ELFIE - The OGC Environmental Linked Features Interoperability Experiment

Kathi Schleidt[1(✉)], Michael O'Grady[2], Sylvain Grellet[3],
Abdelfettah Feliachi[3], and Hylke van der Schaaf[4]

[1] DataCove e.U., Robert Hamerling Gasse 1/14, 1150 Vienna, Austria
`kathi@datacove.eu`
[2] UCD Dublin, Belfield Dublin 4, Ireland
`michael.j.ogrady@ucd.ie`
[3] BRGM, 3 avenue Claude-Guillemin, BP 36009,
45060 Orleans Cedex 2, France
`{s.grellet,a.feliachi}@brgm.fr`
[4] Fraunhofer IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, Germany
`hylke.vanderschaaf@iosb.fraunhofer.de`

Abstract. The OGC Environmental Linked Feature Interoperability Experiment (ELFIE) sought to assess a suite of pre-existing OGC and W3C standards with a view to identifying best practice for exposing cross-domain links between environmental features and observations. Environmental domain models concerning landscape interactions with the hydrologic cycle served as the basis for this study, whilst offering a meaningful constraint on its scope. JSON-LD was selected for serialization; this combines the power of linked data with intuitive encoding. Vocabularies were utilized for the provision of the JSON-LD contexts; these ranged from common vocabularies such as schema.org to semantic representations of OGC/ISO observational standards to domain-specific feature models synonymous with the hydrological and geological domains. Exemplary data for the selected use cases was provided by participants and shared in static form via a GitHub repository. User applications were created to assess the validity of the proposed approach as it pertained to real-world situations. This process resulted in the identification of issues whose resolution is a prerequisite for wide-scale deployment and best practice definition. Addressing these issues will be the focus of future OGC Interoperability Experiments.

Keywords: Open geospatial standards · Linked data · Environmental features

1 Introduction

Environmental data is becoming ever more ubiquitous; this poses many difficulties going forward as data pertaining to an arbitrary real world situation will differ according to the perception of the provider as to what is contextually important. Advances in geospatial technologies continue to resolve many of the problems entailed in spatial data provision; nonetheless, certain domains, for example, the hydrological and geological domain, often require more comprehensive information on how spatial objects interrelate than is innately apparent from their topological relationships alone.

Open Geospatial Consortium (OGC) data models and services are lacking when it comes to exposing cross-domain links between environmental domains and sampling features. The existing suite of OGC Web Services (OWS) follows classic web service design patterns. Data encoding is done in XML, only fragile linkages between spatial objects are possible via xlink; however, the semantic context behind these links is difficult to include. There is no clear guidance on xlink targets and resolution mechanisms. Ultimately, current OWS are still driven by simple feature concepts that assume discrete standalone datasets; there is little support for interlinkage beyond the boundaries of virtual dataset concepts.

Linked open data has the potential to radically transform the normal OGC service pattern (GetCapabilities request followed by lengthy introspection and further DescribeFeatureType requests to understand contents) [1] by encoding associations between features as linked data predicates. The potential of Linked Data in the geospatial domain has been acknowledged by the academic community. To elucidate briefly: A framework for utilizing geospatial Linked Data with the Web Feature Service (WFS) has been proposed and evaluated within the biodiversity domain [2]. Additionally, The Open European Location Services project, a collaboration between the national mapping agencies of Finland, the Netherlands, Norway and Spain, demonstrated the capabilities of linked data for international geospatial data provision [3]. Closely related to this project was an initiative in the Netherlands that focused on the visualization of linked geospatial data [4].

The OGC foresees the concept of Interoperability Experiments (IE) for the structured exploration of potential future topics; thus, as specific implementation conventions and best practices are not available, an OGC IE was initiated to explore how linked data might be best harnessed in OGC services and identify a roadmap for future activities.

The Environmental Linked Features Interoperability Experiment (ELFIE) [5] "focused on encoding relationships between cross-domain features and linking available observations data to sampled domain features" within the hydrogeological domain, adopting and integrating relevant concepts from the semantic community. In order to keep the experiment within scope, the ELFIE specifically focused on linked data requirements and encoding options; in this process, diverse areas of future work were identified and documented for potential future IEs.

2   Problem Statement

The ELFIE seeks to define a method of interlinking domain features and observations of them, whilst maintaining a focus on the semantics of these linkages. An additional objective concerned the provision of a simple solution that would be easily adoptable by developers and users across software platforms whilst leveraging existing standards and best practices (notably SDW BP 2&3 [6]) and, as far as possible, integrating standard taxonomies and ontologies.

At the onset of the ELFIE, work on the specification of an OGC API [7] was still very much a work-in-progress (as WFS3). Thus, the ELFIE also aims at illustrating how RESTful and Linked Data principles can be leveraged to create a reusable

approach for encoding information models specified in cross-disciplinary applications, independent of any specific web-API pattern. In addition, the possibility of defining multiple "views" of the same data resource, with each view providing a specific subset of the linked-data graph, was explored during the IE.

Keeping within the parameters of the IE, environmental domain models were thus limited to landscape interactions with the hydrologic cycle. Data was constrained to surface water, groundwater, well/borehole structure and soil moisture. The following data models, where utilized, serving as proto-ontologies: WaterML2, GroundwaterML 2, GeosciML 4 and SoilML. Topics pertaining to network architecture, default behavior when dereferencing an identifier, and discoverability were all deemed as being of utmost importance, but out-of-scope for this IE.

Initially, a wide range of potential use cases was considered with topics ranging from floods to droughts, water quality and quantity, as well as causes and impacts. Additional data sources ranging from meteorological data (both measurement and forecast), elevation data, and critical infrastructure (transport networks, bridges and so forth) and known critical discharge locations (for example, mines) were also identified and integrated.

These use cases were then iteratively reduced to a smaller but representative set for testing and prototyping of the developed concepts; relevant aspects identified in those use cases not utilized were retained in order to assure a comprehensive solution.

In addition to spanning environmental domains, a key feature of the selected use cases was that their constituent data was often administered within different institutions or even countries.

The following use cases were then examined in greater detail - the final two may be regarded as flagships demonstrating the intrinsic usefulness of linked data in environmental/cross-domain contexts:

1. Water budget summary: integrating water budget data with data on the hydrographic network, watershed boundary and outlet, this use case strives to give the user a summary overview of the water budget for a watershed.
2. Flood risks and impacts: by linking available hydrographic information on a watershed with meteorological and water level information as well as the relevant transport networks, real time information of benefit to decision-makers can be provided.
3. Groundwater level monitoring: integrating boreholes and other monitoring facilities with aquifers, thereby gaining a better understanding of groundwater levels.
4. Surface-groundwater networks interaction: provides a comprehensive overview of a water system by applying a linked data approach to all relevant domain features as well as measurements being taken on these features.
5. Watershed data index: by applying linked data principles to monitoring sites and watersheds, data stemming from water quality and quantity sensors is brought into context with the hydrographic network, allowing for a wide array of linked watershed information use cases.

3 Proposed Solution

OGC data models were used extensively in the ELFIE. The Observations and Measurements (O&M) conceptual model provided a high level organizing framework for most ELFIE documents. The OGC-W3C Spatial Data On the Web Working Group implementation of O&M (Sensor, Observation, Sample, and Actuator (SOSA) and Semantic Sensor Networks (SSN)) [8] was used directly because of its applicability to the linked data technology pursued in the ELFIE. The GeoSPARQL [9] ontology was also used directly for representing geometries and spatial relations between features, and to overcome the technology gap between GeoJSON and JSON-LD.

Domain specific data models such as HY_Features [10], GWML2 [11] and GeoSciML [12] were also used for feature types and relations in the ELFIE linked data documents.

As the ELFIE seeks to provide a pragmatic and implementable solution while leveraging the power of linked data, it was decided to explore the potential of JSON-LD; this decision was further influenced by ongoing work in the OGC towards the adoption of JSON. JSON-LD Context files were created for attributes that were deemed important for the use cases considered; in line with the approach of providing multiple views of the same data object, context files were created to support two exemplary conceptual views - "preview" and "network". Well established vocabularies were referenced within the contexts, ranging from schema.org, skos and geosparql for general concepts, the W3C Semantic Sensor Network (SSN) and Sensor, Observation, Sample, and Actuator (SOSA) ontologies for observational models to various domain models such as GWML2, HY_Features and GeoSciML (Fig. 1).
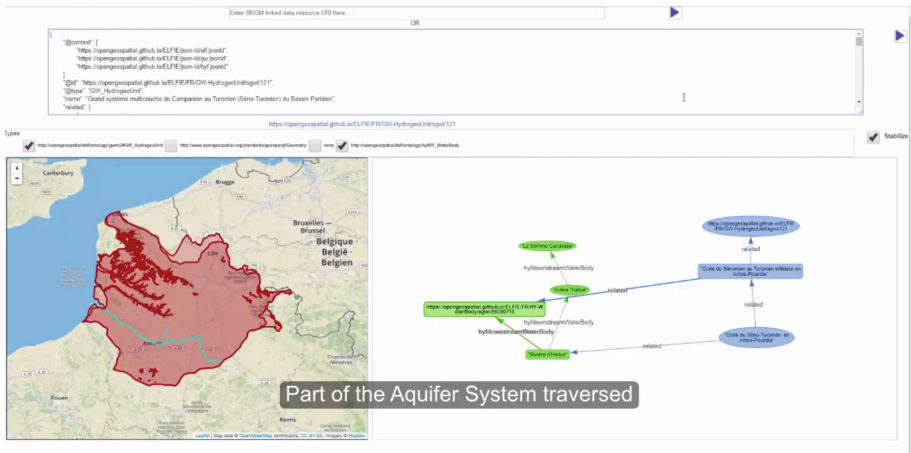
Fig. 1. BLiv viewer for exploration of linked JSON-LD resources

Exemplary data files were created for each use case, whereby in most instances it was decided to provide a small representative set of files, and to only provide these

statically on GitHub. In addition, for the Surface-groundwater Networks Interaction Use Case (Number 4), a wrapper built on top of an OGC SensorThings API deployment was created for the provision of dynamic data. Various GUIs were developed for exploration of the provided data. Exceptional among these is the BLiv viewer developed by BRGM, providing the user with parallel views on the raw data, the underlying semantic graph, as well as a conventional map.

4   Issues and Future Work

While the ELFIE was successful in reaching its primary objectives, this exercise also served to highlight various issues that must be further investigated within future IEs. These issues are summarized as follows:

- Resolvable Identifiers: when utilizing existing OGC services for data provision, a specific feature could only be referenced via a complex and unstable request URI. Rewriting approaches were successfully tested, but there was a consensus that this could only ever be a work-around; APIs allowing resolution of URI based identifiers would be essential.
- Domain Feature Model: while the standard vocabularies utilized in the JSON-LD contexts are well suited for referencing, issues were encountered pertaining to the domain vocabularies only available in conceptual (UML) form, as well as those relying on XML Schema. Ongoing work on the OGC Register should provide valuable insights going forward.
- Spatial Representation: utilization of GeoJSON structures for spatial representation is not possible when using JSON-LD due to the underlying RDF based structure (specifically the unordered arrays). While Point data can be provided in a form valid for both standards, this is not possible for more complex geometries. For this purpose, the ELFIE utilized GeoSPARQL for the provision of geometry information. However, being able to leverage the widespread use of GeoJSON would be valuable
- Multiple Representations of an Object: one real-world-object can have multiple data representations, at times stemming from different organizations, or exposing different facets of the available data. Mechanisms for maintaining alignment must be explored.

Specification work on the OGC API (previously WFS3) has been progressing since the finalization of the ELFIE, with the first prototypes becoming available. Currently, there is work exploring the potential of extending the OGC API with JSON-LD, to allow for implementation of the concepts developed within the ELFIE via standardized OGC services.

At present, the Second Environmental Linked Features Interoperability Experiment (SELFIE) is carrying this work further, focusing on feature identification, referencing real world objects, and URL resolution. Special focus is on the relation between the real-world-objects being described and their digital representations.

References

1. Open Geospatial Consortium: Web Services Context Document (OWS Context). https://www.opengeospatial.org/standards/owc
2. Vilches-Blázquez, L.M., Saavedra, J.: A framework for connecting two interoperability universes: OGC web feature services and linked data. Trans. GIS **23**(1), 22–47 (2019)
3. Ronzhin, S., et al.: Next generation of spatial data infrastructure: lessons from linked data implementations across Europe. Int. J. Spat. Data Infrastruct. Res. **14**, 84–106 (2019)
4. Folmer, E., Beek, W., Rietveld, L.: Linked data viewing as part of the spatial data platform of the future. Int. Arch. Photogrammetry Rem. Sens. Spat. Inf. Sci. **42**(4), 49–52 (2018)
5. Open Geospatial Consortium: Environmental Linked Features Interoperability Experiment. https://opengeospatial.github.io/ELFIE/
6. OGC W3C Spatial Data on the Web Best Practices. https://www.w3.org/TR/sdw-bp/
7. Open Geospatial Consortium: OGC API (OAPI) Common Specification. https://github.com/opengeospatial/oapi_common
8. OGC W3C Semantic Sensor Network Ontology. https://www.w3.org/TR/vocab-ssn/
9. Open Geospatial Consortium: GeoSPARQL - A Geographic Query Language for RDF Data. https://www.opengeospatial.org/standards/geosparql
10. Open Geospatial Consortium: WaterML 2: Part 3 - Surface Hydrology Features (HY_Features) - Conceptual Model. http://docs.opengeospatial.org/is/14-111r6/14-111r6.html
11. Open Geospatial Consortium: OGC WaterML 2: Part 4 – GroundWaterML 2 (GWML2). http://docs.opengeospatial.org/is/16-032r2/16-032r2.html
12. Open Geospatial Consortium: Geoscience Markup Language 4.1 (GeoSciML). https://docs.opengeospatial.org/is/16-008/16-008.html