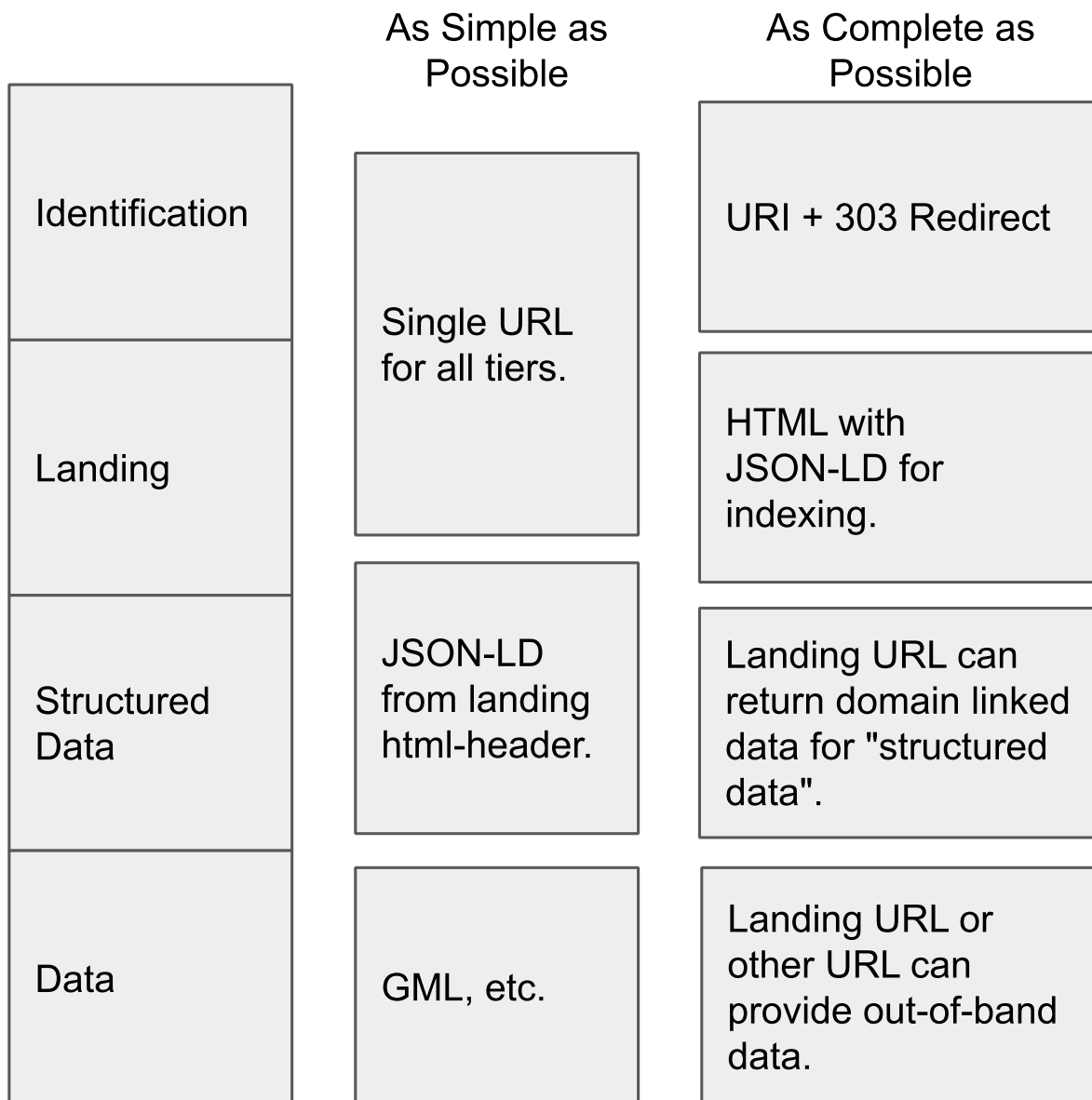


Figure 1. The four functions of the SELFIE general use cases. The most simple implementations, while limited, may use a single resource and content negotiation for all four functions. A complete SELFIE implementation would use separate resources for each function with linked-data hypermedia to facilitate discovery and access.

Chapter 6. Resource / Content Model

Resources are the stuff of the internet. Is *a resource* 1) its identifier, 2) content retrieved by dereferencing its identifier, or 3) some abstract notion identified by a URI and described by dereferenced content? The problems pursued by SELFIE made the importance of these distinctions clear and, to some extent, found some answers. More often than finding answers, the SELFIE found that the technical baseline pertaining to the problem is rich and relatively un-explored for environmental data use cases. That is to say that given the technical baseline available to the community, implementation and building an understanding of modern technologies should proceed in order to better understand pain points and find where additional complexity really is required vs. where existing technologies can satisfy the real need.

6.1. W3C Resources Summary

The SELFIE model for web resources is based on the notion of *information resources* and *non-information resources* as defined by the W3C document ‘[Dereferencing HTTP URIs](https://www.w3.org/2001/tag/doc/httpRange-14/2007-08-31/HttpRange-14.html)’ [https://www.w3.org/2001/tag/doc/httpRange-14/2007-08-31/HttpRange-14.html]. This summarizes the so-called ‘Range-14’ decision that, while unofficial, is useful especially when read in conjunction with ‘[Cool URIs for the Semantic Web](https://www.w3.org/TR/cooluris/)’ [https://www.w3.org/TR/cooluris/].

Information resources are the currency of the web - the pages and data served in a digital form to be consumed by web browsers and applications. Non-information resources are the things these information resources may describe (for example people, mountains, or Aristotelian philosophical constructs). For SELFIE, the critical distinction between them is that an information resource has a location (expressed as a URL) while a non-information resource has identity, expressed as an HTTP URI.

HTTP URIs have two roles:

1. a globally unique identifier that can be presented as an identifier string and matched with other identifier strings to establish sameness; and
2. the location of a resolver that can redirect enquiries to the location of an appropriate information resource.

Several representative information resources describing given non-information resources may exist. For example, information resources available according to various media-types, ontologies and/or content models (e.g. data quality rules, controlled vocabularies or units of measure) may be available. SELFIE sought to refine the classification of these representative information resources to help data-providers clarify what type of resource they publish providing guidance for implementation of resolvers, catalogs, and data services.

The Environmental Linked Features in SELFIE are non-information resources. These can be thought of in the following groups (in the following, all ‘features’ are ‘non-information resources’):

1. Domain features - identifiable environmental things in the world.
2. Sampling features - human-created representative samples of domain features. These features exist to provide metadata about how the feature was described and how robust/representative

that description is. An information resource describing a domain feature would logically use, link to, or summarize data related to sampling features.

3. Semantic resources: the ontologies and vocabularies used to structure and populate the descriptions of domain and sampling features. These are formalized using OWL and RDF.

Each of these groups occupy different meta-levels but can be interlinked. For example, allowing an agent traversing a knowledge graph to move from the description of a domain feature to how that description was obtained (sampling features). The links between the domain and sampling features are described with [ontologies such as SOSA](https://www.w3.org/TR/vocab-ssn/) [https://www.w3.org/TR/vocab-ssn/]. Given that SELFIE has adopted JSON-LD as its RDF encoding syntax, links to semantic resources are provided via JSON-LD contexts that map JSON keys and types to RDF ontology property and class HTTP URIs.

For SELFIE, nodes in a knowledge graph are non-information resources, and the nodes of most interest are those that identify domain features. Sampling and semantic resources provide metadata that describe those features and organize available knowledge about them. Therefore, the SELFIE knowledge graph describes relationships between non-information resources in the 'real world'. The links between information resources that describe them **is an important but separate concern**. In SELFIE, how to transition from non-information resources in the knowledge graph to those in the web of information resources was an important consideration where further work is needed.

6.2. ELFIE Resources

At the outset of SELFIE, the team was thinking in terms of a "three-tiered resource model" where "resource" was a thing identified by a URI. The resource model involved "non-information resources", "meta-information resources", and "data-information resources". Conceptually, the "non-information" "meta" and "data" scheme is useful, but the word "resource" is wrong when applied to the terminology of particular technologies (such as HTTP Resources). As such, the team found a need to change its language to more accurately reflect what was found to be useful ways to describe what is actually a set of use-cases that can be described in terms of non-information, metadata, and data.

Everything in this scheme is identified by a HTTP URI. In general, we have three categories that can be described as follows:

1. non-digital things that are not information,
2. digital things that provide meta-information about non-information things, and
3. digital things that are information representing or characterizing other things.

Tier 1. is clear — there should be URIs that only ever return a 300 series redirect and are identifiers for real-world features. Tier 2. is hard to define precisely. It can only be defined strictly by the application retrieving it rather than by specific characteristics of its content. Tier 3. is clear in most cases but has potential overlap with tier 2 in that some applications may consider metadata about a feature to actually be data representing the feature. Since self-describing data always contains metadata, we would expect most if not all of tier 2 to be contained in tier 3.

If we think about it this way, then tier 2 is a convenience layer to achieve a certain functional goal. In SELFIE, tier 2 is a convenience layer for search-engine crawlers and humans looking for an idea

of what a real-world feature is, what it's related to, and if there's interesting data available representing it.

This should make it clear that resource is the wrong word to describe the distinction being drawn here. It is ok for tier 1, but it breaks down for tier 2 and 3. A single URL may have one or more representations designed to be metadata and one or more representations intended to be data — each intended for a different use pertaining to the same real-world feature. This is not saying that tier 2 and tier 3 are always to be represented variously based on the same URL — they very well may be represented as different resources. This technical diversity is what SELFIE sought to enable.

Consider a typical use case considered by the SELFIE:

As a Web user, I want to find all the information available for an environmental feature, so I can find what I'm looking for and retrieve it.

As the project dissected this use case the HTTP-Range 14 (303 redirects), OGC API - Features (html landing pages for features), and schema.org JSON-LD in a `<script>` tag of a landing page were all embraced as useful technical solutions that serve it. None of the above requires alternative resource representations (media types). With an HTML landing page, a human user or crawler designed around natural language and schema.org are satisfied.

However, the IE recognized the potential to bring structure to data (including semantically enabled data) underlying the landing page content. Going further, the SELFIE was premised on the idea that html landing pages are layered on top of a potentially wide range of data systems that need such a discovery layer. Given this, while not addressed specifically in SELFIE, alternative media types and content negotiation are expected in a system that the SELFIE model is applied to. However, the complexity and lack of broad implementation made making progress on this front difficult.

Participants in the IE agreed that they would avoid specifying how to use content negotiation between the "meta" and "data" tiers. Standards for content negotiation by profile are emerging but we have not been able to evaluate them rigorously. Instead, SELFIE was limited to describing how to advertise that multiple content-types are available for a given URL in structured JSON-LD data. The scope and summary architecture are described in [Figure 2](#).

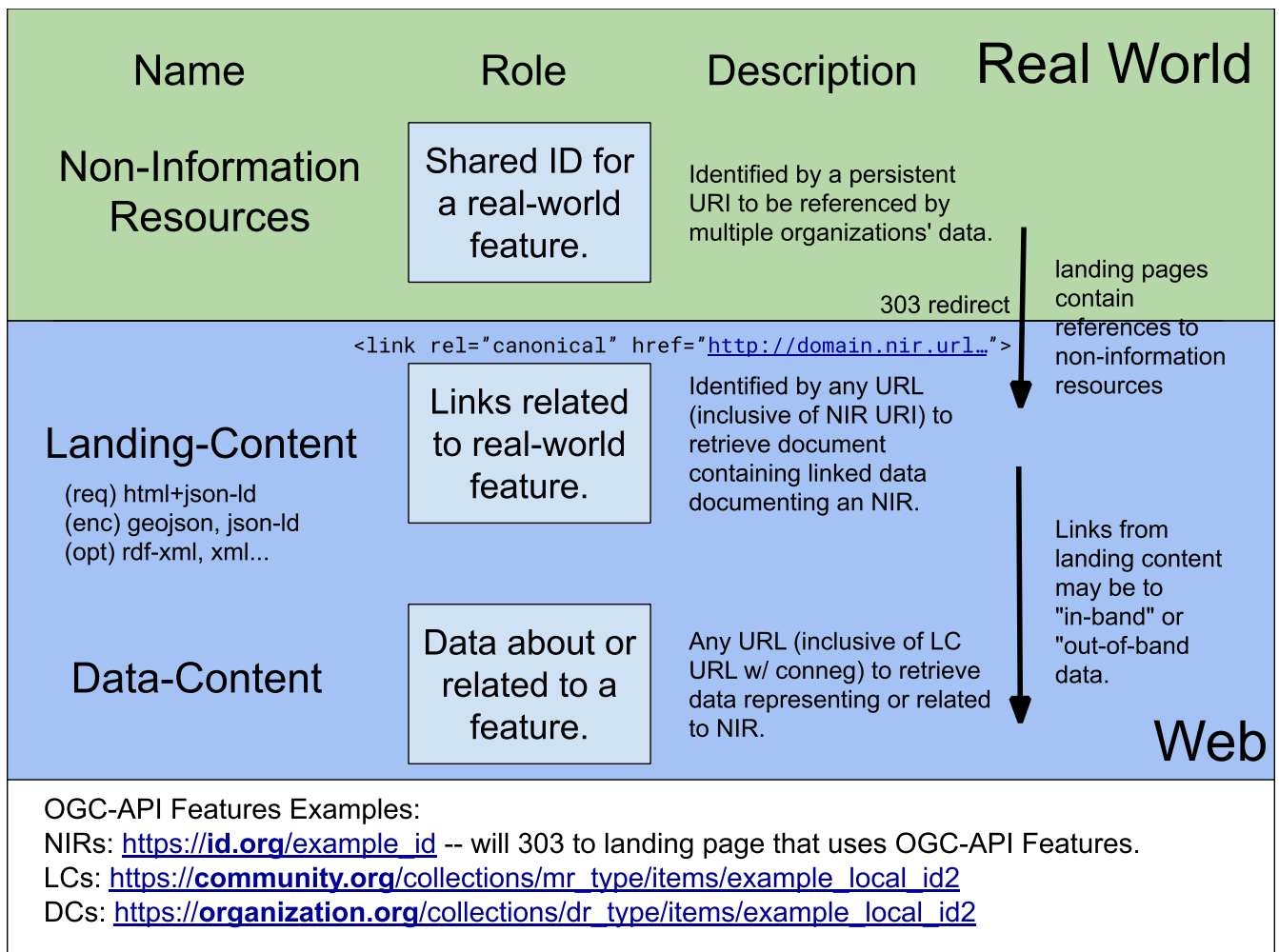


Figure 2. Summary of the SELFIE resource / content model showing that there are Non-information resources which 303 redirect to a resource intended to provide "landing content". The distinction between landing-content and data-content is use-case specific and methods for negotiating between the two is left for future work.

6.3. "In band" and "out of band" resources

The idea of "in-band" and "out-of-band" has been brought up as a useful distinction between resource representations that can provide information that is useful to a given application (in-band) and resource representations that are opaque to an application (out-of-band). In reality, there are many bands that correspond to various applications. Here, we define the SELFIE-band which is intended to foster interoperability toward the goals of the IE.

There are three defining characteristics of the SELFIE "band":

1. *The resources:* ELFIE is a graph of non-information resources.
2. *The access protocol:* The HTTP protocol (with no extensions [perhaps controversial?]) with responses managed according to the range-14 decision.
3. *The encoding:* HTML + JSON-LD and JSON-LD in which ELFIE non-information resources are identified, and linked to, using the JSON-LD @id key.

A SELFIE resource is recognizable because:

1. it has an @id;

2. it has a format property that includes application/ld+json; This limited set of criteria covers the important architectural concerns. It implies an 'architectural profile' that encompasses `@id`, `schema:url`, `dct:format`, and `rdfs:label` and therefore basic resource description and linking.

To illustrate the distinction, consider the following JSON-LD example which has one `schema:sameAs` and one `schema:subjectOf` property for an identified feature:

```
{
  "@id": "https://feature.id",
  "http://schema.org/sameAs":
  {
    "@id": "https://somesresource",
    "http://purl.org/dc/terms/format": "application/ld+json;",
    "http://www.w3.org/2000/01/rdf-schema#label": "A resource that can extend the
    linked data graph."
  },
  "http://schema.org/subjectOf":
  {
    "http://schema.org/url": "https://blobby",
    "http://purl.org/dc/terms/format": "application/xml;",
    "http://www.w3.org/2000/01/rdf-schema#label": "blobby thing with the feature as
    its subject"
  }
}
```

Alternatively, when we resolve `https://feature.id` we might get a more limited document that does not include pre-fetched content about `https://somesresource`:

```
{
  "@id": "https://feature.id",
  "http://schema.org/owl#sameAs":
  {
    "@id": "https://somesresource"
  },
  "http://schema.org/subjectOf": {
    "http://schema.org/url": "https://blobby",
    "http://purl.org/dc/terms/format": "application/xml;",
    "http://www.w3.org/2000/01/rdf-schema#label": "blobby thing with the feature as
    its subject?"
  }
}
```

Which would mean we would need to resolve and interrogate `https://somesresource` to retrieve information needed to decide whether it is of interest, which is possible with the "in-band" `https://somesresource`, and might give us the JSON-LD below, but impossible with the "out-of-band" `https://blobby` which might only return xml or linked data using an unknown ontology.

```

{
  "@id": "https://somerresource",
  "http://www.w3.org/2000/01/rdf-schema#label": "A resource that can extend the linked
data graph.",
  "http://purl.org/dc/terms/format": "application/ld+json;",
  "http://www.w3.org/2000/01/rdf-schema#seeAlso": "https://someOtherThing"
}

```

Note that we have avoided discussing `@type` and `conformsTo`. Use of these properties, while valuable, introduces complexities that were determined to go beyond the scope SELFIE was able to accomplish.

6.4. Resource Resolution Alternatives

The Range-14 decision, to identify real-world features with URIs that HTTP-303 redirect to resources providing information about the real-world feature, was accepted by SELFIE. [Figure 3](#) illustrates the complete solution.



Figure 3. Complete range-14 resolution behavior.

However, to simplify implementation, some landing resource providers skip the 303 redirect entirely, using a URL for a landing resource as an indirect identifier of a real world feature. [Figure 4](#) illustrates this less complicated, but limited approach.

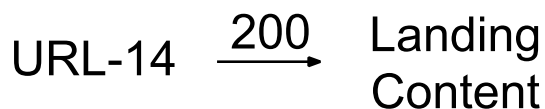


Figure 4. Indirect identification of a feature where a URL is used as an indirect identifier for a real-world feature.

There are two related problems with the indirect identification approach: one technical and one social. Both issues stem from the need to maintain stable identifiers for real world features and very real needs to change URLs to retrieve digital resources.

The technical issue is related to how URLs are used to drive server behavior. Changes to server software implementation often necessitate changes to URL paths or parameters. The requirement to maintain URL stability is in conflict with this and causes needless complexity for server-

implementers.

Socially, real-world feature identification is a process undertaken by a group of people that is likely not the same as those who implement the server software used to retrieve information about those features. Identification of features may work best with a different URI structure than retrieval of digital information about those features; forcing the two groups of people to reconcile these patterns is an unneeded, complicated, and likely fraught interaction that can be eliminated by separating real world feature identification from information index resource identification.

Adding content negotiation to the discussion of resource resolution, a 303 redirect works fine as long as the client passes the same accept header to the redirect target URL. However, there is a common content negotiation override practice involving URL parameters such as `?f=mime-type` or `?format=mime-type` that may be desirable to have passed along as part of a 303 redirect. Some SELFIE participants support such mime-type overrides, but additional experimentation will be required to determine if there is a solution that should be recommended for this in general. Note that this says nothing about [content-negotiation "by profile"](https://www.w3.org/TR/dx-prof-conneg/) [https://www.w3.org/TR/dx-prof-conneg/], an emerging technique that was decided to be beyond the scope SELFIE would be able to address.

Extending the resource resolution use case to include retrieving representations of a feature introduces additional functions that were the subject of some SELFIE experiments. Two such resolution schemes were tested. One required a client to inspect information index hypermedia and make an additional request for an available representation. The other used media-type content negotiation to return a representation available via that media-type directly from a URL-14 indirect identifier without the client needing to review information index hypermedia. These two schemes are illustrated in [Figure 5](#). These alternatives are equally valid and further work is needed to determine if one is preferable to the other.

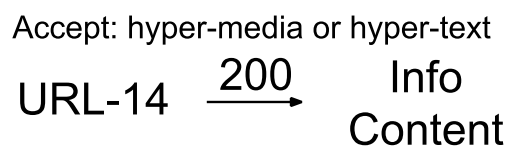
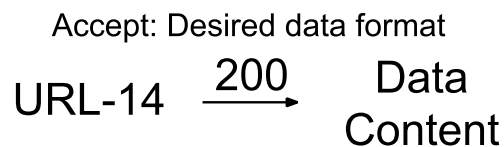
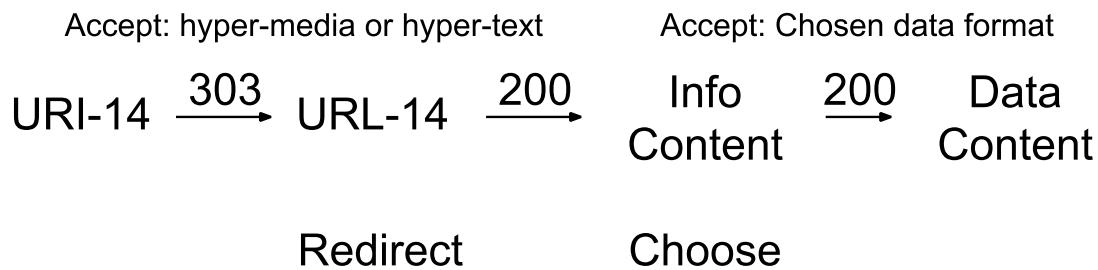


Figure 5. Hypermedia-driven resource resolution (above) versus content negotiation-driven resource resolution (left). While less complex, the content negotiation-driven approach is limited to implementation on a single domain and requires a significantly more complex resolver implementation.

6.5. Contexts

The SELFIE evolved several JSON-LD contexts that were initially created for the first ELFIE. IE participants also worked to establish a number of contexts containing UML classes and associations (used as feature types and property associations in RDF/JSON-LD) from domain feature models such as GeoSciML, GWML2 and HY_Features. The content of the SELFIE-contexts is described in the meta-content section below. The purpose of the domain-specific contexts is to provide semantics that describe the basic associations between non-information resources (e.g. upstream/downstream associations) and domain-specific type (e.g. a feature is a water well). In order to ensure the contexts are usable by a broad audience and to foster adoption, the contexts and ontologies that they reference were kept purposefully focused and minimal; including feature types and associations from the normative UML models only.

With regard to style of contexts themselves, IE participants agreed that we should alias only properties and not classes/feature types. Only namespace URI prefixes (e.g. "schema": "http://schema.org/") and property keys (e.g. "sameAs": "schema:sameAs") are aliased. Classes are not, instead they are returned as compact URIs (CURIs) in the JSON documents: "@type": "gw:GW_HydrogeoUnit".

The argument for namespace is very common practice. They simplify documents, making URIs shorter and clearer in documents. For example, gw:GW_Well is easier to work with than `https://www.opengis.net/def/gwml2#GW_Well`.

The first ELFIE made the decision not to use CURIs for property key names as a simplification to

make JSON-LD documents more approachable and limit colons in JSON keywords. [See this description of ELFIE context design for more... context.](#) [https://opengeospatial.github.io/ELFIE/json-ld/] The downside of this decision is we lose the ability to distinguish between the use of the same label in different JSON-LD contexts. CURIs disambiguate identical property keys (`schema:comment`; `rdfs:comment`) but this can also be accomplished with longer key aliases (`"schemaComment": "schema:comment"`; `"rdfsComment": "rdfs:comment"`). This workaround will be required very rarely and isn't necessary for classes (`@type` values) as they are property values and not used as JSON keys so problematic characters (like colons) are not a problem.

6.6. Non-information resources

The SELFIE content model, with its loosely-defined non-information, meta, and data tiers, is focused on the nature of a resource and not the structure of the identifier used to dereference the resource. However, there is a need to establish expected behavior when dereferencing URIs for information at each of these three tiers. As discussed in [ELFIE Resources](#), the SELFIE project found that the distinction between meta and data tiers is loose and should be the subject of future work. The IE considered two behaviors between non-information and information resources that provide landing content: 1) A 303 redirect from a URI-14 to a URL-14, and 2) Indirect identification of the non-information resource where the URI-14 and URL-14 are identical.

The distinction between non-information resources and resources that provide landing-content located with different URIs introduces the potential for confusion. This confusion was anticipated to produce two issues:

1. Search engine Indexes of landing content might incorrectly record associations with the URL-14 (a less-permanent information identifier) rather than the URI-14 (a more permanent non-information identifier) that the landing content references.
2. URL-14s that are displayed in a browser URL bar after a 303-redirect could easily get copied and used as an indirect identifier when a URI-14 exists and should be used instead. SELFIE participants tested the former by exposing a pilot infrastructure to the public internet and found that, as long as landing content references the URI-14, search engine indexes work appropriately. The second issue can be mitigated by placing prominent URI-14s on the html representation of landing content and structure URL-14s in such a way that it is clear that they are not meant to be used as identifiers.

Indirect identification of URI-14s using URL-14s has been tested by SELFIE participants successfully. It was found that this practice limits flexibility of systems to accommodate multiple agencies and the tight coupling of URI-14 and URL-14 introduces technical challenges for landing content resolution. However, the practice is less complicated from a user perspective and, as long as the URL-14s are stable, is compatible with future decoupling of URI-14 and URL-14 when the separation is needed. These two cases, 303 and 200 response codes to a non-information resource identifier, illustrate the flexibility of the SELFIE content model as well as the nature of landing-content as a convenient resource that encodes content relating URI-14s to each other and data resources.

An important recommendation IE participants agreed to is that when a 303-redirect from a URI-14 to URL-14 exists, the **URL-14 should not be referenced in structured data**. This follows from the fact that the URL-14 is a locator for landing content with one and only one subject—the URI-14. That is, a JSON-LD representation of a URL-14 resource would have one and only one root `@id` that is the URI-14 that the JSON-LD documents. As such, the URL-14 should never appear in a linked data

graph as it will never return linked data about itself.

6.7. Landing content

The following sections describe the purpose and some logic for SELFIE landing content.

As an overview, [Table 1](#) contains the most precise use-case descriptions the SELFIE team was able to craft that pertain to URL-14 landing-content "groups". These groups can be seen as building blocks that are used to construct landing-content structured data.

Table 1. SELFIE landing-content groups, their purpose, and descriptive text.

| Content Group | Purpose | Description |
|-------------------------------|---|--|
| <i>schema.org properties</i> | <i>discovery; indexing</i> | Statements about the non-information resource using schema.org properties. These have literal values (string, number etc) or structured values (collections of sub-properties that would appear as JSON object or an RDF blank node). |
| <i>content properties</i> | <i>discovery; indexing</i> | Statements about the nature of the <i>information</i> resource, e.g. available media types or data profiles. |
| <i>representation links</i> | <i>discovery; indexing; integration</i> | Links to other representations of the <i>same resource</i> . The representations likely come from a different provider and/or system. These links will be objects/blank nodes rather than simple URI/Ls. The object will include a small set of prefetched properties (schema.org, content or descriptive) from the target resource |
| <i>descriptive links</i> | <i>non-information resource description</i> | Links to related non-information resources according to the thing's place in the world. For example a link could describe a spatial, temporal topological or semantic relationship, a sample's feature of interest, or a river's reach's upstream reach. As such these links use properties defined in our domain ontologies. As with representation links, these links will be objects/blank nodes rather than simple URI/Ls. |
| <i>descriptive properties</i> | <i>non-information resource description</i> | Statements about the nature of the non-information resource using domain ontology properties. These have literal values (string, number etc) or structured values (collections of sub-properties that would appear as JSON object or an RDF blank node). |

The content groups described in [Table 1](#) have been encoded in JSON-LD contexts that can be used as a guide for implementers and by-reference in structured data. These contexts are grouped functionally around indexing, documenting data-resources, describing topological feature associations, and providing simple spatial representations of features.

6.7.1. Properties for documenting data resources

`elf-index.jsonld` [<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld>] is intended for search engine indexing and discovery. It contains properties needed for a general preview of a feature. These were chosen to be compatible with search-engine indexing and to give a complete core set of information about a non-information resource.

```
{
  "@context": {
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "schema": "http://schema.org/",
    "skos": "http://www.w3.org/2004/02/skos/core#",
    "name": "schema:name",
    "description": "schema:description",
    "spatialCoverage": "schema:spatialCoverage",
    "geo": "schema:geo",
    "image": "schema:image",
    "subjectOf": "schema:subjectOf",
    "sameAs": "schema:sameAs",
    "note": "skos:note"
  }
}
```

6.7.2. Properties for documenting related data.

The `elf-index.jsonld` [<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld>] context contains a set of properties to be used to document data about a non-information resource. These can be thought of as "pre-fetched" information about an information resource. As described below, these can be used as properties of an "in-band" information resource or properties of a blank node that describes an "out-of-band" information resource. `conformsTo` is included here to be complete but the specifics of their use for interoperability should be the subject of future work.

These properties are useful for documenting any information resource in structured data whether a feature-representation, associated with a `schema:subjectOf` or `schema:sameAs` property, or other related data, associated with a more specific property such as a sampling relationship.

```
{
  "@context": {
    "dct": "http://purl.org/dc/terms/",
    "schema": "http://schema.org/",
    "provider": "schema:provider",
    "url": "schema:url",
    "conformsTo": "dct:conformsTo",
    "format": "dct:format",
  }
}
```

As is shown in examples in the ["In band" and "out of band" resources](#), these properties can be used

in two ways.

An in-band resource might be documented as follows:

```
{
  "@id": "https://in-band-information-resource",
  "about": {
    "schema:description": "Description of the in-band information resource, see
https://schema.org/about."
  },
  "provider": {
    "schema:name": "Name of the provider"
  },
  "conformsTo": "https://identifier-of-conformance-target",
  "format": [
    "mime-type-1",
    "mime-type-2"
  ]
}
```

An out-of-band resource might be documented similarly but using a blank node as follows:

```
{
  "url": "https://out-of-band-information-resource",
  "about": {
    "schema:description": "Description of the out-of-band information resource,
see https://schema.org/about."
  },
  "provider": {
    "schema:name": "Name of the provider"
  },
  "conformsTo": "https://identifier-of-conformance-target",
  "format": [
    "mime-type-1",
    "mime-type-2"
  ]
}
```

6.7.3. Properties for relating non-information resources

In addition to the `elf-index.jsonld` [<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld>] context, SELFIE participants created a `sosa.jsonld` [<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/sosa.jsonld>] context which includes a `isFeatureOfInterestOf` association which can be used to associate observational data with a feature of interest. The following is an extremely minimal example that demonstrates this.

```

{
  "@context": [
    "https://opengeospatial.github.io/ELFIE/contexts/elfie-2/sosa.jsonld",
    "https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-data.jsonld"
  ],
  "@id": "https://non-information-resource",
  "isFeatureOfInterestOf": {
    "@type": "Observation",
    "hasResult": {
      "url": "https://url-to-retrieve-observation-results"
    }
  }
}

```

elf-network.jsonld [<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-network.jsonld>] is a set of spatial and temporal topological properties that can be used to relate non-information resources in space and time.

```

{
  "@context": {
    "gsp": "http://www.opengis.net/ont/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"
  }
}

```

Many other contexts based on domain data models were created as part of the SELFIE. These can be seen at the [SELFIE contexts web page](https://opengeospatial.github.io/ELFIE/contexts/elfie-2/). [<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/>]

6.8. Data content

Resources containing data content are extremely diverse. Examples include but are by no means limited to geospatial feature data whether a feature of interest or a reference feature, monitoring result data, monitoring location data, and related remote sensing data. As described above, such data can be said to be "in-band" or "out-of-band". The former would be a data resource that generally conforms to the system of linked data, GeoJSON, and HTML prescribed by the OGC-W3C Spatial Data on the Web best practices and emerging practices such as is described here. The latter is any other data resource that, while of interest and associated with a non-information resource, does not conform to linked data / semantic web practices.

The distinction between what is landing content and what is data content depends on the context in which the resource is being accessed. That is, in one context, landing content will be seen as data about a non-information resource; in another context, that same landing content will be used merely as hypermedia and metadata to help choose data content of interest. Because of this, semantic annotation of data will look very similar to landing content except that the URL for a resource that is intended to provide landing content (a URL-14) will not appear in the subject or object of linked data. The URL of a resource that provides data content (a URL-200) must appear in the subject or object of linked data.

The SELFIE focused most of its efforts on details of landing content and how to link to data content. The actual structure of linked data or way to architect resources that provide data content is assumed to be either status quo or left for future work. The potential for resources to have multiple media-type formats and potentially multiple profiles that map onto certain use cases is of great interest and documentation of alternate formats and media types is supported by the landing-content concepts described here.

6.9. Relationship to OGC-API

OGC API - Features was evaluated for providing both URL-14 landing-content and URL-200 data-content and found to be largely compatible with both. For provision of landing content, a feature would be provided as JSON-LD and HTML with JSON-LD structured data in a `<script>` element containing the URI-14 as its primary subject. In this case, hypermedia containing links to items in a feature-collection would contain links to URI-14s rather than display any URL-14s. Since OGC API-Features supports feature-level access by item id (`/collections/{collectionid}/items/{itemid}`) or a feature attribute (`/collections/{collectionid}/items?uri=https://id.com/id`) it can be used as a flexible building block in linked data architectures.

During the IE, SELFIE participants funded, drove, tested, enhanced (depending on the cases) the support for JSON-LD in two major implementations of OGC API - Features namely:

- PyGeoAPI
 - Since 0.7.0 version : <https://docs.pygeoapi.io/en/latest/configuration.html#linked-data>
 - <https://github.com/geopython/pygeoapi/pull/246>
- Geoserver
 - JSON-LD community plugin is available since Geoserver 2.16.2 release (<http://blog.geoserver.org/2020/01/22/geoserver-2-16-2-released/>.)

6.10. Properties / Relations

The logic for selection of relations in SELFIE followed from the more content-focused ELFIE. That logic is described in detail in the [ELFIE engineering report](https://docs.opengeospatial.org/per/18-097.html) [https://docs.opengeospatial.org/per/18-097.html] and on the [ELFIE contexts web page](https://opengeospatial.github.io/ELFIE/json-ld/). [https://opengeospatial.github.io/ELFIE/json-ld/] To summarize, broad adoption and commonality on the web was the primary driver of selection of relations. Spatial topology, temporal, and monitoring concepts were taken from established OGC and W3C ontologies, and domain concepts were sought from OGC domain feature models (GeoSciML, GWML2, HY_Features). The following lists relations by source with comments for

selected relations as necessary.

6.10.1. Schema.org:

<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld> [https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld]

- <https://schema.org/name> [https://schema.org/name]
- <https://schema.org/description> [https://schema.org/description]
- <https://schema.org/geo> [https://schema.org/geo]
- <https://schema.org/image> [https://schema.org/image]
- <https://schema.org/subjectOf> [https://schema.org/subjectOf]
 - Used as a generic association between URI-14s and URL-200s, i.e. between a real world feature identifier and a URL that will provide data about the feature.
- <https://schema.org/sameAs> [https://schema.org/sameAs]
 - Used very infrequently to establish that two resources are the same but available via two URIs. A domain-specific association should almost always be used.
- <https://schema.org/provider> [https://schema.org/provider]
- <https://schema.org/url> [https://schema.org/url]

6.10.2. Dublin Core:

<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld> [https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-data.jsonld]

- <http://purl.org/dc/terms/conformsTo> [http://purl.org/dc/terms/conformsTo]
- <http://purl.org/dc/terms/format> [http://purl.org/dc/terms/format]

6.10.3. SKOS:

<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld> [https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld]

- <https://www.w3.org/2009/08/skos-reference/skos.html#note> [https://www.w3.org/2009/08/skos-reference/skos.html#note]

6.10.4. RDFS:

<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld> [https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld]

- <https://www.w3.org/2000/01/rdf-schema#label> [https://www.w3.org/2000/01/rdf-schema#label]

6.10.5. Geosparql:

<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-network.jsonld>

[<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-network.jsonld>]

- <http://www.opengis.net/ont/geosparql#sfIntersects> [<http://www.opengis.net/ont/geosparql#sfIntersects>]
- <http://www.opengis.net/ont/geosparql#sfTouches> [<http://www.opengis.net/ont/geosparql#sfTouches>]
- <http://www.opengis.net/ont/geosparql#sfWithin> [<http://www.opengis.net/ont/geosparql#sfWithin>]

6.10.6. OWL-time:

<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-network.jsonld>

[<https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-network.jsonld>]

- <https://www.w3.org/TR/owl-time/#time:after> [<https://www.w3.org/TR/owl-time/#time:after>]
- <https://www.w3.org/TR/owl-time/#time:before> [<https://www.w3.org/TR/owl-time/#time:before>]
- <https://www.w3.org/TR/owl-time/#time:intervalAfter> [<https://www.w3.org/TR/owl-time/#time:intervalAfter>]
- <https://www.w3.org/TR/owl-time/#time:intervalBefore> [<https://www.w3.org/TR/owl-time/#time:intervalBefore>]
- <https://www.w3.org/TR/owl-time/#time:intervalDuring> [<https://www.w3.org/TR/owl-time/#time:intervalDuring>]

6.11. Domain models Classes / Properties / Relations

In order to use classes and properties (or relations) defined in domain models, the SELFIE project needed minimal ontologies to reference. "Minimal ontologies" means having basic classes and relations (associations) from a model without the constraints and axioms that could otherwise be generated in the process of transforming UML models to OWL ontologies. The focus was on having resolvable names for the classes and properties that are used in simple linked data payloads. The packaging pattern used in organizing UML domain models was also not used in this exercise, i.e. one UML model = one minimal ontology.

The creation of such minimal ontologies for a specific domain model was realized by using [ShapeChange](https://shapechange.net) [<https://shapechange.net>] processes for the every package of that model, with rules that kept the minimum creations of classes from FeatureTypes and properties with their range definitions. Since there is no rule (known to the IE participants) that specifically constrains the creation of properties from associations only, a second step was to filter (using SPARQL queries) properties by their ranges to keep only ObjectProperties that originated from associations in the UML Model. The last step was to merge content from different packages into one simple ontology having one base URI. This exercise was performed on GeoSciML, HY_Features and GWML2 domain models.

While this simplification has its merits with regards to linked data applications, a discussion with the OGC Naming Authority (OGC-NA) about how to push this ontologies behind OGC-NA servers concluded that they cannot be published in this state. These minimal ontologies do not respect the

OGC-NA content negotiation and naming policy [patterns](https://docs.openeospatial.org/pol/09-048r5.html) [https://docs.openeospatial.org/pol/09-048r5.html], which allows having different possible views of domain models. The inclusion of all properties (and not only associations) respecting the packaging and naming pattern as defined following the UML models should be guaranteed before issuing a request for integration in the OGC-NA server. A work is being carried out to capitalize on what have done for the minimal ontologies to have more detailed and well-organized ontologies that respect the OGC-NA specifications. This would provide a better way to have resolvable JSON-LD contexts directly from the OGC-NA server.

6.12. Summary and outcomes

SELFIE participants experienced difficulty communicating about the nature of the issues at hand. Examples included, but were not limited to:

1. complexity introduced by the relationships between domain feature models
2. the nuances of observational feature relations (e.g. monitoring vs feature of interest)
3. the particulars introduced by the RDF data model (e.g. semantic vs non-semantic resources)
4. URIs as identifiers and/or locators
5. json-ld reserved keys and "context" encoding of RDF

As a result, the IE's primary outcomes are a deeper understanding of the problem at hand rather than fully-fledged and tested solutions.

First and foremost, the IE demonstrated that use of the HTTP-Range14 303 redirection between so-called URI-14 URIs and URL-14 resources that provide "landing content" has great utility and is not incompatible with modern web infrastructure. There is a strong logical and technical case for using URI-14 → URL-14 redirection. A critical additional consideration here is that **URL-14 HTTP URIs should almost never be used as identifiers** in and of themselves. That is, landing content should only ever be retrieved by first dereferencing the URI-14 identifier.

The SELFIE refined the linked data predicates defined for use in landing content by the first ELFIE especially as it relates to links between and among URI-14 identifiers and URL-200 data resources. While more work is needed here, the core-relation to link between identification and representation is <https://schema.org/subjectOf>. Other aspects of the ELFIE content model (such as domain-feature model associations between URI-14 identifiers) were vetted, and proved to work in the context of URI-14 → URL-14 redirection. Further work is needed to enrich the semantics between identification and representation — especially in the context of multiple representations of the same feature where each representation is of a different type.

While the SELFIE set out to test content negotiation by profile and potentially other ways to negotiate between various representations of a resource, the complexities and communication difficulties regarding the core-problem proved enough for one IE.

6.13. Issues and recommendations

6.13.1. Landing content type and the URL-14 resource

A major issue the IE could not resolve arises from the URI-14 → URL-14 distinction. The content returned by a URL-14 is ostensibly an information resource about a URI-14 non-information resource. While SELFIE moved forward by agreeing that the URL-14's information content would document the URI-14 resource directly, there may be an argument to treat the URL-14's information content as an identifiable linked-data resource in its own right. An example that could drive further work investigation is the need to type the URL-14 resource itself. This would, effectively, make the URL-14 content part of the linked data graph. This would introduce a level of complexity and buy into semantic web concepts that may be too much for the current state of semantic web technology adoption.

To illustrate this, the outcomes of SELFIE dictate that a URL-14 resource is, at its core, an identified feature with a type as in:

```
{
  "@id": "https://id",
  "@type": "https://type_of_feature"
}
```

However, this content would have been retrieved from a URL such as: <https://info/id> and might just as well have content like:

```
{
  "@id": "https://info/id",
  "@type": "https://feature_info",
  "http://xmlns.com/foaf/0.1/primaryTopic": "https://id"
}
```

Whether the resource providing information about a non-information resource is itself a resource with a type, was deferred in SELFIE. The decision was made to move forward pragmatically—treating the URL-14 resource as a convenience resource whose only job is to provide information about a URI-14 and, therefore, not a resource that would ever be the subject of linked data. The details surrounding this point will undoubtedly need to be explored in more detail.

6.13.2. In band and out of band resources

As discussed in "[In band](#)" and "[out of band](#)" resources, the nature of a resource that is the object of a linked-data triple is important. What qualifies as an "in band" resource or a "semantic" resource is entirely in the eyes of the beholder. In the context of SELFIE, it was decided that a URL-14 resource that could provide an HTML representation with JSON-LD linked data in a `<script>` tag is considered "in band". Format and/or profile content negotiation, while of great interest, was deemed beyond the scope of what SELFIE could make firm recommendations about.

Extension of what is "in band" for environmental linked features should be the subject of future work. How content negotiation, either on format or profile, layers on top of 303 redirection is an open question. How links to resources that have both "in band" and "out of band" representations

should be handled is another. The key here is that future work will need to have implementation of the outcomes of SELFIE as hindsight to treat as evidence for future experiments.

6.13.3. Domain Ontologies and JSON-LD Contexts

A major outcome of the first ELFIE was noting the lack of domain ontologies upon which to build links between environmental features. Work to further this in SELFIE was largely unsuccessful. While additional "minimal ontologies" were generated, the procedures for publishing them as resolvable URIs proved too cumbersome to complete in the period of the IE. Additionally, foundational details of adapting these UML-based class models, and their numerous dependencies, to use in linked data remain unsolved and untested. While progress was made during the IE, the process is ongoing at the time of writing. Ontologies that correspond to a more detailed and adapted organization with respect to the packaging and naming pattern imposed by the UML domain model are currently under construction. Discussions were held with the OGC-NA to ensure that these ontologies follow a more generic controlled process and will provide easy and natural access to JSON-LD contexts from domain ontologies by negotiation.

Appendix A: Domain Use Cases

At the time of publication, these use cases were also available as web resources associated with the SELFIE project here: <https://opengeospatial.github.io/ELFIE/> [https://opengeospatial.github.io/ELFIE/]. The use cases documented in this report are only a sample of the use cases considered in SELFIE.

A.1. US Internet of Water Distributed Data and Observations

A.1.1. Use Case Description

The Internet of Water (IoW) is an initiative [proposed by a multi-stakeholder workshop](https://www.aspeninstitute.org/tag/internet-of-water/) [https://www.aspeninstitute.org/tag/internet-of-water/] that has been supported by a foundation and has a [non-profit organization](https://internetofwater.org/about/) [https://internetofwater.org/about/] by the same name working toward the initiative's goals. The initiative proposes a federated system of data providers and consumers that are aggregated into domain-specific "hub" organizations and coordinated by a single "umbrella" organization. Data producers are to maintain ownership and control of their information and federation is intended to be cache-only in that information ownership should never be transferred if the data producer wants to maintain ownership.

This use case applies to all types of water data—hydrography, hydrometric observations, hydrologic model results, water quality, hydrodynamics, etc. Ultimately, the hub and umbrella model requires construction of a linked-data system that references common environmental features throughout. Data providers' services should use consistent methods to reference those features to facilitate automated discovery of newly available or changed data.

[See this GitHub repository for more.](https://github.com/internetofwater/geoconnex.us) [https://github.com/internetofwater/geoconnex.us] [See the demo resources here for initial SELFIE content in this use case.](https://github.com/opengeospatial/SELFIE/tree/master/docs/usgs) [https://github.com/opengeospatial/SELFIE/tree/master/docs/usgs] [Use this JSON-LD demo page as an entry-point.](https://geoconnex.us/SELFIE/usgs/huc/huc12obs/070900020601) [https://geoconnex.us/SELFIE/usgs/huc/huc12obs/070900020601]

User Story

As a user of water data, I need to discover and access water information relevant to the environmental feature I care about from all the organizations that hold data about it, so I don't have to have special knowledge to access some information and so I don't miss some potentially relevant information.

Datasets and Sources

- USGS Reference Hydrography
- State and local data and observations
- University consortia aggregated data services
- Federal aggregated data and services

In the long run, this demonstration should have a very broad scope. Initially, it will focus on

building a catalog of hydrographic and observed data associated with hydrologic units.

A.1.2. Demo Description and Links

For this demonstration, the primary entry point is a [single hydrologic unit](https://geoconnex.us/SELFIE/usgs/huc/huc12obs/070900020601) [https://geoconnex.us/SELFIE/usgs/huc/huc12obs/070900020601] with three realizations: 1) a [hydrographic network](https://geoconnex.us/SELFIE/usgs/nhdplusflowline/huc12obs/070900020601) [https://geoconnex.us/SELFIE/usgs/nhdplusflowline/huc12obs/070900020601] of flowlines, 2) a [catchment divide](https://geoconnex.us/SELFIE/usgs/hucboundary/huc12obs/070900020601) [https://geoconnex.us/SELFIE/usgs/hucboundary/huc12obs/070900020601] containing a polygon representation, and 3) a [hydrometric network](https://geoconnex.us/SELFIE/usgs/hydrometricnetwork/huc12obs/070900020601) [https://geoconnex.us/SELFIE/usgs/hydrometricnetwork/huc12obs/070900020601] index of monitoring data.

These three realizations can be seen in the example below:

```
{
  "@context": [
    "https://opengeospatial.github.io/ELFIE/contexts/elfie-2/elf-index.jsonld",
    "https://opengeospatial.github.io/ELFIE/contexts/elfie-2/hy_features.jsonld"
  ],
  "@id": "https://geoconnex.us/SELFIE/usgs/huc/huc12obs/070900020601",
  "@type": "https://www.opengis.net/def/appschema/hy_features/hyf/HY_Catchment",
  "name": "Waunakee Marsh-Sixmile Creek",
  "description": "USGS Watershed Boundary Dataset Twelve Digit Hydrologic Unit Code Watershed",
  "catchmentRealization": [
    {
      "@id": "https://geoconnex.us/SELFIE/usgs/nhdplusflowline/huc12obs/070900020601",
      "@type": "https://www.opengis.net/def/appschema/hy_features/hyf/HY_HydrographicNetwork"
    },
    {
      "@id": "https://geoconnex.us/SELFIE/usgs/hucboundary/huc12obs/070900020601",
      "@type": "https://www.opengis.net/def/appschema/hy_features/hyf/HY_CatchmentDivide"
    },
    {
      "@id": "https://geoconnex.us/SELFIE/usgs/hydrometricnetwork/huc12obs/070900020601",
      "@type": "https://www.opengis.net/def/appschema/hy_features/hyf/HY_HydrometricNetwork"
    }
  ]
}
```

Links to Demo Resources

The demo was created using the R code found here: <https://github.com/opengeospatial/SELFIE/tree/master/tools/R>

While it will expand into other systems, the following resources have been contributed directly to the SELFIE space.

The top level Hydrologic Unit "HY\Catchment": <https://geoconnex.us/SELFIE/usgs/huc/huc12obs/070900020601>

The flowlines "HY_HydrographicNetwork": <https://geoconnex.us/SELFIE/usgs/nhdplusflowline/huc12obs/070900020601>

The boundary "HY\CatchmentDivide": <https://geoconnex.us/SELFIE/usgs/hucboundary/huc12obs/070900020601>

The monitoring sites "HY_HydrometricNetwork": <https://geoconnex.us/SELFIE/usgs/hydrometricnetwork/huc12obs/070900020601>

flowlines and boundary are intended to provide a visual representation and could also be used for geoprocessing workflows. Monitoring sites are a potentially long list of well-documented monitoring for the hydrologic unit. The state of "well-documented" in this use case is be a work in progress.

A.1.3. Demo findings and potential next steps

This demo demonstrates that the core SELFIE technical solution rooted in URI-14 → URL-14 redirection works well. Links to representations and associated features are operable but additional implementations will be needed to gain needed experience before any strong conclusions can be made. Availability of domain feature models (classes and associations) continues to be an issue the community needs to address before full-fledged implementation of domain-data-model linked data will be possible.

A.2. Groundwater Surface Water Interoperability Pilot - SELFIE Demonstration

A.2.1. Use Case Description

Link various heterogeneous datasets from various sources under a common water ontology using linked data across Canada and USA. The web application demonstrates how a linked data enabled application can use a predefined ontology to navigate across water related real world features potentially managed by various organizations within and across national jurisdictions.

User Story

I want to discover all the data related to water features (real world features relevant to surface water and groundwater) and how they are connected to other water features. From a map, I want to select a feature and be able to traverse to another water feature by following a link. I expect the application to understand key actionable properties of feature, such as the difference between a link to another water feature and a link to representations. I also expect the application to recognize useful representations and perform specific operation on them (such as geojson).
Datasets and Sources

- Watersheds delimitation of the Richelieu-Lake Champlain (NRCan-CCMEO for canadian portion and USGS for american portion)

- Aquifer systems description (NRCan-GSC)
- Water wells NRCan-GSC (GIN) and SIH (Système d'information hydrogéologique) Ministère de l'Environnement et de la lutte aux changements climatique du Québec.
- Stream gauges (ECCC, Water office, Meteorological Service of Canada and Centre d'expertise hydrique du Québec)
- Bedrock geology NRCan-GSC (based on various compilations)
- Cross-border USGS hydrologic units

Demo Description and Links

The demo is a web map application showing water related features. The map application operates on top of a linked data infrastructure (node) hosted on both side on the US-Canada border. Each Groundwater Surface water Interoperability Pilot (GSIP) node exposes a catalogue of water features from their respective jurisdiction, some cross-border (shared) features and establishes relations between water features on its side and features on the other side.

The GSIP resolver is built on top of an RDF catalog, containing water features description and links to other features and representation. GSIP deals with content negotiation and 303 redirect of NIR or other representation if necessary. This figure show the overall interaction with GSIP node

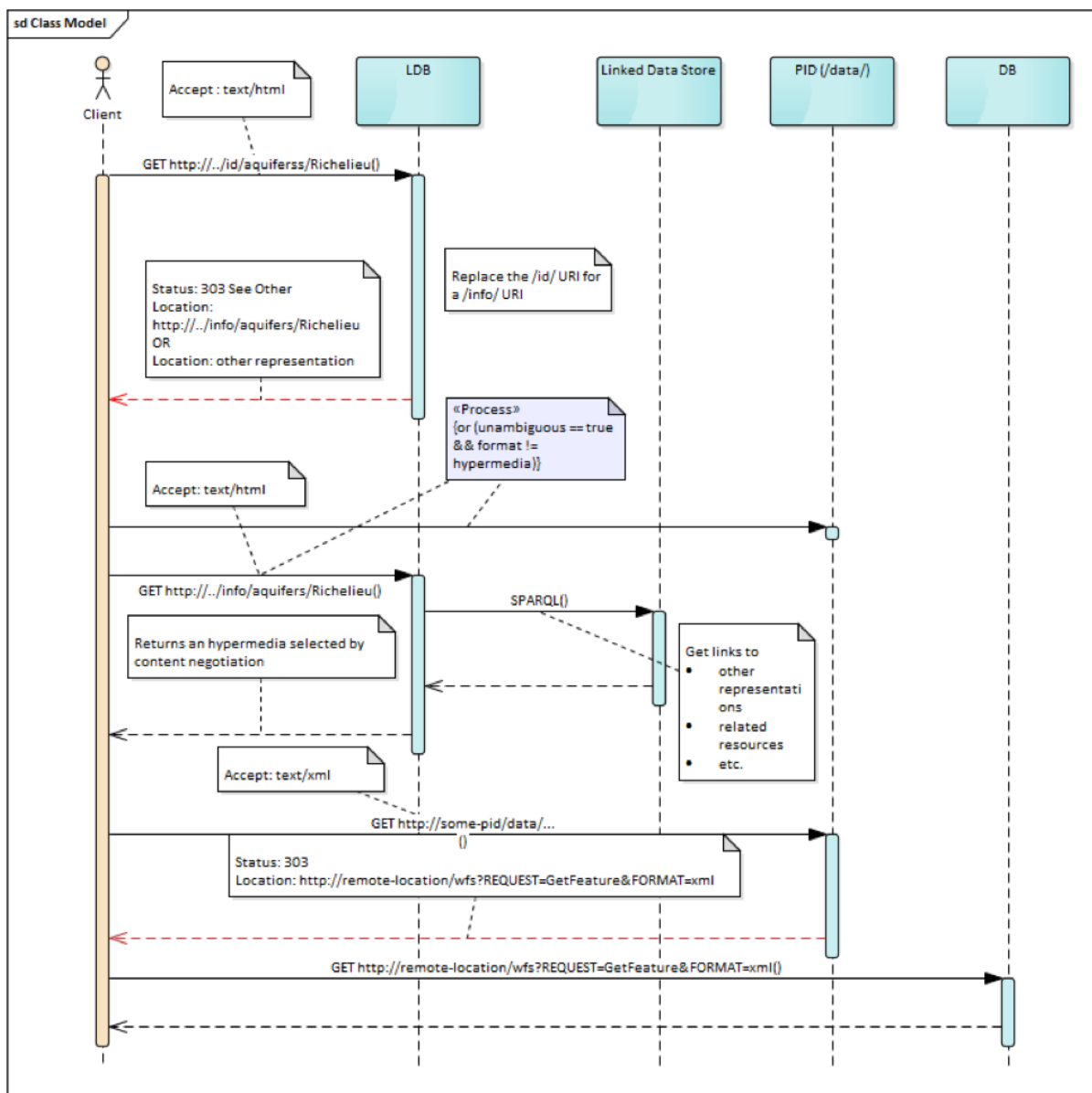


Figure 6. GSIP sequence diagram

The linked data infrastructure operates on its own and can be accessed using a regular browser. The map application leverages this infrastructure by adding new functionalities.

Features on the map are spatial representations of "real things" (Non-Information) in the world and assigned a URI as their identifier. At this point, this is all the web application has (feature with their NIR). The map is pre-loaded with a set of watersheds around Lake Champlain. Canadian version of the application shows feature north of the border and vice-versa for US. Note that NIR of water features can point anywhere (the Canadian version can consume data from the US node and vice and versa). When requested, the application attempts to resolve the feature's NIR by issuing a HTTP GET request using the NIR URI and process the document and do something useful with it. The application expects an RDF document conformant to the model (ontology) defined in this interoperability experiment. The application is robust as it will try to process whatever is returned by the resolution of the NIR. If the RDF document does not contain any `schema.org` or `RDFS` or `HY_Feature`, etc., it will simply do nothing.

The application then offers the option of:

- traverse a link to another water feature;

- open a representation in one of the proposed format;
- leave the map application and browse resources directly in the browser.

The application recognizes specifically GeoJSON representations and when available, the application can load it and add the content to the map. If the uploaded feature has a `uri` property, the application assumes it's a NIR and behave accordingly.

A typical session goes through a variation of these steps

1. User loads the map application
2. User clicks on a feature
3. Application displays information about the resources in an info bubble.
4. User can
5. click on a link (a predicate) and traverse to another resources
 - a. click on HTML icon and pop a new browser page/tab loaded with a landing page of the feature. At the point, the user “left” the application (although it remains available in the original tab if the user want to return).
 - b. Click on the GeoJSON “push” pin and load the feature on the map
6. repeat from step #2

Harvesting

While not explicitly demonstrated in the web application, the architecture relies on a series of nodes to resolve NIR. The current demonstration relies on 2 nodes (a.k.a `geoconnex.ca` and `geoconnex.us`) but it is totally agnostic of the number of nodes that can eventually be use while traversing from one water feature to another. Nodes are autonomous and are not “aware” that a statement refers to a resource managed by another “GSIP” node (they are just NIR that will be resolved by the client). But the node might be interested in those statements, especially when it can create a reciprocate statement (if A in US is upstream of B in Canada, B is downstream of A). When such a statement is added to either node, the node of interest needs to be updated to reflect that change. The GSIP architecture includes a harvester that probe known nodes (the Canadian harvester knows the location of the US node) and extract relevant cross border (cross node) statements and update its copy of the catalog. The harvesting is done periodically.

A.2.2. Links to Demo Resources

The demo instances are located at <https://geoconnex.ca/gsip/app/index.html> and <https://info.geoconnex.us/chyld-pilot/app/index.html> . The web application is a map where water related features are shown. The map has watershed delimitation permanently displayed. Other features are displayed if a GeoJSON representation is available and the user requests it.

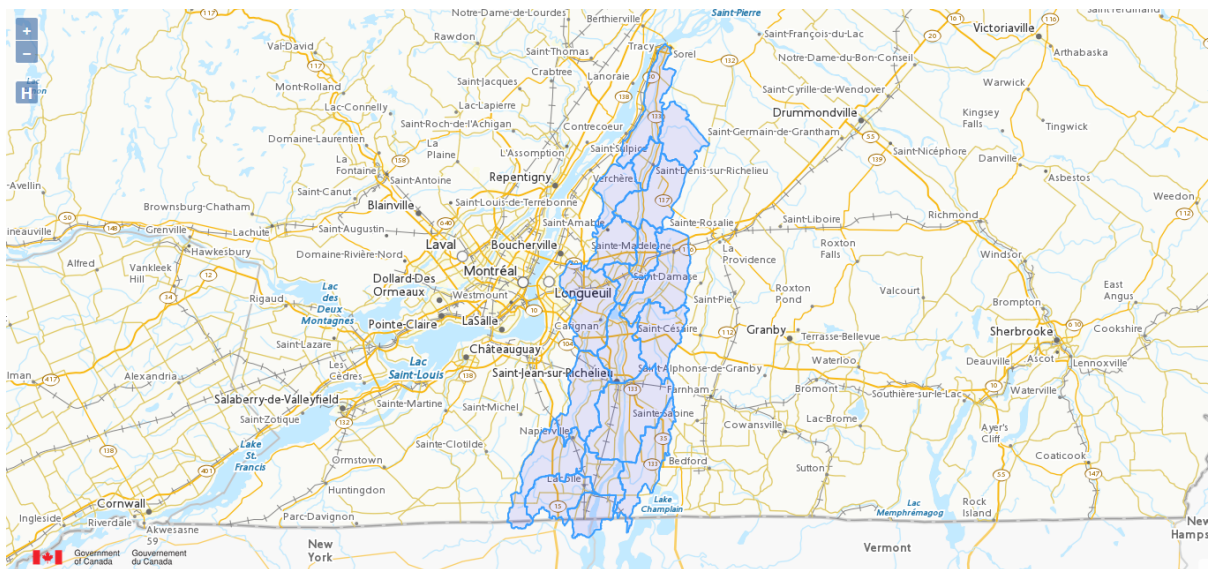


Figure 7. GSIP web application

Clicking on a feature displays an information bubble containing names and linkages to other resources (a.k.a, in band). The information bubble is built from the MIR received from GSIP. Each resource has a link to its landing page (HTML icon). Clicking on the link loads the page in a different tab, outside the web application. Interacting with the landing page is done outside the context of the map application and behave similarly to other demos in this report.

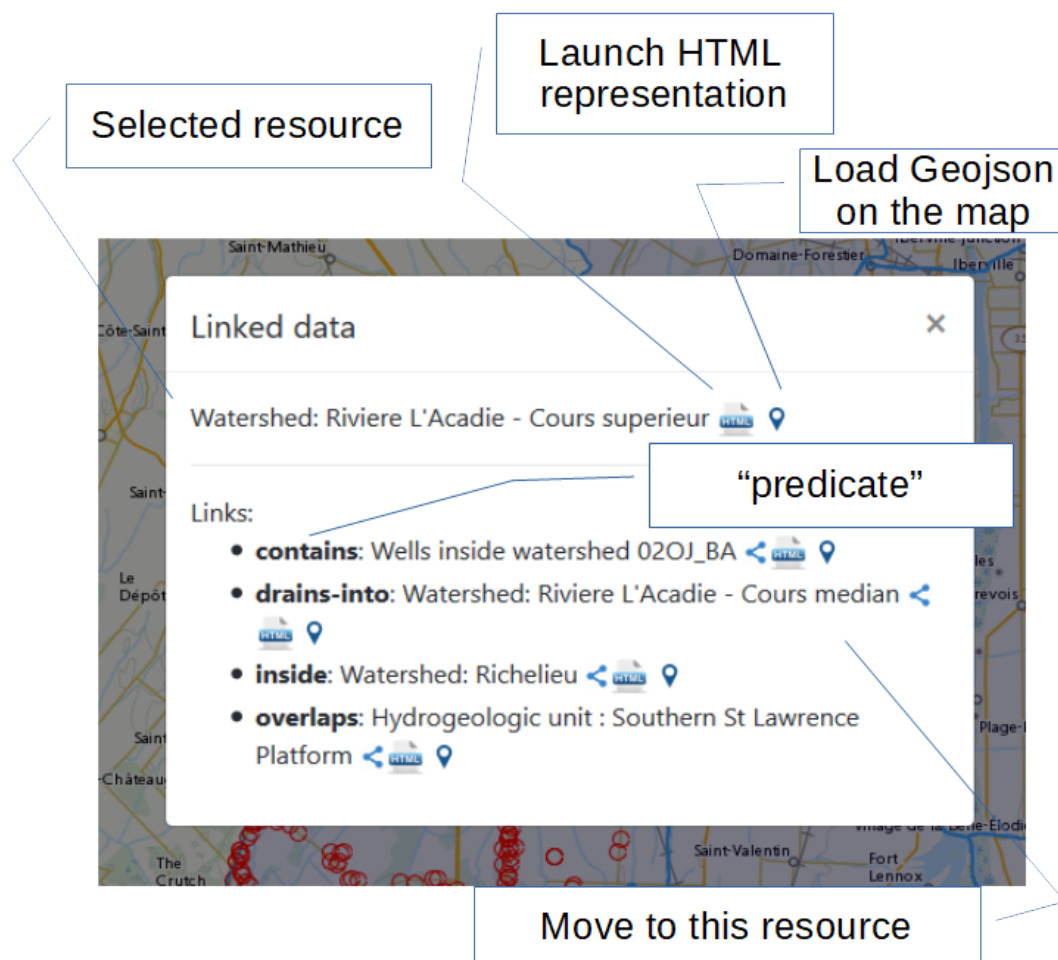


Figure 8. GSIP information bubble

Resources having a GeoJSON representation show a “pushpin” icon. Clicking the pushpin loads the

GeoJSON representation of that resource into the map in red [figure 3]. The loaded feature can also be clicked and if it has a “uri” property, the application will try to dereference it. If it succeeds, an information bubble is displayed.

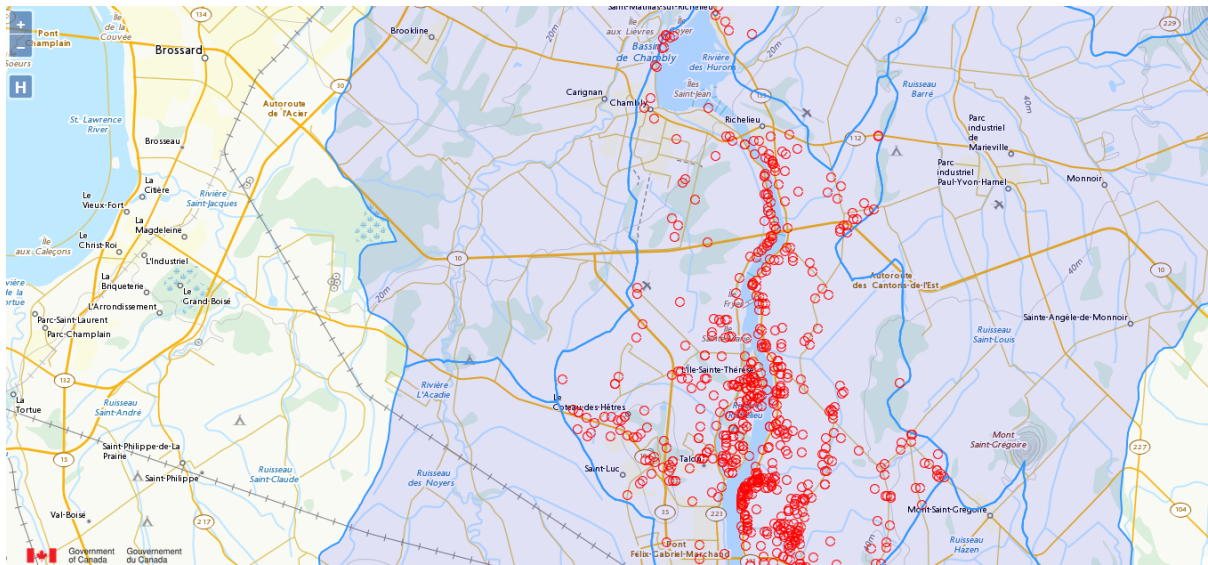


Figure 9. GeoJSON loaded on the map

A typical NIR URI : https://geoconnex.ca/gsip/id/catchment/02OJ*BA will be redirected to MIR https://geoconnex.ca/gsip/info/catchment/02OJ*BA containing, among other things, the information used to populate the information bubble.

The key elements of interest that is reflected in the information bubble are presented here in RDF TTL (a full MIR is available in annex. Note that statements were manually reorganized here to illustrate the key statements).


```

@prefix schema: <http://schema.org/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix hy: <http://geosciences.ca/def/hydraulic#>.

# "in band" data = actionable information
<https://geoconnex.ca/id/catchment/020J*BA>
  a hy:HY_Catchment> , <http://www.w3.org/2002/07/owl#Thing> , rdfs:Resource ;
  rdfs:label "Watershed: Riviere L'Acadie - Cours superieur"@en , "Bassin
versant : Riviere L'Acadie - Cours superieur"@fr ;
  hy:contains
    <https://geoconnex.ca/id/featureCollection/wellsIn020J_BA> ;
  hy:drains-into
    <https://geoconnex.ca/id/catchment/020J*BB> ;
  hy:inside
    <https://geoconnex.ca/id/catchment/020J> ;
  hy:overlaps
    <https://geoconnex.ca/id/hydrogeounits/Richelieu1> ;
  schema:name "Watershed : Riviere L'Acadie - Cours superieur" ,
"020J*BA".

# links to other representations (one 1 here), either out of band or in band,
depending of dct:conformsTo value. The following example does not announce any
conformance and therefore is "out-of-band" by default.

<https://geoconnex.ca/id/catchment/020J*BA> schema:subjectOf
<https://geoconnex.ca/data/catchment/HYF/WSCSSSDA/NRCAN/020J*BA>.

# description of that representation.
<https://geoconnex.ca/data/catchment/HYF/WSCSSSDA/NRCAN/020J*BA>
  dct:format "application/vnd.geo+json" , "text/html" ;
  schema:provider <http://gin.gw-info.net> .

```

For example, a US watershed might state that it is upstream of a Canadian watershed.

eg:

```

<https://geoconnex.us/chyld-pilot/id/hu/041504081604>
  <https://www.opengis.net/def/hy_features/ontology/hyf/lowerCatchment>
    <https://geoconnex.ca/id/catchment/020J*CA> .

```

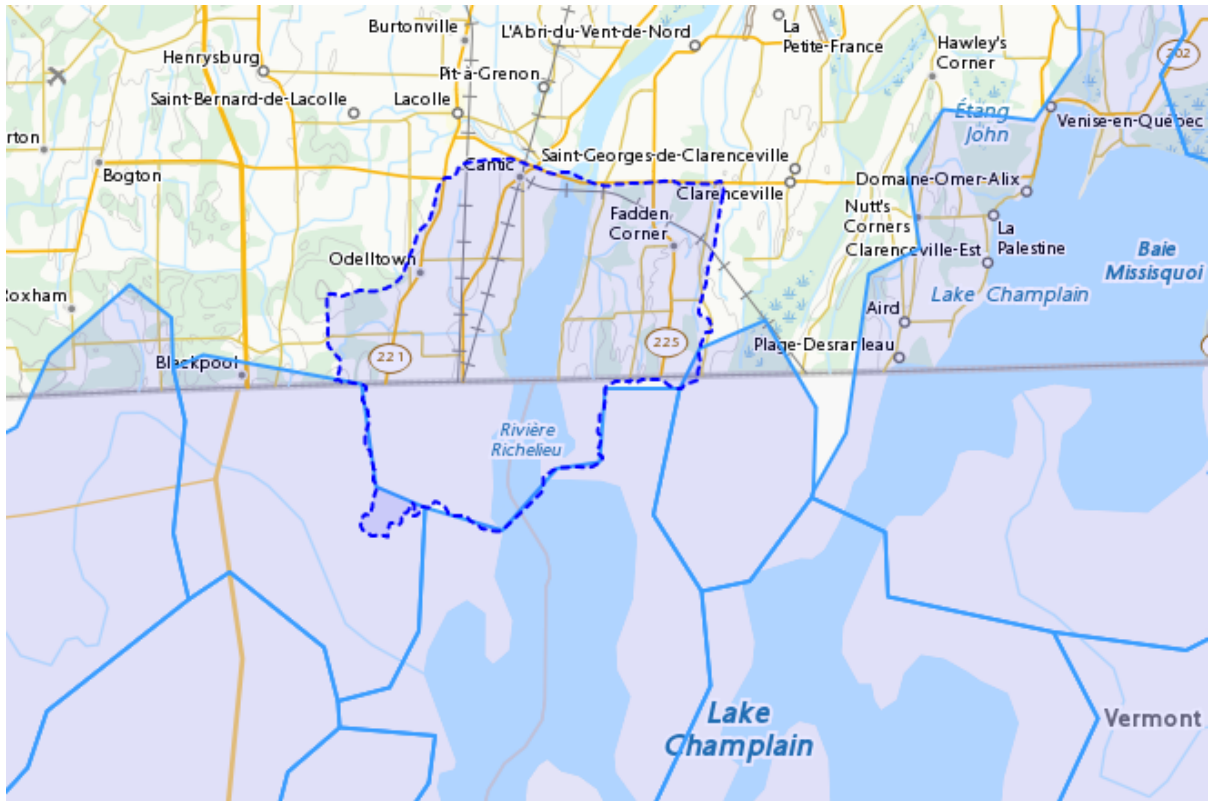


Figure 10. Canadian watershed pulled from Canadian node shown on US instance

A.2.3. Demo findings and potential next steps

This demo explores the possibility of developing software on top of linked data infrastructure. It is slightly different from the common web heavy demonstration of search engines + web browser + html. The web application performs operations that are not usually done by browsers alone (such as manipulation spatial data and display them). Another demo implementation has been created in a non-browser environment (QGIS), with the same capabilities

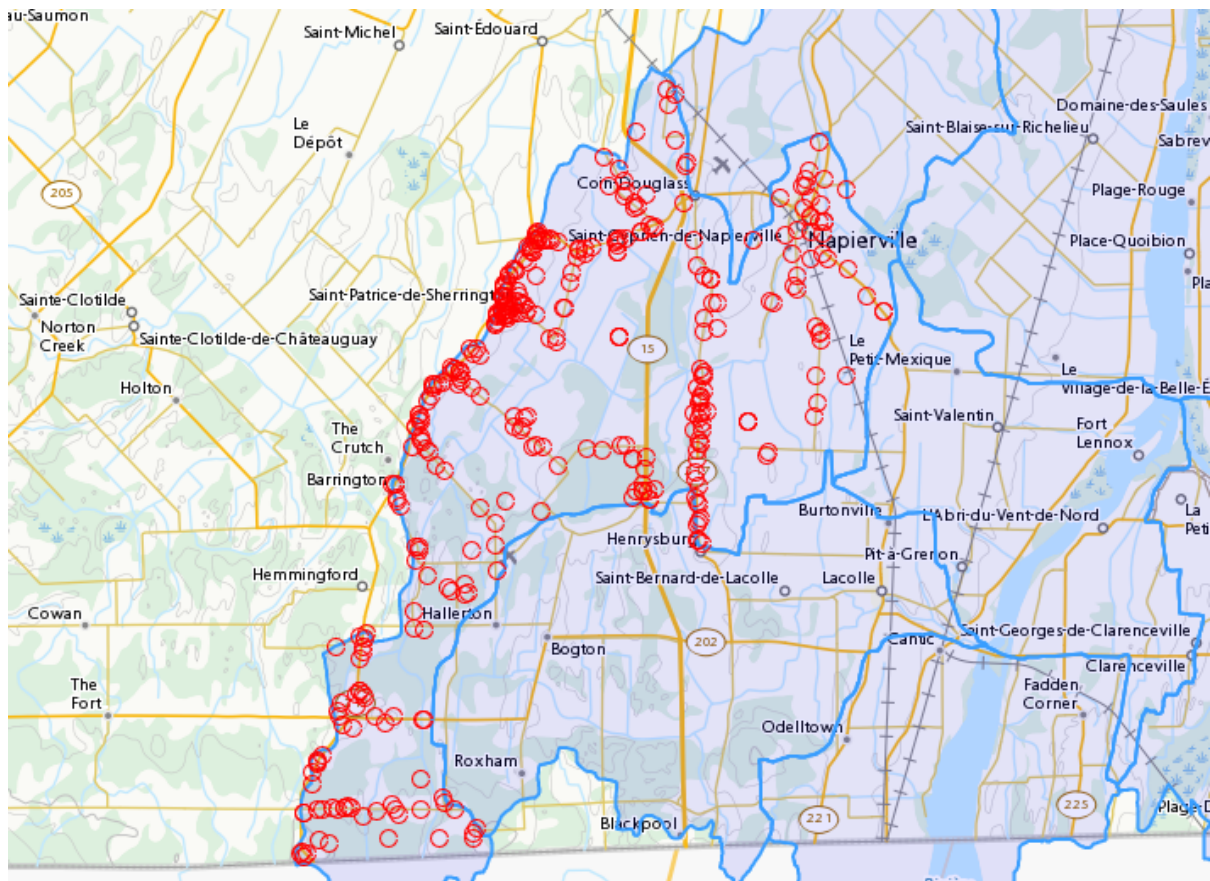


Figure 11. Same demo implemented in QGIS

Because the application has prior knowledge of the model, or more accurately put, is able to recognize some specialized content, it can act upon it. We could envision more specific application, such as a tool that is HY_Feature aware rebuilding complete watershed from one point by traversing an “upperCatchment” predicate or a GWML aware application locating a recharge area from of an aquifer, or a GeoSciML aware application loading all datasets that are relevant to stratigraphic columns (as an ‘aspatial’ exemple). Because the landing page can mix predicates and classes from many ontology, many application can be built on over the same linked data infrastructure.

A.2.4. Annex

```

@prefix schema: <http://schema.org/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<https://geoconnex.ca/data/catchment/HYF/WSCSSDA/NRCAN/020J*BA>
  dct:conformsTo <https://www.opengis.net/def/gwml2> ;
  dct:format "application/vnd.geo+json" , "text/html" ;
  schema:provider <http://gin.gw-info.net> .

rdfs:Resource a rdfs:Class , <http://www.w3.org/2002/07/owl#Class> ,
rdfs:Resource ;
  rdfs:subClassOf rdfs:Resource ;
  <http://www.w3.org/2002/07/owl#equivalentClass>
    rdfs:Resource .

```

```

<https://geoconnex.ca/id/hydrogeounits/Richelieu1>
  a
    <http://geosciences.ca/def/groundwater#GW_HydrogeoUnit> ,
<http://www.w3.org/2002/07/owl#Thing> , rdfs:Resource ;
  rdfs:label
    "Unite hydrogeologique : Plate-forme du St-Laurent sud"@fr
, "Hydrogeologic unit : Southern St Lawrence Platform"@en ;
  <http://geosciences.ca/def/groundwater#gwAquiferSystem>
    <https://geoconnex.ca/id/aquiferSystems/Richelieu> ;
  hy:contains>
    <https://geoconnex.ca/id/swmonitoring/WSC_020J026> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53537> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53515> ,
<https://geoconnex.ca/id/swmonitoring/MDDELCC_030430> ,
<https://geoconnex.ca/id/swmonitoring/MDDELCC_030421> ,
<https://geoconnex.ca/id/swmonitoring/WSC_020J024> ,
<https://geoconnex.ca/id/swmonitoring/WSC_020J007> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53529> ,
<https://geoconnex.ca/id/swmonitoring/WSC_020J016> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53545> ,
<https://geoconnex.ca/id/swmonitoring/MDDELCC_030415> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53517> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53544> ;
  hy:measuredBy>
    <https://geoconnex.ca/id/gwmonitoring/prj_27.53515> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53537> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53517> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53544> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53545> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53529> ;
  hy:overlaps>
    <https://geoconnex.ca/id/catchment/020J*AB> ,
<https://geoconnex.ca/id/geologicUnits/008000/GSCC00053008397> ,
<https://geoconnex.ca/id/catchment/020J*DB> ,
<https://geoconnex.ca/id/catchment/020J*DA> ,
<https://geoconnex.ca/id/geologicUnits/006000/GSCC00053006880> ,
<https://geoconnex.ca/id/geologicUnits/014000/GSCC00053014607> ,
<https://geoconnex.ca/id/catchment/020J*CA> ,
<https://geoconnex.ca/id/geologicUnits/011000/GSCC00053011490> ,
<https://geoconnex.ca/id/catchment/020J*CC> ,
<https://geoconnex.ca/id/catchment/020J*DC> ,
<https://geoconnex.ca/id/geologicUnits/010000/GSCC00053010067> ,
<https://geoconnex.ca/id/geologicUnits/010000/GSCC00053010658> ,
<https://geoconnex.ca/id/catchment/020J*CB> ,
<https://geoconnex.ca/id/geologicUnits/003000/GSCC00053015117> ,
<https://geoconnex.ca/id/catchment/020J*BB> ,
<https://geoconnex.ca/id/catchment/020J*BA> ,
<https://geoconnex.ca/id/geologicUnits/017000/GSCC00053017020> ,
<https://geoconnex.ca/id/catchment/020J*CE> ,
<https://geoconnex.ca/id/catchment/020J*CD> ,
<https://geoconnex.ca/id/geologicUnits/010000/GSCC00053010757> ,
<https://geoconnex.ca/id/catchment/020J*AA> ,

```

```

<https://geoconnex.ca/id/geologicUnits/008000/GSCC00053008293> ,
<https://geoconnex.ca/id/geologicUnits/015000/GSCC00053015716> ,
<https://geoconnex.ca/id/geologicUnits/008000/GSCC00053008833> ,
<https://geoconnex.ca/id/geologicUnits/001000/GSCC00053001039> ,
<https://geoconnex.ca/id/geologicUnits/012000/GSCC00053012027> ,
<https://geoconnex.ca/id/geologicUnits/000000/GSCC00053000990> ,
<https://geoconnex.ca/id/catchment/020J*BC> ,
<https://geoconnex.ca/id/catchment/020J*CF> ;
    schema:description "\r\nIn the context of the southern area of the St.
Lawrence Platform of (south Lowlands), the clay unit is generally not continuous or
thick. The bedrock is rather covered by a till unit of at least 10 m thick which may
allow significant bedrock aquifer recharge rates. This limited sedimentary cover
suggests that there would be links between the bedrock aquifer and streams,
particularly along some sections of the Richelieu River, which constitute discharge
areas. The flow is oriented east-west, from the recharge areas to Richelieu River or
others discharge areas. The surficial permeable sediments with significant thickness
have small spatial extension, thus that the aquifer potential is mainly based on
fractured bedrock aquifer. In the unit, there is a significant use of groundwater as
water supply. The predominant semi-confined conditions involve a moderate
vulnerability of the bedrock aquifer. Groundwater exceeds frequently some aesthetic
criteria as Fe, Mn, S, Na, and F in the central area of the hydrogeological unit.\r\n"
;
    schema:image "http://gin.gw-info.net/service/ngwds//en/wms/ngwd-
wms/inset?REQUEST=GetMap&SERVICE=WMS&VERSION=1.1.1&LAYERS=area&STYLES=&FORMAT=image/pn
g&BGCOLOR=0xFFFFF&TRANSPARENT=TRUE&SRS=EPSG:4326&BBOX=-
73.6883387829505,44.9741147159004,-
72.8050177950318,45.6366054568393&WIDTH=400&HEIGHT=300&TABLE=gw_data.hydrogeological_u
nits&FIELD=id&ID=1" ;
    schema:name "Hydrogeologic unit : Southern St Lawrence Platform" ;
    <http://www.opengeospatial.org/standards/geosparql/sfIntersects>
    <https://geoconnex.us/chyld-pilot/id/hu/041504081507-drainage_basin> ,
<https://geoconnex.us/chyld-pilot/id/hu/041504081102-drainage_basin> ,
<https://geoconnex.us/chyld-pilot/id/hu/041504081007-drainage_basin> ,
<https://geoconnex.us/chyld-pilot/id/hu/041504081006-drainage_basin> ,
<https://geoconnex.us/chyld-pilot/id/hu/041504081005> , <https://geoconnex.us/chyld-
pilot/id/hu/041504081507> , <https://geoconnex.us/chyld-pilot/id/hu/041504081505-
drainage_basin> , <https://geoconnex.us/chyld-pilot/id/hu/041504081203> ,
<https://geoconnex.us/chyld-pilot/id/hu/041504081006> , <https://geoconnex.us/chyld-
pilot/id/hu/041504081007> , <https://geoconnex.us/chyld-pilot/id/hu/041504081005-
drainage_basin> , <https://geoconnex.us/chyld-pilot/id/hu/041504081505> ,
<https://geoconnex.us/chyld-pilot/id/hu/041504081203-drainage_basin> ;
    <http://www.w3.org/2002/07/owl#sameAs>
    <https://geoconnex.ca/id/hydrogeounits/Richelieu1> .

<https://geoconnex.ca/id/catchment/020J>
    a <http://www.w3.org/2002/07/owl#Thing> , hy:HY_Catchment> ,
rdfs:Resource ;
    rdfs:label "Watershed: Richelieu"@en , "Bassin versant: Richelieu"@fr ;
    hy:contains>
    <https://geoconnex.ca/id/gwmonitoring/prj_27.53523> ,
<https://geoconnex.ca/id/catchment/020J*DD> ,

```

<https://geoconnex.ca/id/featureCollection/wellsIn020J_DC> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_CC> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_CE> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53541> ,
<https://geoconnex.ca/id/swmonitoring/WSC_020J026> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_DH> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53517> ,
<https://geoconnex.ca/id/catchment/020J*DH> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53544> ,
<https://geoconnex.ca/id/catchment/020J*CC> ,
<https://geoconnex.ca/id/catchment/020J*BC> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_BB> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_AB> ,
<https://geoconnex.ca/id/swmonitoring/MDDELCC_030421> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53510> ,
<https://geoconnex.ca/id/catchment/020J*DA> ,
<https://geoconnex.ca/id/catchment/020J*DC> ,
<https://geoconnex.ca/id/swmonitoring/WSC_020J007> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_DA> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_DB> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_CB> ,
<https://geoconnex.ca/id/catchment/020J*AB> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53515> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_DE> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_DG> ,
<https://geoconnex.ca/id/catchment/020J*DG> ,
<https://geoconnex.ca/id/catchment/020J*CB> ,
<https://geoconnex.ca/id/catchment/020J*BB> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53545> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_BA> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_AA> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53529> ,
<https://geoconnex.ca/id/swmonitoring/MDDELCC_030430> ,
<https://geoconnex.ca/id/catchment/020J*DB> ,
<https://geoconnex.ca/id/swmonitoring/WSC_020J016> ,
<https://geoconnex.ca/id/catchment/020J*CF> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_CA> ,
<https://geoconnex.ca/id/catchment/020J*AA> ,
<https://geoconnex.ca/id/catchment/020J*DE> ,
<https://geoconnex.ca/id/swmonitoring/WSC_020J024> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53632> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_DD> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_CD> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_DF> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_CF> ,
<https://geoconnex.ca/id/catchment/020J*DF> ,
<https://geoconnex.ca/id/catchment/020J*CA> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53542> ,
<https://geoconnex.ca/id/catchment/020J*BA> ,
<https://geoconnex.ca/id/swmonitoring/MDDELCC_030429> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53518> ,

```

<https://geoconnex.ca/id/catchment/020J*CD> ,
<https://geoconnex.ca/id/gwmonitoring/prj_27.53537> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_BC> ,
<https://geoconnex.ca/id/swmonitoring/MDELCC_030415> ,
<https://geoconnex.ca/id/catchment/020J*CE> ;
    schema:name "Watershed : Richelieu" ;
    <http://www.w3.org/2002/07/owl#sameAs>
        <https://geoconnex.ca/id/catchment/020J> .

<https://geoconnex.ca/id/catchment/020J*BA>
    a          hy:HY_Catchment> , <http://www.w3.org/2002/07/owl#Thing> ,
rdfs:Resource ;
    rdfs:label "Watershed: Riviere L'Acadie - Cours superieur"@en , "Bassin
versant : Riviere L'Acadie - Cours superieur"@fr ;
    hy:contains>
        <https://geoconnex.ca/id/featureCollection/wellsIn020J_BA> ;
    hy:drains-into>
        <https://geoconnex.ca/id/catchment/020J*BB> ;
    hy:inside>
        <https://geoconnex.ca/id/catchment/020J> ;
    hy:overlaps>
        <https://geoconnex.ca/id/hydrogeounits/Richelieu1> ;
    schema:name "Watershed : Riviere L'Acadie - Cours superieur" , "020J*BA"
;
    schema:subjectOf
<https://geoconnex.ca/data/catchment/HYF/WSCSSSDA/NRCAN/020J*BA> ;
    <http://www.w3.org/2002/07/owl#sameAs>
        <https://geoconnex.ca/id/catchment/020J*BA> .

<https://geoconnex.ca/id/featureCollection/wellsIn020J_BA>
    a          <http://www.w3.org/2002/07/owl#Thing> , rdfs:Resource ;
    rdfs:label "Wells inside watershed 020J_BA"@en , "Puits a l'interieur du
bassin 020J_BA"@fr ;
    hy:inside>
        <https://geoconnex.ca/id/catchment/020J> ,
<https://geoconnex.ca/id/catchment/020J*BA> ;
    <http://www.w3.org/2002/07/owl#sameAs>
        <https://geoconnex.ca/id/featureCollection/wellsIn020J_BA> .

hy:HY_Catchment>
    a          rdfs:Class , <http://www.w3.org/2002/07/owl#Class> ,
rdfs:Resource ;
    rdfs:label "Bassin de drainage"@fr , "Catchment"@en ;
    rdfs:subClassOf hy:HY_Catchment> , <http://www.w3.org/2002/07/owl#Thing> ,
rdfs:Resource ;
    <http://www.w3.org/2002/07/owl#equivalentClass>
        hy:HY_Catchment> .

<https://geoconnex.ca/id/catchment/020J*BB>
    a          <http://www.w3.org/2002/07/owl#Thing> , rdfs:Resource ,
hy:HY_Catchment> ;

```

```

    rdfs:label    "Bassin versant: Riviere L'Acadie - Cours median"@fr ,
"Watershed: Riviere L'Acadie - Cours median"@en ;
    hy:contains>
        <https://geoconnex.ca/id/gwmonitoring/prj_27.53537> ,
<https://geoconnex.ca/id/featureCollection/wellsIn020J_BB> ;
    hy:drains>
        <https://geoconnex.ca/id/catchment/020J*BA> ;
    hy:drains-into>
        <https://geoconnex.ca/id/catchment/020J*BC> ;
    hy:inside>
        <https://geoconnex.ca/id/catchment/020J> ;
    hy:overlaps>
        <https://geoconnex.ca/id/hydrogeounits/Richelieu1> ;
    schema:name  "Watershed : Riviere L'Acadie - Cours median" ;
    <http://www.w3.org/2002/07/owl#sameAs>
        <https://geoconnex.ca/id/catchment/020J*BB> .

<http://www.w3.org/2002/07/owl#Thing>
    a          rdfs:Class , <http://www.w3.org/2002/07/owl#Class> ,
rdfs:Resource ;
    rdfs:subClassOf <http://www.w3.org/2002/07/owl#Thing> , rdfs:Resource ;
    <http://www.w3.org/2002/07/owl#equivalentClass>
        <http://www.w3.org/2002/07/owl#Thing> .

```

A.3. BRGM - INSIDE research center - Surface / Ground water linked data gazetteer

A.3.1. Use Case Description

This use case builds on the one set up for the OGC ELFIE and tries to test a system coping with two specific requirements:

- A. clients that dereference a URI asking for a specific media-type, content model etc.
- B. clients that dereference the same URI without knowing beforehand the available media-type, content model etc. in a kind of "probing" or "discovery" behavior.

User Story

Clients from Group A can be very diverse

- linked data centered application : in this case BLiV (BRGM Linked data Viewer) is considered. BLiV (<https://data.geoscience.fr/Bliv/>) is developed to ingest
 - natively, linked data serialization
 - if not available, ask the end-user (human being), if he wants to interact with the other representations/serializations available
- desktop GIS : QGIS with GML Application Schema Toolbox which expects responses that are GML application schema compliant (e.g. <https://plugins.qgis.org/plugins/>)

[gml_application_schema_toolbox/](#))

- search engines crawlers : which expect HTML with JSON-LD in the `<script>` header

Clients from Group B do not correspond to the vast majority of the clients considered in Linked Data oriented approaches because, most of the time, specific implementation environments are considered (e.g. linking Non-Information Resources together and somehow expecting a specific serialization content-model paradigm, web interfaces, reasoning).

However, this need should not be overlooked as in running production environments, especially when linking information about Non-Information Resources together with another system, (linked) data manager and their associated system(s) need to know what's is available 'behind' a URI. At least to cross-check they are linking to the relevant resource. Just knowing the URI-14 is not always sufficient for disambiguation.

The client dereferences a URI without specifying any media-type, content model, etc. and retrieves what combination of information is available.

From this

- Client from Group B: checks that they are linking to the relevant resource.
- Client from Group A: assesses whether they can consider the data content available provided in-band or out-of-band and interact with what suits them the most (or not).

Clients from Group A follow a 1 step approach from URI-14 to the data content skipping the URL-14 landing content whereas clients from group B may apply the complete URI-14 → URL-14 → data content pattern.

Datasets & APIs

- BRGM BD LISA : French Aquifer dataset provided using international semantic and technical interoperability approaches. BRGM already provides an OGC API Feature implementation on top of BD LISA that exposes the content according to OGC:GWML2 in GML, GeoJSON, JSON-LD with a URI resolver on top.
- SANDRE Aquifer reference dataset : The French Aquifer dataset is provided according to French Water Information System conceptual model (semantics), and interchange format(XML serialization) etc... Note : in both cases the source dataset (instances) is the same.

A.3.2. Demo Description and Links

The use case corresponds somehow to set up what could be called a linked-data-gazetteer along with the corresponding URI configuration

In order to fulfill it, the following steps have been carried out

- define a model that allows to know in which media-type, content model etc. a given instance is available. This mode is a level of abstraction higher than SELFIE content model (<https://data.geoscience.fr/def/LinkedDataGazetteer.xsd>)
- populate it for the French Aquifer dataset

- implement that model on an OGC API – Features interface using GeoServer
- fund and drive the necessary evolutions so that GeoServer’s OGC API - Features implementation is capable of serving JSON-LD content and ultimately deploy within BRGM infrastructure.

The API endpoint is the following: <https://data.geoscience.fr/api/LDGazetteerFAPI> and allows to search within the French Aquifer dataset

Demo Screenshot(s)

Group B use case leads to dereferencing the same URI asking for nothing (ex: using cURL) and loading URI-14 landing content into BLIV

Then traversing "in-band" to fetch the Aquifer description in BLIV (ex : JSON-LD) or move to other content model, media-types. (GWML2 in GML, SANDRE in XML) and use it in another application.

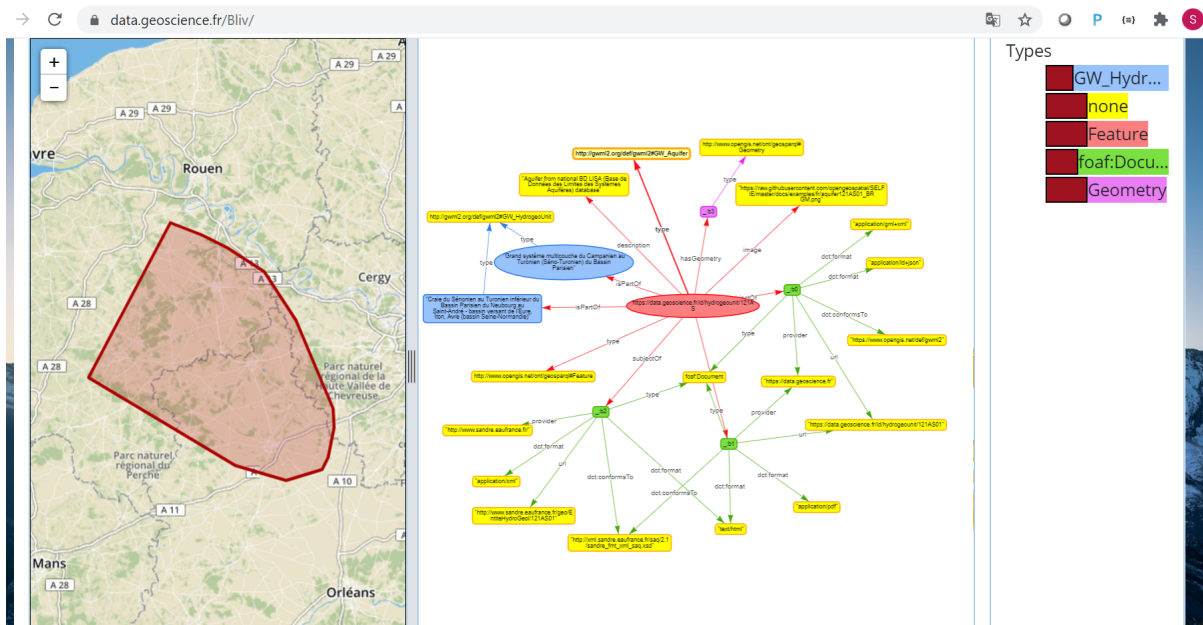


Figure 12. French Hydrogeologic unit 121AS01 landing data in BLiV

Group A use case leads to dereferencing the same URI asking for

- `application/ld+json`: that Aquifer in BLiV in JSON-LD (see ELFIE demos on <https://opengeospatial.github.io/ELFIE/>)
- `application/gml+xml`: that Aquifer in QGIS with GML Application Schema Toolbox (see demos under https://github.com/BRGM/gml_application_schema_toolbox/tree/master/presentations)
- `text/html` (or variations around this): resolver logs clearly show that Google Bots (or Bing or other crawlers) ask for text/html (or variations around this) ask for and receive the corresponding content. The plan is then to refine the embeded JSON-LD content in the `<script>` header of this representation.

A.3.3. Demo findings and potential next steps

The implementation of the linked-data-gazetteer serving SELFIE content model compliant payload clearly demonstrates

- that SELFIE proposed content model fills a gap in linked data systems enabling description of various representations available for a given URI-14
- the importance of the in-band / out-of-band distinction to better understand what client actually expects. The introduction of that terminology within SELFIE members discussions clearly helped move forward towards a common understanding
- the evolution of how content negotiation is being considered (ex: adding notion of profiles), will change how linked data systems are designed and implemented (on both client & server side). When this evolution becomes mainstream in the implementations, it will become more natural for (linked) data managers to link resource together using a URI-14.

Several potential next steps have been identified:

- Hypermedia-driven resource resolution versus content negotiation-driven resource resolution needs to be further discussed as both approaches have their pros & cons. An issue is created to compare SELFIE content model with W3C DXWG ConnegP. A quick cross-check seems to identify the following elements (but more in depth work is required)
 - connegP 'type' seems to map to SELFIE choice of 'format'
 - connegP 'profile' seems to map to SELFIE choice of 'conformsTo'
 - but there is no notion of "primaryTopic" and no "provider" in connegP
- the 'dc:partOf' in SELFIE content model could also be a specific view/profile (a data content) and not embedded in the URL-14 landing content. It seems to us that we are mixing two different use cases : "URI-14 to data content probing and discovery" VS "exploring a domain graph of linked feature". We may consider serving this domain graph in data content later on.
- what response should be provided to a client trying to dereference a URI-14 and asking for specific media-type, content model etc. that cannot be provided? It would make sense to respond back with what is actually known to be available using SELFIE content model but have not been implemented in our system.
- in its current version (1.2.0), QGIS GML Application Schema Toolbox plugin "identifies" itself when dereferencing a URI (default behavior is to have the HTTP Use-Agent named "QGIS GML Application Schema Toolbox"). This allows to know at the resolver level that the expected media-type to be used is `application/gml+xml`. This mechanism, should be revised and it is already planned to have this QGIS plugin evolve to a better implementation of content negotiation, explicitly stating what the client wants to retrieve

A.4. CeRDI VVG - SELFIE Demonstration

A.4.1. Use Case Description

To provide a means whereby all the relevant information (resources) about a real-world feature (in this case a borehole or well) can be brought together via machine readable (and indexable) web available information.

CeRDI SELFIE testbed [<https://geo.org.au/info/well/46217>]

User Story

Groundwater borehole data management in Victoria is split across a number of Government departments, research agencies and community groups. Information about the same real-world entity borehole may exist in multiple databases. The VVG web portal partly addressed this problem by federating these disparate data services into a spatial web portal that allows the user to access ALL the information regardless of the source or duplication. [History of Bore data in Victoria](https://www.vvg.org.au/cb_pages/history_of_bore_data.php) [https://www.vvg.org.au/cb_pages/history_of_bore_data.php]

Information about a borehole exists at one or more of:

- Geological Survey of Victoria (GSV)
- Department of Environment, Land, Water and Planning (DELWP)
- State Library of Victoria (SLV)
- Federation University Australia (FedUni)

These are services that deliver one or more of:

- HTML
- GML
- JSON
- Documents / multimedia

The data from these services may be about:

- Geology / Aquifers
- Groundwater (water quality, levels)
- Borehole construction
- Reports
- Observations made on things intersected in the bore

Currently a person or automated client must individually discover and access these different data services and compile the relevant information about a Borehole manually. Where the same borehole exists across multiple data sources it is not readily apparent that they are the same real-world feature (there is no common identifier across these services). Additionally, there is no mechanism to identify the different types of information available.

Through this demonstration, a user should be able to use a standard search engine to discover the availability of these various sources, formats and contents via URL-14 landing content. The user (including machines) can navigate via the links in the landing content to request data from the various providers in one of the available formats.

Datasets

- Geological Survey of Victoria (GSV) - [Geological Exploration and Development Information System \(GEDIS\)](https://www.vvg.org.au/cb_pages/gedis.php) [https://www.vvg.org.au/cb_pages/gedis.php]

- Department of Environment, Land, Water and Planning (DELWP) - [Water Measurement Information System \(WMIS\)](https://www.vvg.org.au/cb_pages/wmis.php) [https://www.vvg.org.au/cb_pages/wmis.php]
- Federation University Australia - [Groundwater research bores](https://www.vvg.org.au/cb_pages/ub_spatial_bore_database.php) [https://www.vvg.org.au/cb_pages/ub_spatial_bore_database.php]
- State Library of Victoria (SLV) - [Boring Records](https://www.vvg.org.au/cb_pages/digitisation_of_historical_records.php) [https://www.vvg.org.au/cb_pages/digitisation_of_historical_records.php]
- Environmental Protection Authority Victoria (EPA) - TBC

A.4.2. Demo Description and Links

What have we done?:

- The demo is currently designed to expose a single borehole via its real-world identifier.
- The application resolves a URI-14 URI pattern for the real-world-feature of the form <https://geo.org.au/id/well/46217> which performs a 303 redirect to URL-14 landing content at: <https://geo.org.au/info/46217>
- The application then uses a lookup tool (rosetta stone) to determine which of the various data providers have a corresponding borehole and the source-specific ID needed to access URL-200 data resource for that borehole.
- Basic Content Negotiation via accepts header caters for both HTML (with embedded JSON-LD) and straight JSON-LD. The format can be overridden with either a .json or ?f=json in the URI For the JSON-LD the landing-content encodes links to various representations as URL-200s in a SubjectOf block of associations.

Demo Screenshot(s)

Well 46217

Meta Information Resource (MIR) for <https://geo.org.au/id/well/46217>

| | | |
|------------|---|-------------------------|
| Type | feature,well | JSON-LD |
| Identifier | https://geo.org.au/id/well/46217 | |
| Name(s): | 46217,ARDNO 2 | |

Representations

Information from GEDIS - Geoscience Victoria (GSV) ([view on GeoVic](#))

| | |
|----------|---|
| Borehole | XML (gsml) HTML JSON-LD |
|----------|---|

Water Management Information System (WMIS) - Victorian Government (DELWP) ([view on WMIS](#))

| | |
|------------------------------|--|
| Well/Borehole | XML (gwml2) HTML JSON-LD |
| Observations (Water quality) | XML JSON-LD |

Relationships

Figure 13. CeRDI Landing Content Screenshot

Links to Demo Resources

Table 2. FedUni Meta Resource landing-content Examples

| Demo | Link |
|------------------------|--|
| JSON LD Example | https://geo.org.au/info/well/46217?f=json [https://geo.org.au/info/well/46217?f=json] |
| Photos/Reports | https://geo.org.au/info/well/WA1 [https://geo.org.au/info/well/WA1] |
| State library Archives | https://geo.org.au/info/well/326217 [https://geo.org.au/info/well/326217] |

For the URL-200 data resources, in most cases, we were starting with existing WFS services delivering complex features as GML. We have made use of Alistair Ritchie’s WFS mediator to allow on-the-fly conversion of the GML into JSON-LD and HTML (as implemented in ELFIE) We have not been able to validate these yet apart from checking that they generate something that looks like JSON-LD.

These data resources follow a URI scheme `/sourceprovider/data/datatype/featuretype/id`

Table 3. FedUni Data Resource Examples

| Demo | Link |
|---|---|
| <i>WMIS Service</i> | https://id.cerdi.edu.au/wmis/data/gwml2/well/46217?f=json [https://id.cerdi.edu.au/wmis/data/gwml2/well/46217?f=json] |
| <i>GSV service</i> | https://id.cerdi.edu.au/gsv/data/gsm12/borehole/46217?f=json [https://id.cerdi.edu.au/gsv/data/gsm12/borehole/46217?f=json] |
| <i>Lab data (water quality) ObservationCollection via bore ID</i> | https://id.cerdi.edu.au/wmis/data/sosa/observationcollection/46217?f=json [https://id.cerdi.edu.au/wmis/data/sosa/observationcollection/46217?f=json] |

A.4.3. Demo findings and potential next steps

The Federation University Use Case was that information about a single real-world feature (a Non-Information Resource, which in this case was a ‘Borehole’) was available online, in multiple formats and representations; and from multiple authoritative sources. SELFIE sought a mechanism whereby these various resources could be related to each other in a way that was discoverable and machine navigable e.g. search engines. The proposed SELFIE solution was JSON-LD landing-content that linked these various online resources together. SELFIE achieved all Federation University’s Use Case requirements, except in one crucial area. There was no satisfactory solution to encoding landing content for different data content. For example, different resources about a single borehole may contain data that is structured according to various domain models, i.e. different ‘profiles’, such as GeoSciML:Borehole, GroundWaterML2:Borehole or GroundWaterML2: GW_Well. Four options were considered as part of the experiment:

1. Use schema.org:sameAs

- The understanding of the IE was that sameAs, whether schema.org or the more rigorous owl:sameAs, asserts that the two resources are literally the same. That is, either resource could be used and the same outcome would result. This is clearly not the case in the Federation University example where the three domain models are providing significantly different data content about the same resource.

2. Use schema.org:subjectOf

- Other representations of the URI-14 (borehole), such as links to images, pdf reports, html pages etc. all comfortably fit under this property. However, FedUni felt that it did not adequately capture the fact that the domain model data was structured according to defined models and that the relationship between them was significant, and potentially navigable, compared to the other representations, which were more ‘here is a link, but we don’t know what you will find there’. Nor is there any one of the three profiles that could be considered a ‘primary’ representation for the URI-14 (borehole) and the other two subsidiary and under the ‘subjectOf’ property. Rather, all three are primary representations of the URI-14 (borehole).

3. Use W3C profiles (e.g. <https://w3c.github.io/dxwg/profiles/>)

- The Federation University Use Case could be met by using content negotiation by profile. However, this was considered outside of SELFIE scope. Even so, it is unclear how practical

this solution is. It requires domain groups to establishing and maintain (govern) domain models (already a difficult and perhaps unsuccessful endeavor), and then establishing and maintaining profiles of these domain models that can be referenced by data providers. During the IE even the relatively simple case of generating a JSON-LD context of a simplified GeoSciML model ran afoul of the Domain Group who manage the XML encoded model. Establishing and maintaining JSON-LD contexts for the required profiles is well beyond the capacity and remit of FedUni.

4. Use domain model properties

- This was suggested as a mechanism to encode collections of observations about the URI-14. For example, the property `sosa:isFeatureOfInterestOf` could contain the link to `sosa:ObservationCollection`. For the FedUni Use Case, `gwWellConstruction` could contain `gwml:Borehole`, but it is unclear what the property is that would contain the top level domain features such as `gwml:GW_Well` and `gsml:Borehole`. Investigating this further was considered out of scope for the IE.

What we have not done: There are some fairly major parts still missing from this Demo.

- We have not exposed the catalogue of resources (bores). There is no provision for a search engine to crawl and index all bores and their data resources.
- We are only in the preliminary stages of generating RDF or TTL format options.

A.5. SELFIE Demonstration Write Up: Loc-I

A.5.1. Use Case Description

The [Location Index](http://locationindex.org) [http://locationindex.org] (Loc-I) project aims to enable capability for integrating and analysis of spatio-temporal data in a reliable, effective and efficient manner across information domains and organizations, initially focusing on the public sector agencies in Australia. These include social, economic, business, and environmental data with location references embedded within the data (e.g. census district, water drainage regions, and address identifiers). Loc-I is part of the Data Integration Partnership for Australia (DIPA) initiative, which seeks to maximize government data to improve policy advice.

Table 4. LOCI Demo URLs

| Demo | Link |
|---|---|
| <i>Loc-I Explorer</i> [https://explorer.loci.cat/] - interactive demonstrator for user discovery of location features by a geo-point, or a location label. | https://explorer.loci.cat/ |
| <i>Loc-I Integration API Search by label</i> | https://api2.loci.cat/api/v1/location/find-by-label?query=50055290000 |
| <i>Location resource (ASGS) landing page</i> | http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000 |
| <i>Location resource (ASGS) alternates views</i> | http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000?_view=alternates |

| Demo | Link |
|--|---|
| Location resource (ASGS) RDF/Turtle view | http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000?_format=text/turtle |
| Location resource (ASGS) JSON-LD view | http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000?_format=application/ld+json |

User Story

Helga is an **Enterprise data warehouse manager** that manages data warehouse of a public sector organization. The data warehouse contains observation data which is captured by researchers in her organization. The observation database captures data and includes a field for location. The location information values captured is varied and can include: textual description, place name, region according to a specific geographic classification, GPS location (lat-long). Helga would like to harmonize location information so that they are consistently and precisely captured and requires a tool for searching, resolving and consistently referring to the location.

Datasets and Sources

- [Australian Statistical Geography Standard](https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS)) [https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS)] (ASGS) published by the Australian Bureau of Statistics.
- [ASGS MeshBlocks 2016](https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1270.0.55.001~July%202016~Main%20Features~Mesh%20Blocks%20(MB)~10012) [https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1270.0.55.001~July%202016~Main%20Features~Mesh%20Blocks%20(MB)~10012] is a set of geographical areas defined by the ABS and are used as building blocks for the larger regions of the ASGS. MeshBlock area identifiers are embedded in observation data such as the [2074.0 - Census of Population and Housing: Mesh Block Counts, Australia, 2016](https://www.abs.gov.au/ausstats/abs@.nsf/mf/2074.0) [https://www.abs.gov.au/ausstats/abs@.nsf/mf/2074.0].

A.5.2. Demo Description and Links

Helga, the Enterprise Data Warehouse manager, is creating ETL scripts to append Loc-I identifiers to the data warehouse she is managing for references to location by label or ID.

Links to Demo Resources

Helga uses the [Loc-I Explorer](https://explorer.loci.cat/) [https://explorer.loci.cat/] app to discover location features by label.

The Loc-I Explorer app fires off a query to the Loc-I Integration API, specifically the Search by label API at <https://api2.loci.cat/api/v1/location/find-by-label?query=50055290000> [https://api2.loci.cat/api/v1/location/find-by-label?query=50055290000]

Helga gets a list of matching resources in the results page of the Loc-I Explorer.

Helga clicks on the Loc-I resource link to verify that it's the right one and gets the landing page: <http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000> which is redirected to the info resource <https://asgsld.net/2016/object?uri=http%3A%2F%2Flinked.data.gov.au%2Fdataset%2Fasgs2016%2Fmeshblock%2F50055290000>.

Helga clicks on the link to alternates views to check-out what other views and formats are available: http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000?_view=alternates which 303 redirects to https://asgslid.net/2016/object?uri=http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000&_view=alternates.

Helga views the MB Location resource in RDF/Turtle format: http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000?_format=text/turtle which 303 redirects to https://asgslid.net/2016/object?_view=asgs&_format=text%2Fturtle&uri=http%3A%2F%2Flinked.data.gov.au%2Fdataset%2Fasgs2016%2Fmeshblock%2F50055290000.

Helga views the MB Location resource in JSON-LD format: Location resource (ASGS) JSON-LD view: http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000?_format=application/ld+json which 303 redirects to https://asgslid.net/2016/object?_view=asgs&_format=application%2Fld+json&uri=http%3A%2F%2Flinked.data.gov.au%2Fdataset%2Fasgs2016%2Fmeshblock%2F50055290000.

Helga resolves the `wfs` view for the MB Location resource in: http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000?_view=wfs which 303 redirects to https://geo.abs.gov.au/arcgis/services/ASGS2016/MB/MapServer/WFSServer?service=wfs&version=2.0.0&request=GetFeature&typeName=MB:MB&Filter=%3Cogc:Filter%3E%3Cogc:PropertyIsEqualTo%3E%3Cogc:PropertyName%3EMB:MB_CODE_2016%3C/ogc:PropertyName%3E%3Cogc:Literal%3E50055290000%3C/ogc:Literal%3E%3C/ogc:PropertyIsEqualTo%3E%3C/ogc:Filter%3E.

Helga is satisfied with the resource and embeds the Loc-I identifier (<http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000>) in the data warehouse and makes a note in the row about the WFS view along with its link (http://linked.data.gov.au/dataset/asgs2016/meshblock/50055290000?_view=wfs).

A.5.3. Demo findings and potential next steps

The demo uses NIR for identifying the resource (MeshBlock), its views, and its formats. This provides a separation between the NIR, which has been setup with <http://linked.data.gov.au/dataset/asgs2016/> namespace prefixes arranged with the Australian Linked Data Working Group (ALDWG), and the implementation site <http://asgslid.net>. The intention was to provide flexibility in case the implementation sites needed to change or move. The status of `asgslid.net` is a research operations grade resource, and we are exploring production operations grade hosting arrangements. However, the advantage of the way Loc-I NIRs are minted means that we should not need to change once the production operations grade hosting arrangements are determined.

In this case, it is important that any downstream applications and clients to use the NIRs and resolve them so that the applications and clients are not affected by any changes to the underlying implementations.

A limitation of the current demo is that the current ASGS landing page doesn't embed any JSON-LD (with `schema.org` tags). For indexing by search engines for discovery by search engines (e.g. Google), we would like to explore adding this in as per the ELFIE/SELFIE recommended discovery profile view.

Appendix B: Revision History

Table 5. Revision History

| Date | Editor | Release | Primary clauses modified | Descriptions |
|-----------------|---------------|----------------|---------------------------------|-----------------------------------|
| July 4, 2020 | D. Blodgett | .1 | all | transfer from google doc |
| July 6, 2020 | D. Blodgett | .2 | all | fix formatting and add images |
| July 19, 2020 | E. Boisvert | .2 | all | comments for editor consideration |
| October 8, 2020 | G. Hobona | .3 | all | Final staff review |

Appendix C: Bibliography

[1] Blodgett, D., Cochrane, B., Atkinson, R., Grellet, S., Feliachi, A., Ritchi, A.: OGC Environmental Linked Features Interoperability Experiment Engineering Report. OGC 18-097, Open Geospatial Consortium, <https://docs.opengeospatial.org/per/18-097.html> (2019).