A framework for connecting two interoperability universes: OGC Web Feature Services and Linked Data

Luis M. Vilches-Blázquez¹ | Jhonny Saavedra²

¹Javeriana University, Bogota, Colombia
²Universidad Politécnica de Madrid, Spain

Abstract
Diverse studies have shown that about 80% of all available data are related to a spatial location. Most of these geospatial data are available as structured and semi-structured datasets, and often use distinct data models, are encoded using ad-hoc vocabularies, and sometimes are being published in non-standard formats. Hence, these data are isolated within silos and cannot be shared and integrated across organizations and communities. Spatial Data Infrastructures (SDIs) have emerged and contributed to significantly enhance data discovery and accessibility based on OGC (Open Geospatial Consortium) Web services. However, finding, accessing, and using data disseminated through SDIs are still difficult for non-expert users. Overcoming the current geospatial data challenges involves adopting the best practices to expose, share, and integrate data on the Web, that is, Linked Data. In this article, we have developed a framework for generating, enriching, and exploiting geospatial Linked Data from multiple and heterogeneous geospatial data sources. This proposal allows connecting two interoperability universes (SDIs, more specifically Web Feature Services, WFS, and Semantic Web technologies), which is evaluated through a study case in the (geo)biodiversity domain.
INTRODUCTION

Diverse studies have shown that about 80% of all available data are related to a spatial location (Densham & Goodchild, 1989; Li & Li, 2014). Nowadays, this statement has been strengthened, due to the fact that Web 2.0, mobile devices, citizen participation, and crowdsourcing have become a relevant stream in the collection of geo-related data with the advancement of network deployment (Salk, Sturn, See, Fritz, & Perger, 2016). These new streams are significant and have become complementary efforts to the traditional and official organizations associated with geospatial information production. Likewise, in-situ sensor network data, GPS trace data from mobile devices, geo-social media data (e.g. Twitter), and crowdsourcing/volunteered geographic information data (e.g. OpenStreetMap) contribute to the increased volume and variety in data (Guo et al., 2017).

Most of these geospatial data are available as structured¹ (e.g. CSV, Microsoft Excel, relational databases, etc.) and semi-structured² (e.g. shapefile, XML, JSON, Web services, etc.) datasets. Besides, these data often use distinct data models, are encoded using ad-hoc vocabularies, and are sometimes being published in non-standard formats. Hence, these resources are isolated in silos and cannot be shared and integrated across organizations and communities. As a result, data use and analysis become a privilege owned by some domain-specific well-educated users, and much data may not be analyzed and sufficiently exploited (Zhao, Foerster, & Yue, 2012).

In this scenario, Spatial Data Infrastructures (SDIs) have emerged and contributed to integrate data sources, systems, network linkages, standards, and institutional issues to deliver geospatial data from heterogeneous sources to the widest possible audience (Giuliani et al., 2017). These issues motivated such initiatives (SDIs) to focus on interoperability to significantly enhance data discovery and accessibility based on OGC (Open Geospatial Consortium) Web services. However, although SDIs constitute a mature technology widely developed over the last 20 years in almost every country in the world, in practice, relatively simple tasks such as finding, accessing, and using data disseminated through SDIs are still difficult for non-expert users (Tandy, van den Brink, & Barnaghi, 2017). Nevertheless, several efforts in the SDIs context have been made to significantly facilitate access to and use of geospatial data in recent years. GEOSS (http://www.geoportal.org/) and INSPIRE (https://inspire-geoportal.ec.europa.eu/) are some representative examples of these efforts. In spite of these exceptions, the aforementioned issues originate based on the fact that these initiatives have been successful in addressing syntactic and structural interoperability; however, semantic interoperability, which may solve problems related to reusing and integration, remains a major challenge (Kuhn, 2005; Janowicz et al., 2010).

In order to overcome these challenges and achieve geospatial data available on the Web in an appropriate way, the (geospatial) resources described need to be identified using Hypertext Transfer Protocol (HTTP) URIs, published in such a way that they are indexable by search engines, and connected, or linked, to other resources (Tandy et al., 2017). This entails adopting best practice for exposing, sharing, and integrating data via de-referenceable URIs on the Web (Heath & Bizer, 2011), that is, Linked Data. Such best practice is being adopted by an increasing number of data providers, leading to the creation of a global data space containing billions of assertions—the Web of Data (Heath & Bizer, 2011).

Although SDIs and Linked Data are quite different in respect of their architectures, workflows, and applied standards, they can be regarded as complementary (Wiemann & Bernard, 2016). Therefore, this combination (SDI and Semantic Web) can make the data easy to find and easy to access for non-specialist users, since geospatial data becomes integrated within the wider Web of Data. Under this approach, different work has appeared combining SDI Web services and Linked Data, especially for bringing to SDIs the integration and reasoning capabilities of the Semantic Web (Lutz & Klien, 2006; Roman & Klien, 2007; Janowicz et al., 2013, 2010; Yue, Gong, Di, He, & Wei, 2011; Hietanen, Lehto, & Latvala, 2016; Wiemann & Bernard, 2016).
The main contribution of this article is a framework for generating, enriching, and exploiting geospatial Linked Data from multiple and heterogeneous data sources. This proposal allows a connection between two interoperability universes, that is, SDIs and Semantic Web technologies, offering a promising approach to link SDIs, and more specifically Web Feature Services (WFS), to mainstream information technologies, to enable further areas of SDI application, and to facilitate the reuse and connection of multiple and heterogeneous geospatial information. Herein, we describe the process of generation, enrichment, and exploitation, illustrating them with a case study in the (geo)biodiversity domain. Furthermore, an important innovation of this work is that we consider geometries, which are the key component for connecting geospatial data, allowing that they can be retrieved and interlinked on an unprecedented level in the context of these two universes of interoperability (SDIs and Linked Data).

The remainder of this article is structured as follows. We start by providing a description of a case study, which is the application context of our proposal, in Section 2. In Section 3, we present a brief overview of the WFS and geospatial Linked Data, and related work. Our framework for generating, enriching, and exploiting Linked Data from geospatial data sources is described in Section 4, detailed using the proposed case study. Section 5 provides a discussion. Finally, we summarize some conclusions and identify future work in Section 6.

2 | A CASE STUDY WITH BIODIVERSITY DATA SOURCES

Global biodiversity is changing at an unprecedented rate as a complex response to several human-induced changes in the global environment. The magnitude of this change is so large and so strongly linked to ecosystem processes and society's use of natural resources that biodiversity change is now considered an important global change in its own right (Sala et al., 2000).

According to the Convention on Biological Diversity (https://www.cbd.int/), Colombia is listed as one of the world's "megadiverse" countries, hosting close to 10% of the planet's biodiversity. Worldwide, it ranks first in bird and orchid species diversity and second in plants, butterflies, freshwater fishes, and amphibians. With 314 types of ecosystems, Colombia possesses a rich complexity of ecological, climatic, and biological components. This varied richness represents a significant challenge for implementing sustainable development initiatives.

In recent years, the biodiversity informatics community has made tremendous strides by creating shared common vocabularies such as the Darwin Core (DwC) (Wieczorek et al., 2012) and publishing mechanisms such as the Integrated Publishing Toolkit (IPT) (Robertson et al., 2014). However, data published in this kind of system are stored in an isolated way for each dataset, which limits its analysis in an integrated way with context information. Hence, coordinating, integrating, and performing effective use of the vast amount of environmental data, which are generated from many different providers and across research domains, remain challenging (Stucky et al., 2014; Walls et al., 2014).

Driven by this scenario, we present a case study where heterogeneous (geo)biodiversity data sources are semantically integrated, enriched, and exploited in order to connect data from different domains and communities. These data sources are created and maintained by diverse national and international institutions associated with the biodiversity domain, such as the Alexander von Humboldt Research Institute (https://www.humboldt.org.co/en), the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM; https://www.ideam.gov.co), the Global Biodiversity Information Facility (GBIF; https://www.gbif.org), the Colombian Biodiversity Information System (Sistema de Información sobre Biodiversidad de Colombia, SIB Colombia), and the European Nature Information System (EUNIS; https://eunis.eea.europa.eu/). This case study focuses on (geo)biodiversity information of the Cundinamarca Department, which was selected for its diversity of ecosystems and because it has a large amount of data, widely gathered.
The heterogeneity of these data sources arises from its diversity of subdisciplines (e.g. ecosystems/community, marine/freshwater/terrestrial, and plant/animal/microbial biodiversity). In addition, adjacent disciplines in earth and life science, as well as relevant disciplines in the social sciences and humanities, have their own terminologies, specialized measurements, data models, and formats that generate heterogeneity (Reichman, Jones, & Schildhauer, 2011). Therefore, the main challenges of this case study are related to how to deal with such data sources in order to prevent data isolation and enable cross-dataset analysis.

In order to address these challenges, spatial data fusion techniques play a crucial role in fostering an integrated view of distributed spatial data sources on the Web. In addition, flexibility and interoperability are key factors in such data integration. Hence, the use of standards is a basic requirement (Wiemann & Bernard, 2016). In this sense, progress on Linked Data, ontologies, and geospatial semantics (Kuhn, 2005) are considered in this study case, due to the fact that they provide the decisive glue between models, data, and users (Janowicz & Hitzler, 2012). Moreover, the advantages of Linked Data technologies are recognized not only in biodiversity research and its related disciplines (e.g. Madin, Bowers, Schildhauer, & Jones, 2008; Roderic & Page, 2008; Deans, Yoder, & Balhoff, 2012; Parr, Guralnick, Cellinese, & Page, 2012), but also throughout the life sciences (Stevens, 2002; Blake & Bult, 2006; Good & Wilkinson, 2006; Antezana, Kuiper, & Mironov, 2009; Chen, Yu, & Chen, 2013).

Thus, we take different “traditional” (geo)biodiversity datasets and OGC Web Services (OWS) and transform them into machine-processable data. After that, we link these data with other sources of the Web of Data and OWS, using geometrical information related to (geo)biodiversity features for performing spatial comparisons between them. We also enrich existing biodiversity Linked Data with OWS, concretely with WFS such that, after transforming, linking, and enriching the (geo)biodiversity data, we exploit these data, setting connections between two interoperability universes, that is, SDIs and geospatial Linked Data.

In short, this work presents a process for generating, enriching, and exploiting (geo)biodiversity Linked Data from multiple and heterogeneous (national and international) data sources using our developed framework. Furthermore, a connection between SDIs and Semantic Web technologies is established to facilitate the reuse and connections of contained information within these dissociated universes.

3 | BACKGROUND AND RELATED WORK

In this section, we present a brief introduction to WFS and geospatial Linked Data. Also, we provide a description of the main proposals related to our work.

3.1 | Background

3.1.1 | WFS

WFS (https://www.opengeospatial.org/standards/wfs; Sinha, 2008) is an interface specified by the OGC (https://www.opengeospatial.org/) that allows the exchange of geographic data across the Web. It defines the rules for requesting and retrieving geographic information using HTTP. The interface describes the data manipulation operations on geographic features. Extensible Markup Language (XML)-based Geographic Markup Language (GML) is used for exchange of information. It should be noted that WFS supports the vector data model.

Within this kind of service, a feature may contain one or many geometries, optionally with attribute values. The standard operations of WFS are based on the GetCapabilities, DescribeFeatureType, and GetFeature requests (Vretanos, 2010).

The GetCapabilities operation generates a service metadata document describing the WFS provided by a server. This document is an XML one composed of the following sections:
• Service. This section provides information about the service itself. Therefore, it contains service metadata, such as name, title, abstract, keyword, etc.
• Capabilities. This section describes the operations that are supported by a specific service. According to the WFS specification (Vretanos, 2010), this service must support, at least, GetCapabilities, DescribeFeatureType, and GetFeature operations.
• Feature type. This section contains a list of feature types that the WFS can return. Besides, it defines transactions and query elements that are supported by each feature type.

Listing 1 shows an example of how a GetCapabilities request can be sent to the WFS server.

**LISTING 1** An example of GetCapabilities request

https://www.example.com/wfs?SERVICE=WFS&REQUEST=GetCapabilities

The DescribeFeatureType operation returns a schema description of feature types offered by an instance of the WFS. All features are composed of a set of properties, which are described in an XML schema. This schema description can be obtained through the DescribeFeatureType operation, as shown in Listing 2.

**LISTING 2** An example of DescribeFeatureType request

https://www.example.com/wfs?SERVICE=WFS&VERSION=1.0.0&REQUEST=DescribeFeatureType&TYPENAME=FEATURE_ID&SRSNAME=EPSG:4326

The GetFeature operation returns a selection of features from a data store. WFS processes a GetFeature request and returns a response document to the client that contains zero or more feature instances that satisfy the query expressions specified in the request. In this operation, the client asks for one or more features in a specific WFS encoding version, as shown in Listing 3. Besides, this operation has the capability to perform queries using their own attributes or parameters, such as bounding box, type name, feature ID, etc. Finally, the client receives a document, which can be encoded in different formats (e.g. KVP [key/value pair], XML, or GML), that contains the feature and its attribute information.

**LISTING 3** An example of GetFeature request

https://www.example.com/wfs?SERVICE=WFS&VERSION=1.0.0&REQUEST=GetFeature&TYPENAME=FEATURE_ID&SRSNAME=EPSG:4326

An extension of WFS—the Transactional Web Feature Service (WFS-T) (Vretanos, 2014)—allows the creation, deleting, and updating of features, but these functionalities are currently not addressed in this work.

### 3.1.2 Geospatial Linked Data

The World Wide Web Consortium (W3C) has been working on a new generation of standards with capabilities for semantic expression and data integration to make the Semantic Web a reality. This Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation (Berners-Lee, Hendler, & Lassila, 2001). In this context, Linked Data has been identified as one of the best practices of the Semantic Web for creating a shared information space through interlinking single, context-specific data fragments within a distributed environment (Bizer, 2009; Auer, Lehmann, Ngomo, & Zaveri, 2013).

Since the representation and publication of geospatial data as Linked Data has only recently been addressed, we provide some background on the Linked Data initiative. The principles of Linked Data were first outlined by Berners-Lee (2006) using the following four guidelines: (1) use of URLs as names for things; (2) use of HTTP URLs so that people can look up those names; (3) when someone looks up a URI, provide useful information, using
standards, such as the Resource Description Framework (RDF) and SPARQL Query Language for RDF (SPARQL); and (4) include links to other URIs, so that they can discover more things.

In these aforementioned guidelines, two key elements stand out in the Linked Data scenario: RDF and SPARQL. On the one hand, RDF (https://www.w3.org/TR/rdf11-concepts/) is the standard knowledge representation language for the Semantic Web, an evolution of the World Wide Web that aims to provide a well-founded infrastructure for publishing, sharing, and querying structured data on the Semantic Web (McDonald & Levine-Clark, 2017). RDF provides a framework for representing information on the Web, of which syntax (a data model) has two key data structures: (1) RDF graphs are sets of subject-predicate-object triples, where the elements may be URIs, blank nodes, or datatyped literals — these graphs are used to express descriptions of resources, among other things personal information, social networks, metadata about digital artifacts, and so on, as well as providing a means of integration over disparate sources of information; and (2) RDF datasets are used to organize collections of RDF graphs, and comprise a default graph and zero or more named graphs.

On the other hand, SPARQL (https://www.w3.org/TR/sparql11-query/) is a semantic query language able to retrieve and manipulate data stored in RDF. This language can be used to express queries across diverse data sources, whether the data are stored natively as RDF or viewed as RDF via middleware. Besides, SPARQL contains capabilities for querying required and optional graph patterns, along with their conjunctions and disjunctions. SPARQL also supports aggregation, subqueries, negation, creating values by expressions, extensible value testing, and constraining queries by source RDF graph. The results of SPARQL queries can be result sets or RDF graphs. Further details about these standards (RDF and SPARQL) can be found in the W3C specifications: https://www.w3.org/RDF/ and https://www.w3.org/TR/sparql11-overview/, respectively.

In the geospatial context, the transformation and publication of geographical data as Linked Data was pioneered by initiatives from GeoNames (https://www.geonames.org/ontology/documentation.html), OpenStreetMap (Auer, Lehmann, & Hellmann, 2009), and Ordnance Survey (Goodwin, Dolbear, & Hart, 2009). After these initiatives, an increasing number of geospatial datasets have been published in the Linked Data cloud (https://lod-cloud.net/). Currently, Basic Geo Vocabulary (https://www.w3.org/2003/01/geo/) (WGS84 lat/long) is one of the top five vocabularies by usage in datasets (https://stats.lod2.eu/stats) in the context of the Web of Data among the existing vocabularies for describing geospatial Linked Data. This is a basic vocabulary that provides a namespace for representing lat(itude), long(itude), and other information about spatially located things, using WGS84 as a reference datum. Listing 4 shows an example using the WGS84 vocabulary.

Listing 4 An example of a point (lat/long) described with the WGS84 vocabulary

```sparql
@prefix geo: <https://www.w3.org/2003/01/geo/wgs84_pos#> .
a geo:Point ;
geolat "55.701" ;
geolong "12.552" .
```

OGC has defined an approach for representing and querying geospatial data on the Semantic Web. This effort, called GeoSPARQL (Perry & Herring, 2012), defines a vocabulary for representing geospatial data in RDF, and an extension to the SPARQL query language for processing geospatial data. In addition, GeoSPARQL is designed to accommodate systems based on qualitative spatial reasoning and systems based on quantitative spatial computations. This proposal allows distinct kinds of geometry (e.g. polygons, lines, points, multipoints, etc.) to be described, adding multiple coordinate reference systems, and bringing to the Linked Data cloud the possibility of expressing spatial relations for querying geographic datasets (e.g. intersects, touches, overlaps, etc.). Listing 5 shows an example of a geometry expressed according to GeoSPARQL. Geometries are defined by the class Geometry and the coordinates can be encoded in an RDF literal of type Well-Known-Text (WKT), which corresponds to Herring (2010), using a single RDF property, namely asWKT.

Listing 5 An example of a geometry expressed according to GeoSPARQL.

```sparql
a Geometry .
geo:asWKT "MULTIPOLYGON (((
-7.307 43.655,
-7.307 43.657,
-7.306 43.657,
-7.306 43.657,
-7.306 43.657,
-7.307 43.655,
-7.307 43.655)))" .
```
LISTING 5 An example of a geometry described with the GeoSPARQL vocabulary

```turtle
@prefix geosparql: <https://www.opengis.net/ont/geosparql#> .
@prefix sf: <https://www.opengis.net/ont/sf#> .
@prefix ex: <https://example.org#> .
ex:GEOMETRY_1 a geosparql:Geometry ;
geosparql:asWKT "<https://www.opengis.net/def/crs/EPSG/0/4326>
LINESTRING(4 5, 6 7, 7 4, 6 2)"^^sf:wktLiteral.
```

GeoSPARQL also offers the possibility of using GML to encode geometries. In this case, the data type (GMLLiteral), property (asGML), and the URL for the geometry type have to be changed accordingly. Further details about GeoSPARQL are described in Perry and Herring (2012).

The geospatial Linked Data is stored and managed by triple stores, also known as RDF store or knowledge bases. According to Battle and Kolas (2012), they are capable of better handling several types of problems which relational databases struggle with, or are not intended to perform: queries with many joins across entities (Weiss, Karras, & Bernstein, 2008), queries with variable properties (Weiss et al., 2008), and ontological inference on datasets. The W3C has collected a compendium of existing triple stores (https://www.w3.org/2001/sw/wiki/Category:Triple_Store), where different implementations related to geospatial domains can be found, for instance Apache Marmotta (https://marmotta.apache.org/), OpenLink Virtuoso (https://virtuoso.openlinksw.com/rdf/), Parliament (https://parliament.semwebcentral.org/), Strabon (https://www.w3.org/2001/sw/wiki/Strabon), or USeekM (https://www.w3.org/2001/sw/wiki/USeekM).

### 3.2 Related work

Several works have contributed to provide geospatial context to the Web of Data. Thus, different proposals appear in the state-of-the-art associated with geospatial reference data (Varanka, 2008; Auer et al., 2009; Goodwin et al., 2009; Battle & Kolas, 2012; Usery & Varanka, 2012; GADM-RDF (https://gadm.geovocab.org/); Feliachi, Abadie, Hamdi, & Atemezing, 2013; Athanasiou, Giannopoulos, Garcia-Rojas, & Le Grange, 2014; Koubarakis et al., 2014; Vilches-Blázquez, Villazón-Terrazas, Corcho, & Gómez-Pérez, 2014; Lehmann et al., 2015), sensor data (Barnaghi, Presser, & Moessner, 2010; Patni, Henson, & Sheth, 2010), and geospatial provenance (Celino, 2013; Yuan, Yue, Gong, & Zhang, 2013).

In this context, different proposals have focused concretely on the transformation process from geospatial data to RDF. Schade and Cox (2010) present a proposal for transforming GML to RDF and for extending data types offered through WFS to other types, such as GML, RDF, and HTML. Usery and Varanka (2012) describe a conversion from vector and raster data to RDF. The authors encode vector data in GML format, by means of pre-computing of topological relations, and then convert them into RDF triples. Another proposal is described in van den Brink, Janssen, Quak, Stoter, and Kadaraster (2014), where the authors present a semi-automatic transformation from geographic information models and GML data to RDF data.

Likewise, different tools have appeared in order to make the transformation of “traditional” geospatial data (shapefiles or geospatial databases) into RDF a little easier. Thus, in Vilches-Blázquez et al. (2010), Geometry2RDF (https://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/technologies/151-geometry2rdf/index.html) is presented, a library for generating RDF files from geometrical information available in GML or WKT using the WGS84 vocabulary. Sparqlify (https://github.com/SmartDataAnalytics/Sparclify) is a scalable SPARQL-SQL rewriter whose development is associated with the LinkedGeoData project (https://linkedgeodata.org/). In the context of this project, this tool is used for generating RDF data from mapping information, which is stored in a relational database (PostGIS) together with the OpenStreetMap data. TripleGeo (https://github.com/GeoKnow/TripleGeo) (Patroumpas, Alexakis, Giannopoulos, & Athanasiou, 2014) transforms geometries from shapefiles, GML, KML, INSPIRE-aligned datasets for seven data themes, and several geospatially enabled DBMSs (i.e. PostGIS, MySQL, PostgreSQL).
Oracle spatial, etc.). This tool is based on Geometry2RDF and uses the WGS84 vocabulary and several geometric types of GeoSPARQL. Kyzirakos, Vlachopoulos, Savva, Manegold, and Koubarakis (2014) describe GeoTriples, a tool that allows the transformation of geospatial data stored in spatially enabled relational databases and raw files. This tool is implemented as an extension to the D2RQ platform (https://d2rq.org/) and uses GeoSPARQL and stSPARQL as the target vocabulary. The main limitation of these tools is that they deal only with “traditional” geospatial data (shapefiles, GML, and geospatial databases) and do not support OGC Web services.

Regarding interlinking of datasets, which constitutes a key issue of Linked Data, we mention some works that have studied the interconnection of geospatial Linked Data by extending linking criteria and adding geospatial characteristics, such as the use of distance and semantic criteria (Auer et al., 2009; Patni et al., 2010; Salas & Harth, 2011; Vilches-Blázquez, Saquicela, & Corcho, 2012; Feliachi et al., 2013; Ngonga, 2013; Yuan et al., 2013), or the discovery and exploitation of topological relations (Smeros & Koubarakis, 2016; Grütter, Purves, & Wotruba, 2017; Sherif, Dreßler, Smeros, & Ngomoa, 2017). However, these approaches deal only with RDF datasets and do not support or exploit complex geometries gathered in other (geospatial) formats.

On the other hand, there are some efforts where interactions between SDIs and the Semantic Web have been collected. In Lutz and Klien (2006), the authors present an approach to ontology-based geographic information retrieval that contributes to solving problems related to semantic heterogeneity using a graphical user interface (GUI) and a well-known domain vocabulary. Roman and Klien (2007) describe a semantic framework that facilitates the employment of geospatial Web services to solve a specific task in geospatial decision-making. Maué, Schade, and Duchesne (2009) introduce a proposal for semantic annotations for OGC standards, which explores how semantic annotations can be carried out at three levels: metadata, data models, and data instances. This proposal is extended in Maué, Michels, and Roth (2012), where a lightweight RESTful application programming interface (API) deployed as a free service injects semantic annotations into existing Web service descriptions without breaking the standards. In Janowicz et al. (2010), the authors propose creating profiles of existing services that implement a transparent semantic layer between the OGC Web services and the Semantic Web world. Wiemann and Bernard (2016) propose a service-oriented decomposition of fusion processes and strategies for feature relation management using Linked Data in combination with SDIs. This proposal is validated with use cases, which focus on the fusion of road networks, the versioning of datasets with change detection, and the linking of different feature representations.

Finally, in the context of specific OGC Web services, several previous works have proposed combining them with the Semantic Web. In Yue et al. (2011), the authors describe a proposal to facilitate the management and discovery of services through a semantic search middleware developed for the geospatial catalogue service. Janowicz et al. (2013) propose a Linked Data model and a RESTful proxy for OGC Sensor Observation Services (SOS) to improve integration and interlinkage of observation data. Jones, Kuhn, Keßler, and Scheider (2014) present an adapter that enables access to geographic Linked Open Data (LOD) datasets from within OGC WFS and links them as a data source. Hietanen, Lehto, and Latvala (2016) present a prototype service to provide a geographic names dataset as Linked Data. This proposal makes an on-the-fly transformation process from GML to RDF to provide URIs to spatial objects using existing WFS as source. Finally, Yue, Guo, Zhang, Jiang, and Zhai (2016) explore the integration of Linked Data and Web geoprocessing workflows to complement SDI by discovering geospatial resources in the Web of Data to build geoprocessing workflows. This work describes and links semantically various resources (sensors, observations, services, etc.), which provide an integration example of Semantic Web, services, provenance, and workflows.

Most of the existing proposals in the state-of-the-art have focused on the development of new capabilities or technology extensions (OGC Web services and geospatial semantics). Nevertheless, our proposal is based on the (re)use of standards from both universes (SDI and Semantic Web), without performing changes within them, such that it is possible to combine and exploit multiple and heterogeneous geospatial information paying attention to geometries.
4 | A FRAMEWORK FOR GENERATING, ENRICHING, AND EXPLOITING LINKED DATA FROM GEOSPATIAL DATA SOURCES

In this section, we present the core of our approach for the generation, enrichment, and exploitation of geospatial information using Linked Data principles in combination with SDIs. We have developed a framework, called GeoLOD (https://github.com/jasaavedra/GeoLOD), that gives support to this approach, and which is summarized graphically in Figure 1. Our framework consists of the following five main components: (1) generation, which performs a transformation from geospatial data to RDF; (2) linking, which sets connections with (geospatial) Linked Data and WFS; (3) enrichment, which enables us to enrich Linked Data with attributes collected from WFS; (4) repository, which stores the obtained results after applying our framework components; and (5) exploitation, which allows the display and querying of geospatial data using Linked Data principles in combination with SDIs.

In order to guide the application of our framework, two scenarios have been defined: generation and connection, and exploitation (see Figure 1). These scenarios consider access restrictions and needs related to data. The first scenario uses four components (generation, linking, enrichment, and repository) of our framework. On the other hand, the second scenario of our proposal is associated with exploitation, which is implemented in our framework through visualization and data query.

Next, we briefly describe these components, which are flexible and were tested with a case study in the (geo) biodiversity domain, as a running example, using the workflow shown in Figure 1 and data sources described in Section 2.

4.1 | Generation and connection

In this scenario, transformation, linking, and enrichment components are performed. The applicability of the different components in the workflow depends on two aspects: (1) original data may be transformed and published without restrictions, or (2) there exist some data access limitations, but partial transformation of some relevant attributes may be performed. Next, we provide details associated with the components of this scenario.

4.1.1 | Transforming geospatial data to RDF

A transformation process of geospatial data from Geographic Information Systems (GIS) and SDIs, concretely shapefiles and WFS, to RDF is performed by this component of our framework. We opted for the use of RDF as the normal form for the geospatial datasets to be published, since we want to harmonize different formats of our datasets (databases, shapefiles, Web services, etc.), to avoid using proprietary formats, and because we are pursuing a Linked Data approach. As described in Section 3, RDF is one of the standard languages in which information has to be made available, according to the Linked Data principles. The reason for this is that it offers several advantages, such as the provision of an extensible schema, de-referenceable URIs, and as RDF links are typed, safe merging (linking) of different datasets (Omitolola et al., 2010).

In this transformation process, we recommend exploiting all the advantages of the Linked Data through a complete transformation from original geospatial data to RDF, when there is complete access to data. Also, our approach proposes two different alternatives, described in Sections 4.1.2 and 4.1.3, when data access is limited.

We recommend performing this transformation process by using, reusing, or developing ontologies, although Linked Data can be generated with or without the use of a specific vocabulary, since just transforming data to...
FIGURE 1  A framework for generating, enriching, and exploiting geospatial Linked Data
RDF does not incorporate any semantics, as pointed out by Jain, Hitzler, Yeh, Verma, and Sheth (2010). Moreover, adding ontologies in this process allows making explicit the meaning of concepts in the datasets used, creating a harmonized model for the considered datasets (using common and shared vocabularies) and making it easier to search and access geospatial information (Vilches-Blázquez et al., 2014).

In order to perform the transformation process, our framework contains two elements, called SHP2GeoSPARQL and WFS2GeoSPARQL, which allow transforming shapefiles and WFS into RDF according to GeoSPARQL vocabulary (Perry & Herring, 2012). These tools are based on: (1) the open-source geometry2RDF (https://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/technologies/151-geometry2rdf/index.html) library (Vilches-Blázquez et al., 2010), with notable modifications and substantial enhancements to meet interoperability needs in other geospatial data formats and services; (2) GeoTools (https://www.geotools.org/), which offer Java implementations of OGC specifications (Herring, 2010) for geospatial data management comparable to GIS desktop applications and Web services—its rich API supports feature access to many file formats (like CSV, DXF, GeoJSON, Esri shapefiles, etc.) and spatial DBMSs, as well as coordinate transformations between CRS—and (3) Apache Jena (https://jena.apache.org/), a widely used Java framework for developing Semantic Web applications, tools, and servers.

In our system, both elements (SHP2GeoSPARQL and WFS2GeoSPARQL) work as Web applications, where geometry and spatial relations associated with geographical features are transformed into RDF following the GeoSPARQL vocabulary through different configuration formularies (Web forms). These tools allow uploading ontologies related to a knowledge domain to generate RDF with explicit meaning of concepts in the datasets used (using common and shared vocabularies). Furthermore, SHP2GeoSPARQL and WFS2GeoSPARQL have some optional capabilities, such as: reprojection of geometries into another spatial reference system, set spatial relations between different geographical features, and calculating centroids (only for polygons). These operations are carried out thanks to the integrated GeoTools library and according to user specifications for the source and target operation. An example of the input and output of this transformation process using WFS2GeoSPARQL is shown in Figure 2.

Using the aforementioned elements, we have created RDF data from different data sources. With respect to shapefiles, they come from different national and international bodies associated with the biodiversity domain, such as the Alexander von Humboldt Research Institute, the GBIF, and SIB Colombia. These data sources conform to DwC terms (https://rs.tdwg.org/dwc/terms/) and contain information related to taxonomic classification, scientific name, phylum, kingdom, taxon, coordinates, date, municipality, and so on. On the other hand, WFS are obtained from the Alexander von Humboldt Research Institute (https://www.humboldt.org.co/en). This institute provides access to a catalog of OGC Web services (https://www.humboldt.org.co/es/servicios/infraestructura-institucional-de-datos#geoservicios), where multiple WFS with information about Colombia’s ecosystems, paramos, wetlands, and similar items can be requested.

In all these cases, RDF data are generated, using SHP2GeoSPARQL and WFS2GeoSPARQL, respectively, according to common and shared vocabularies used in each domain, such as the DwC vocabulary, developed by GBIF and used for describing biodiversity resources, GeoSPARQL ontology, Dublin Core Metadata (https://dublincore.org/documents/dcmi-terms/), upper-level DOLCE (https://www.loa.istc.cnr.it/old/DOLCE.html) ontology, and cross-domain DBpedia (https://wiki.dbpedia.org/services-resources/ontology) ontology. Figure 3 depicts a detailed view of the generated RDF through the main elements of the ontologies used, whose details are collected in Listing 6. Both figure and listing show how biodiversity, geographical and geometrical information are integrated for our framework.
**FIGURE 2** Input and output of a WFS2GeoSPARQL transformation
FIGURE 3 A detailed view of the generated RDF
LISTING 6 An RDF excerpt

```r
@prefix geosparql: <https://www.opengis.net/ont/geosparql#> .
@prefix dwc: <https://rs.tdwg.org/dwc/terms/> .
@prefix sf: <https://www.opengis.net/ont/sf#> .
@prefix dcterms: <https://purl.org/dc/terms/> .
@prefix dul: <https://www.ontologydesignpatterns.org/ont/dul/DUL.owl#> .
@prefix DBpedia: <https://dbpedia.org/ontology/> .
@prefix sib: <https://rs.tdwg.org/dwc/iri/sib/> .
@sib:1465972 a dwc:occurrence ;
dcterms:accessRights "Este trabajo está bajo una licencia Creative Commons zero (cc0) 1.0 https://creativecommons.org/publicdomain/zero/1.0/legalcode" ;
dwc:scientificNameAuthorship "" ;
dwc:basisOfRecord "Observacion humana" ;
dwc:class "Aves" ;
dwc:country "Colombia" ;
dwc:decimalLatitude "4.76039"^^<https://www.w3.org/2001/XMLSchema#double> ;
dwc:decimalLongitude "-74.1082"^^<https://www.w3.org/2001/XMLSchema#double> ;
dwc:eventDate "08/07/2006" ;
dwc:family "Ardeidae" ;
dwc:genus "Nycticorax" ;
dwc:institutionCode "Calidris" ;
dwc:kingdom "Animalia" ;
dwc:locality "Humedal la conejera" ;
dwc:occurrenceRemarks "El censo neotropical de aves acuáticas en Colombia: 2002 - 2011" ;
dwc:occurrenceID "1465972" ;
dwc:order "Pelecaniformes" ;
dwc:phylum "Chordata" ;
dcterms:publisher sib:SIB_Colombia ;
dwc:scientificName "Nycticorax nycticorax" ;
dwc:taxonRank "Species" ;
dul:hasLocation sib:Cundinamarca ;
geosparql:defaultGeometry <https://rs.tdwg.org/dwc/iri/sib/Geometry/a3f55c7f5cc41835aabfd-b68509279aa871bb211> .
@sib:Geometry/a3f55c7f5cc41835aabfd68509279aa871bb211 a sf:Point ;
geosparql:asWKT "POINT (−74.10819999999905 4.760390000000029)" ^^^sf:wktLiteral .
@sib:Cundinamarca a DBpedia:Departament .
```

After RDF generation, the system stores them into our repository, an RDF triple store (see Figure 1), where outputs can be queried through a SPARQL Endpoint. In the context of our study case, we have deployed Parliament (Kolas, Emmons, & Dean, 2009), since we have found that it offers an excellent tradeoff between load and query performance and besides the support of the capabilities that GeoSPARQL provides (Battle & Kolas, 2012). This triple store is available at https://ec2-54-94-208-47.sa-east-1.compute.amazonaws.com:8080/parliament/. However, as mentioned previously, there exist other RDF triple stores that support GeoSPARQL, such as Apache Marmotta, OpenLink Virtuoso, Strabon, and USeekM.

4.1.2 | Linking data from geospatial data sources

The fourth Linked Data principle is: "Include links to other URIs, so that they can discover more things." Thus, an increasing number of thematic datasets are published as RDF graphs and linked to other datasets by identifying equivalent resources in other datasets (Felliachi et al., 2013). It brings to light that the value of data and its utility increase when it is more interrelated with other data (Heath & Bizer, 2011). The result of this interlinking process is
often a list of owl:sameAs links between entities of each dataset. These relationships can be discovered using several tools that provide technological support, such as SILK (https://silkframework.org) or LIMES (https://aksw.org/Projects/LIMES.html), which have also started to include some geospatial metrics in the links discovery process. Therefore, we have taken advantage of the benefits of SILK to provide additional advantages to our framework in the linking process. This tool is used for discovering links between generated RDF from geospatial data and other data of the Web of Data.

Thus, in the context of our study case, we have used SILK for interconnecting our generated RDF previously with information about species, habitats, and sites from EUNIS (https://eunis.eea.europa.eu/about) and DBpedia (https://wiki.dbpedia.org/). For that, we have configured SILK using different similarity metrics (such as Jaro, Jaro-Winkler, Levenshtein, and qGrams) and, based on the "scientific name" of species, this discovery process of links is performed. An example of the setting links between both RDF data sources is shown in Listing 7.

**LISTING 7 An RDF instance with a link to EUNIS**

```xml
<https://rs.tdwg.org/dwc/iri/sib/1465972>
  rdf:type dwc:occurrence;
  dwc:eventDate "08/07/2006";
  dwc:scientificName "Nycticorax";
  dwc:family "Ardeidae";
  dwc:genus "Nycticorax";
  dwc:kingdom "Animalia";
  ...
  owl:sameAs <https://eunis.eea.europa.eu/species/1162>;
  owl:sameAs <https://es-la.dbpedia.org/resource/Nycticorax_nycticorax>;
  owl:sameAs <https://dbpedia.org/resource/Black‐crowned_night_heron>.
```

It is important to notice that SILK focuses just on RDF data, sets owl:sameAs links, and deals with different spatial distances (centroid distance and minimum distance between two geometries) and relations (contains, intersects, touches, crosses, etc.). However, within the geospatial information domain, many features are represented by complex geometries and are collected in different types of geospatial resources (formats), for instance, WFS. Our framework takes into account these issues by allowing the establishment of connections between two interoperability universes, such as (geospatial) Linked Data and SDIs.

SDIs often maintain directories of public geospatial Web services built from Web services listed in their registries and play the role of discovery node (Lopez-Pellicer, Renteria-Agualimpia, Nogueras-Iso, ZarazagaSoria, & Muro-Medrano, 2012). Thus, the first step of the workflow starts by discovering WFS in SDI catalogue services (https://www.opengeospatial.org/standards/cat) and/or public lists of geospatial Web services. Some examples of these nodes are Esri Geoportal Server (https://arcgis/2BeR5kS), Data.gov (https://bit.ly/2DhA77), the Global Earth Observing System of System (GEOSS) portal (https://bit.ly/2FWNZFO), or the INSPIRE Geoportal (https://inspire-geoportal.ec.europa.eu/discovery/). Once we have selected the WFS related to our interest domain, the URL of a service is provided to our framework. Such URLs may have been discovered by a user by browsing some of the aforementioned sites or may have been sent to the user by a peer, or may have been crawled automatically.

Given such URLs, our system uses the **WFS2LD_Connector** component, which allows connecting WFS and Linked Data using geometrical information, to try invoking the WFS using the GetCapabilities operation. An example of a URL pattern to invoke WFS is shown in Listing 8 and an excerpt of the output associated with this GetCapabilities operation is shown in Listing 9. This WFS comes from the Alexander von Humboldt Research Institute and contains places for bird conservation.

**LISTING 8 A GetCapabilities request related to WFS used**

```
https://geoservicios.humboldt.org.co/geoserver/Historicos/wfs?&version=2.0.0&request=GetCapabilities
```
LISTING 9 An excerpt of the output related to the GetCapabilities request

```
<wfs:WFS_Capabilities xmlns:xsi="https://www.w3.org/2001/XMLSchema-instance" xmlns="https://www.opengis.net/wfs/2.0" xmlns:wfs="https://www.opengis.net/wfs/2.0" xmlns:ows="https://www.opengis.net/ows/1.1" xmlns:gml="https://www.opengis.net/gml/3.2" xmlns:fes="https://www.opengis.net/fes/2.0" xmlns:xlink="https://www.w3.org/1999/xlink" xmlns:xs="https://www.w3.org/2001/XMLSchema" xmlns:inspire_dls="https://inspire.ec.europa.eu/schemas/inspire_dls/1.0" xmlns:inspire_common="https://inspire.ec.europa.eu/schemas/common/1.0" xmlns:Historicos="https://www.humboldt.org.co/es/servicios/infraestructura-institucional-de-datos" version="2.0.0" xsi:schemaLocation="https://www.opengis.net/wfs/2.0 https://geoservicios.humboldt.org.co:80/geoserver/schemas/wfs/2.0/wfs.xsd https://inspire.ec.europa.eu/schemas/inspire_dls/1.0 https://inspire.ec.europa.eu/schemas/inspire_dls/1.0/inspire_dls.xsd" updateSequence="574" hc="a0" hcx="0">
  <ows:ServiceIdentification>
    <ows:Title>WFS:IAvH:Historicos</ows:Title>
    <ows:Abstract>WFS:Instituto de Investigación de Recursos Biológicos Alexander von Humboldt:Historicos</ows:Abstract>
    <ows:ServiceType>WFS</ows:ServiceType>
    <ows:ServiceTypeVersion>2.0.0</ows:ServiceTypeVersion>
    <ows:Fees/>
    <ows:AccessConstraints/>
  </ows:ServiceIdentification>
  <ows:ServiceProvider>
    <ows:ProviderName>Instituto de Investigación de Recursos Biológicos Alexander von Humboldt</ows:ProviderName>
    <ows:ServiceContact>
      <ows:IndividualName>Grupo I2D</ows:IndividualName>
      <ows:PositionName>Admin I2D</ows:PositionName>
      <ows:ContactInfo>
        <ows:Phone>
          <ows:Voice>(+57) 1 320 2767 Ext. 1160</ows:Voice>
        </ows:Phone>
      </ows:ContactInfo>
    </ows:ServiceContact>
  </ows:ServiceProvider>
  <ows:OperationsMetadata>
    <ows:Operation name="GetCapabilities">
      <ows:DCP>
        <ows:HTTP>
          <ows:Get xlink:href="https://geoservicios.humboldt.org.co:80/geoserver/Historicos/wfs"/>
          <ows:Post xlink:href="https://geoservicios.humboldt.org.co:80/geoserver/Historicos/wfs"/>
        </ows:HTTP>
      </ows:DCP>
    </ows:Operation>
    <ows:Parameter name="AcceptVersions">
      <ows:AllowedValues>
        <ows:Value>1.0.0</ows:Value>
        <ows:Value>1.1.0</ows:Value>
        <ows:Value>2.0.0</ows:Value>
      </ows:AllowedValues>
    </ows:Parameter>
  </ows:OperationsMetadata>
</wfs:WFS_Capabilities>
```
Moreover, the WFS2LD_Connector allows us to return a selection of features from selected WFS, including geometrical and attribute values through the \textit{GetFeature} operation. If the provided URL cannot be invoked successfully (e.g. the service is not available anymore), our system cannot continue without further human intervention. An instance of the \textit{GetFeature} operation is shown in Listing 10, filtered by \texttt{typeName} and \texttt{FeatureID}, and an excerpt of the output associated with this \textit{GetFeature} operation is shown in Listing 11.

**LISTING 10** A \textit{GetFeature} request related to WFS used

```
https://geoservicios.humboldt.org.co/geoserver/Historicos/wfs?&version=2.0.0&request=GetFeature&typeName=Historicos:aicas&featureID=68
```

**LISTING 11** An excerpt of the output related to the \textit{GetFeature} request

```
<wfs:FeatureCollection xmlns:xs="https://www.w3.org/2001/XMLSchema" xmlns:wfs="https://www.opengis.net/wfs/2.0" xmlns:Historicos="https://www.humboldt.org.co/es/servicios/infraestructura-institucional-de-datos" xmlns:gml="https://www.opengis.net/gml/3.2" xmlns:xsi="https://www.w3.org/2001/XMLSchema-instance" numberMatched="1" numberReturned="1" timeStamp="2018-09-18T21:52:41.130Z" xsi:schemaLocation="https://www.opengis.net/wfs/2.0 https://geoservicios.humboldt.org.co:80/geoserver/schemas/wfs/2.0/wfs.xsd https://www.humboldt.org.co/es/servicios/infraestructura-institucional-de-datos https://geoservicios.humboldt.org.co:80/geoserver/Historicos/wfs?service=WFS&version=2.0.0&request=DescribeFeatureType&typeName=Historicos%3Aaicas https://www.opengis.net/gml/3.2 https://geoservicios.humboldt.org.co:80/geoserver/schemas/gml/3.2.1/gml.xsd">
<wfs:member>
  <Historicos:aicas gml:id="aicas.68">
    <Historicos:area>7.61631E8</Historicos:area>
    <Historicos:perimeter>382516.0</Historicos:perimeter>
    <Historicos:codigo_bli>CO079</Historicos:codigo_bli>
    <Historicos:aica>PNN Chingaza</Historicos:aica>
    <Historicos:geom>
      <gml:MultiSurface srsName="urn:ogc:def:crs:EPSG::4686" srsDimension="2">
        <gml:surfaceMember>
          <gml:Polygon>
            <gml:exterior>
              <gml:LinearRing>
                <gml:posList>
                  4.778397 -73.851327 4.778593 -73.851459 4.778823 -73.851591 4.779056 -73.851826 4.779274 -73.851912 4.77955 -73.852032 4.779723 -73.852143 4.779839 -73.852232 4.780046 -73.852342 4.780311 -73.852496 4.780437 -73.852562 4.780633 -73.85266 4.780759 -73.852704 4.780975 -73.852791 4.781073 -73.852849 4.781082 -73.852954 4.780999 -73.852916 4.780962 -73.852048 4.780981 -73.8519 4.78122 -73.851871 4.781358 -73.851937 4.781781 -73.852031 4.781999 -73.852083 4.782332 -73.852248 4.782587 -73.852448 4.782853 -73.852648 4.783105 -73.852734 4.78331 -73.852752 4.78365 -73.852642 4.783861 -73.852421 4.78396 -73.852225 4.784093 -73.852073 4.784172 -73.85206 4.784286 -73.852046 4.784423 -73.852054 4.784515 -73.852098 4.784652 -73.852107 4.784754 -73.852105 4.784843 -73.852023 4.784919 -73.851827 4.78487 -73.851635 4.784968 -73.85145 4.785071 -73.851447 4.785265 -73.851467 4.785505 -73.851529 4.785653 -73.851526 4.785...
```
```
After that, our framework performs a links discovery process, using \textit{WFS2LD} Connector, between gathered data from the WFS and a dataset selected from the LOD cloud. For that, after obtaining the data of the WFS, we have to configure our framework to also gather data of the aforementioned cloud. Thus, our framework requires connection parameters to the SPARQL Endpoint of the selected dataset from the LOD cloud and setting a (Geo) SPARQL query in order to collect existing features and geometries within this dataset. An example of a (Geo) SPARQL query is shown in Listing 12.

\textbf{LISTING 12} An example of a (Geo)SPARQL query

\begin{verbatim}
SELECT DISTINCT ?uri ?wkt WHERE {
  ?uri a <https://rs.tdwg.org/dwc/terms/ocurrence> .
  ?uri <https://www.opengis.net/ont/geosparql#defaultGeometry> ?geo.
  ?geo <https://www.opengis.net/ont/geosparql#asWKT> ?wkt.
}
\end{verbatim}

Once we have collected both geospatial datasets (WFS and RDF), the \textit{WFS2LD} Connector performs the links discovery based on a spatial analysis process, which allows setting relations (e.g. contain, intersect, touch, etc.) between both datasets. Currently, \textit{WFS2LD} Connector needs that geometries are serialized as WKT to carry out this comparison. The output of this process is an RDF file where spatial matching is recorded according to the GeoSPARQL vocabulary. Furthermore, geospatial features of the WFS used are also added to this new RDF file, where pointers to different GetFeature requests are included as URIs. In this way, our system allows us to maintain a connection with the original data (WFS) and obtain these data in GML format, which is very useful for a “traditional” use of this information, for instance, in GIS.

Following the previous example, we performed a spatial comparison between the previously generated RDF data (associated with biodiversity registers) and WFS from the Alexander von Humboldt Research Institute, gathering information on areas of bird conservation in Colombia. In this case, our framework identifies what biodiversity registers are contained within a specific area for bird conservation. When a result of this spatial comparison is set, our framework defines explicitly a simple features topological relation (Perry & Herring, 2012) using, concretely, “Within.” Furthermore, geospatial features of the WFS used are also added to this new RDF file, adding GetFeature requests that are included as URIs. Listing 13 shows an example of the output associated with this process.

Finally, the outputs (RDF files) are stored in a new graph into the RDF triple store associated with our framework (see “Repository” in Figure 1), where results can be queried.

\textbf{LISTING 13} An example of linking process between WFS and RDF dataset

\begin{verbatim}
<https://rs.tdwg.org/dwc/iri/sib/766006>
<https://www.opengis.net/ont/geosparql#sfWithin>
<https://geoservicios.humboldt.org.co/geoserver/Historicos/wfs?version=2.0.0&request=GetFeature&typeName=Historicos:aicas&featureID=68>
\end{verbatim}

\subsection{Enriching Linked Data with WFS}

Any data source often contains implicit references to other data types, and geographic data sources are no exception. These references are more relevant when we deal with expert domain data sources. However, it is often difficult for non-expert users to understand and use these hidden data (Tandy et al., 2017). In this sense, according to Lehmann et al. (2015), the aim of the spatial enrichment process is to retrieve such information and make it explicit.
Our framework carries out a materialization of this viewpoint by means of gathering new data from available WFS, and then these data are used to enrich existing Linked Data (following the workflow presented in Figure 1). In order to achieve this goal, the workflow relies on WFS2LD_Connector to perform the aforementioned spatial analysis process. In the context of our study case, we have invoked the following WFS: https://geoapps.ideam.gov.co:8080/geoserver/Clima/wfs?, which was developed by IDEAM (https://www.ideam.gov.co/geoservicios-institucionales) and contains Lang weather sections (1971–2000); we have also used the previously generated RDF data.

When spatial relations between resources of both data sources (WFS and RDF data) are established, the workflow uses the WFS2LD_Richer component of our framework, which identifies common attributes within both data sources (exact matching of attribute names) and collects geospatial attributes that are different from the WFS through GetFeature requests. The mismatched attributes can be completely transformed to RDF or they can be shown to an end user, who selects those attributes that are transformed to RDF, since there may be many attributes from WFS that are not suitable for enriching the original Linked Data. These new (selected) attributes enrich geospatial information (attributes) of the original Linked Data.

Moreover, WFS2LD_Richer adds a link to the GetFeature operation of the WFS used in order to keep data provenance and connect both data sources. An example of the output in this enrichment process with data from the aforementioned WFS is shown in Listing 14. In this example, a link to the GetFeature operation and two attributes of the WFS are included in the previously generated RDF data. The selected attributes are id and name of the weather classification, which are shown as a representative example of the possibilities of our framework in this enrichment process of the original Linked Data with suitable attributes of WFS.

LISTING 14 An example of enrichment of an RDF dataset with attributes from WFS

```xml
<!-- https://rs.tdwg.org/dwc/iri/sib/715618 -->
<https://datos.clima.co/clima/clasif>
<Superhúmedo> ;
<https://datos.clima.co/clima/id>
<400> ;
<https://www.opengis.net/ont/geosparql#sfWithin>
<https://geoapps.ideam.gov.co:8080/geoserver/Clima/wfs?&version=1.0.0&request=GetFeature&
typeName=Clima:GDBIDEAM.CN2_CLSF_CLM_LANG_1971_2000&featureID=GDBIDEAM.CN2_CLSF_CLM_ LANG_1971_2000.254> .
```

Finally, following the aforementioned workflow (see Figure 1), the obtained outputs (RDF files) are stored in a new graph in the RDF triple store associated with our framework, where results can be queried. Likewise, when RDF data are managed by third parties, we recommend publishing these outputs in the same Linked Data repository where collected.

4.2 | Exploitation

Data are ready to be exploited once they have been transformed, linked, and enriched (see Figure 1). The final goal of connecting and opening these data is that users can use semantics tools and SDI technologies in a coordinated way to search, analyze, visualize, or absorb programmatically the whole data available. Therefore, applications deployment on top of the generated (geo)biodiversity Linked Data is required to exploit these data and provide rich GUIs to users (Vilches-Blázquez et al., 2014). There are several tools available for exploiting geospatial Linked Data through graphical interfaces such as LOD4WFS (https://github.com/jimjonesbr/iod4wfs) (Jones et al., 2014), Sextant (https://sextant.di.uoa.gr/) (Nikolaou et al., 2014), Map4RDF (https://oeg-upm.github.io/map4rdf/) (Llaves, Corcho, & Fernandez-Carrera, 2014), LinkedGeoData Browser (browser.inkedgeodata.org), OsOpenSpace
(https://www.ordnancesurvey.co.uk/business-and-government/products/os-openspace/), or Mappify (https://github.com/GeoKnow/Mappify) (Lehmann et al., 2015). However, it is important to notice that most of these tools create map views and transform Linked Data into map services. Therefore, these services cannot be queried using the SPARQL protocol directly through the mentioned tools.

In this scenario, our system reuses LOD4WFS Adapter, a service that listens to WFS requests and converts these into SPARQL. After the SPARQL query is processed, the LOD4WFS Adapter receives the RDF result set from a triple store, encodes it as a WFS XML document, and returns it to the client (e.g. GIS). Based on this tool, we perform an enhancement in the query processing, since our framework allows setting GeoSPARQL queries on an RDF triple store and showing an updating of the results on WFS automatically. This is another way in which our framework allows setting connections between two interoperability universes (geospatial Linked Data and SDIs).

In order to facilitate access to this connection, our framework has a Web-based application based on LOD4WFS Adapter, Leaflet (https://leafletjs.com/), and Apache Jena libraries. This application allows us to display data published as OGC Web services, concretely WFS, and query Linked Data sources using GeoSPARQL. These GeoSPARQL queries allow the generation of “dynamics” maps, since each interaction with data (query) displays an updated map related to the obtained results of each query.

Moreover, the deployed architecture associated with our framework allows geospatial queries to be performed on data. For instance, the SPARQL query in Listing 15 would get species within the municipality of Funza, with labels associated with their scientific names.

LISTING 15 An example of GeoSPARQL query

```sparql
@prefix geosparql: <https://www.opengis.net/ont/geosparql#> .
@prefix dwc: <https://rs.tdwg.org/dwc/terms/> .
@prefix rdf: <https://www.w3.org/1999/02/22-rdf-syntax-ns#> .
SELECT ?uri ?wkt ?SN
WHERE {GRAPH <https://rs.tdwg.org/dwc/iri/sib/>
  {?uri rdf:type dwc:ocurrence .
   ?uri geosparql:defaultGeometry ?geo2 .
   ?uri dwc:scientificName ?SN.
   ?geo2 geosparql:asWKT ?wkt.
   <https://datos.igac.gov.co/resource/municipio/id/25286>
   geosparql:defaultGeometry ?metry.
   ?metry geosparql:asWKT ?seri.
   FILTER (geof:sfWithin(?wkt, ?seri))}.}
```

Nevertheless, the true potential of our proposal is noticeable when we carry out integration of the published datasets and, hence, connection between both interoperability universes. As an example, a user may combine meteorological (WFS) and biodiversity (RDF) data and exploit the geospatial (GeoSPARQL) component associated with them. This way, the user may retrieve all species located in the study area (i.e. Cundinamarca Department) with their scientific name, their related description within EUNIS (dataset in Linked Data), their association with different climate classification (WFS) present in the study area, and their spatial location (see the SPARQL query in Listing 16 and the obtained results in Figure 4).
LISTING 16 An example of SPARQL query integrating different datasets

```sparql
@prefix geosparql: <https://www.opengis.net/ont/geosparq#> .
@prefix dwc: <https://rs.tdwg.org/dwc/terms/> .
@prefix rdf: <https://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <https://www.w3.org/2002/07/owl#> .
SELECT ?uri ?wkt ?scientificName ?WeatherClass ?EUNIS
WHERE
{GRAPH <https://rs.tdwg.org/dwc/iri/sib/>
{?uri rdf:type dwc:occurrence .
 ?uri geosparql:defaultGeometry ?geo2 .
 ?uri dwc:scientificName ?scientificName .
 ?uri <https://datos.clima.co/clima/clasif> ?WeatherClass .
 ?owl:sameAs ?EUNIS .
 ?geo2 geosparq:asWKT ?wkt .
 <https://datos.igac.gov.co/resource/municipio/id/25286>
 geosparql:defaultGeometry ?metry .
 ?metry geosparq:asWKT ?seri .
FILTER (geof:sfWithin(?wkt, ?seri))}}
```

This query may easily be modified to provide analysis focused on a specific species, associated with a concrete kingdom, or different climate types. In addition, the data can be used by exploiting its geospatial component through geoprocessing and spatial analysis (e.g., measure the evolution of the species in different ecosystems to analyze historical evolution, check climate types and their modifications in different areas, how this affects the species presence, etc.).
The integration of datasets and interoperability universes (WFS and Linked Data), supported by the semantics provided by the links, is one of the main advantages of our proposal. This makes it easier to address problems associated with global biodiversity, which require a cross-disciplinary approach, and allows unlocking data silos facilitating their reuse across organizations and communities.

5  |  DISCUSSION

The need to achieve geospatial data available on the Web in an appropriate way (i.e. [geospatial] resources identified using HTTP URIs, published in such a way that they are indexable by search engines, and connected, or linked, to other resources) is acknowledged (Tandy et al., 2017). However, the combination of SDIs and Linked Data, which can facilitate the reuse and connection of multiple and heterogeneous geospatial information, is still a somewhat isolated approach. In this sense, our work proposes a framework to overcome the current limitations and increase the connection between SDIs and Linked Data.

In terms of generation and connection, the current state of tools for transforming and connecting OGC Web services, and especially WFS, to Linked Data is quite limited. In this sense, our framework provides elements for invoking WFS and transforming them to RDF; however, our proposal is only limited to WFS and does not consider other OGC Web services, although it can serve as a starting point for new efforts. On the other hand, we take into account the key role of geometrical information in order to connect WFS and Linked Data. This allows our framework to discover links between both interoperability universes (WFS and Linked Data, and vice versa) based on spatial relations. This way, the connection is performed by setting relations (e.g. contain, intersect, touch, etc.) between both datasets. In this context, our approach focuses on the GeoSPARQL vocabulary for querying and processing geospatial data and, hence, it is not prepared for handling RDF data according to the WGS84 vocabulary, which is one of the top five vocabularies for describing geospatial Linked Data.

With respect to enrichment issues, our framework allows enriching of geospatial Linked Data by means of collecting new data from WFS. In order to perform this process, our proposal needs to discard common attributes between both data sources; for that, the framework uses an exact matching of attribute names. However, taking into account different, codified and not, sufficiently descriptive attribute names is required to consider different string distance metrics to enhance the identification of common attributes between sources.

Regarding the exploitation component, our framework provides an element that allows setting GeoSPARQL queries on an RDF triple store. The results of each query are displayed on the WFS automatically, which allows the creation of "dynamics" maps in each interaction with data. Moreover, the SPARQL Endpoint of the published triple store (our repository) allows the retrieval and analysis of resource relations in a standardized manner.

Finally, this work provides different benefits because of the connection between two interoperability universes. On the one hand, data silos are unlocked and users may interact with structured, non-proprietary, and harmonized data. On the other hand, “traditional” geospatial data, WFS, and (geo)biodiversity Linked Data are semantically integrated as a unified view; therefore, users may take potential advantage of integration for problems that require connecting SDI services and Linked Data. Besides, these results (data and developed tools) are freely available for the (geospatial) community. In short, these benefits help to facilitate reusing this information, allowing the public to use these data in new and innovative ways, improving the application of geospatial data, and increasing the geo(biodiversity) knowledge available to society (Vilches-Blázquez et al., 2014). However, additional challenges remain in this work for trying to bridge SDI and Linked Data, such as supporting other OGC Web services (e.g. Web coverage service, catalogue services, etc.), performing linking and enrichment processes using queries across diverse data sources (distributed queries) in the Linked Data cloud, and testing our proposal deeply with large volumes of verbose (geometrical) data.
This article presents a framework for generating, linking, enriching, and exploiting geospatial Linked Data from multiple and heterogeneous geospatial data sources. The proposal has allowed us to connect two interoperability universes (SDIs and Semantic Web technologies). This is achieved using open standards from both the geospatial (WFS) and Semantic Web (ontologies and RDF data) domains.

Moreover, the work in this article has linked various datasets in the context of a study case in the (geo)diversity domain. In this context, we deal with geometrical information from WFS and RDF data, since it stands as the key component for connecting geospatial data, allowing that they can be retrieved and interlinked on an unprecedented level in the context of SDIs and Linked Data.

The presented implementation of our framework represents an interesting option aimed to guide the transition from SDIs to best practice for publishing spatial data on the Web (Tandy et al., 2017), that is, the adoption of Linked Data principles in the geospatial data world. We also consider that it may help non-expert users to overcome current problems related to finding, accessing, and using data disseminated through SDIs, and to achieve the goal of finding geospatial data available on the Web in an appropriate way.

Future work will focus on supporting other OGC Web services in order to contribute to the aforementioned transition. We will also test our proposal with large volumes of verbose (geometrical) data and will continue publishing geospatial Linked Data for other related themes and providers (e.g. environmental indicators, climate change, etc.). Furthermore, we plan to improve the exploitation component of our framework by developing hybrid applications capable of combining semantic reasoning, machine learning, and traditional geospatial analysis.

ACKNOWLEDGMENTS

This work was partially sponsored by Colfuturo and Ministerio de Tecnologías de la Información y las Comunicaciones de Colombia. Additionally, we are grateful to Victor Saquicela for his constructive comments and suggestions.

NOTES

1Structured data usually reside in relational databases (RDBMS). Such data are eminently searchable, both with human-generated queries and via algorithms using type of data and field names, such as alphabetical or numeric, currency or date. Source: https://bit.ly/2rdzwI9

2Semi-structured data maintain internal tags and markings that identify separate data elements, enabling information grouping and hierarchies. Both documents and databases can be semi-structured. Source: https://bit.ly/2rdzwI9

3For instance, we have created shapefiles and/or WFS associated with a project, or we can download or access these kinds of files or services without restrictions.

4Note that throughout this article we use the terms vocabulary and ontology without distinction.

5typeName and FeatureID are two parameters (attributes) contained within selected WFS, which can be used for filtering and recovering specific data of this service through the GetFeature operation.

6Datasets that have been published in the Linked Data format (https://lod-cloud.net/).

7Funza is a municipality of the Department of Cundinamarca in Colombia. This municipality is located in the Bogotá Savanna, the southwestern part of the Altiplano Cundiboyacense.

ORCID

Luis M. Vilches-Blázquez http://orcid.org/0000-0001-5799-469X
REFERENCES

Antezana, E., Kuiper, M., & Mironov, V. (2009). Biological knowledge management: The emerging role of the Semantic Web technologies. Briefings in Bioinformatics, 10(4), 392–407.

Athanasiou, S., Giannopoulou, G., Garcia-Rojas, A., & LeGrange, J. (2014). Deliverable 1.2.1: Revised common requirements specification and GeoKnow architecture & system design. Retrieved from http://svn.aksw.org/projects/GeoKnow/Public/D1.2.1_Revised_common_requirements_specification_and_GeoKnow_architecture_system_design.pdf

Auer, S., Lehmann, J., & Hellmann, S. (2009). LinkedGeoData: Adding a spatial dimension to the web of data. In A. Bernstein, D. R. Karger, T. Heath, L. Feigenbaum, D. Maynard, & E. Motta (Eds.), The semantic web – ISWC 2009 (Lecture Notes in Computer Science, Vol. 5823, pp. 731–746). Berlin, Germany: Springer.

Auer, S., Lehmann, J., Ngomo, A., & Zaveri, A. (2013). Introduction to linked data and its lifecycle on the web. In S. Rudolph, G. Gottlob, H. Ian, & F. van Harmelen (Eds.), Reasoning web: Semantic technologies for intelligent data (Lecture Notes in Computer Science, Vol. 8067, pp. 1–90). Berlin, Germany: Springer.

Barnaghi, P., Presser, M., & Moessner, K. (2010). Publishing linked sensor data. In Proceedings of the 3rd International Workshop on Semantic Sensor Networks. Shanghai, China.

Battle, R., & Kolas, D. (2012). Enabling the geospatial semantic web with Parliament and GeoSPARQL. Semantic Web, 3(4), 355–370.

Berners-Lee, T. (2006). Linked data design issues. Retrieved from https://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. Scientific American, 284(5), 28–37.

Bizer, C. (2009). The emerging web of linked data. IEEE Intelligent Systems, 24(5), 87–92.

Blake, J. A., & Bult, C. J. (2006). Beyond the data deluge: Data integration and bio-ontologies. Journal of Biomedical Informatics, 39, 314–320.

Celino, I. (2013). Human computation VGI provenance: Semantic web-based representation and publishing. IEEE Transactions on Geosciences & Remote Sensing, 51(11), 5137–5144.

Chen, H., Yu, T., & Chen, J. Y. (2013). Semantic web meets integrative biology: A survey. Briefings in Bioinformatics, 14, 109–125.

Deans, A. R., Yoder, M. J., & Balhoff, J. P. (2012). Time to change how we describe biodiversity. Trends in Ecology & Evolution, 27, 78–84.

Densham, P. J., & Goodchild, M. F. (1989). Spatial decision support systems: A research agenda. In Proceedings of the GIS/LIS'89 Symposium. Orlando, FL.

Feliachi, A., Abadie, N., Hamdi, F., & Atemezing, G. A. (2013). Interlinking and visualizing linked open data with geospatial reference data. In Proceedings of the 8th International Conference on Ontology Matching (Vol. 1111, pp. 237–238). Sydney, Australia: CEUR.

Giuliani, G., Lacroix, P., Guigoz, Y., Roncella, R., Bigagli, L., Santoro, M., ..., & Lehmann, A. (2017). Bringing GEOSS services into practice: A capacity building resource on spatial data infrastructures. Transactions in GIS, 21(4), 811–824.

Good, B. M., & Wilkinson, M. D. (2006). The life sciences semantic web is full of creeps! Briefings in Bioinformatics, 7, 275–286.

Goodwin, J., Dolbear, C., & Hart, G. (2009). Geographical linked data: The administrative geography of Great Britain on the semantic web. Transactions in GIS, 12(1), 19–30.

Grüttner, R., Purves, R. S., & Wotрубa, L. (2017). Evaluating topological queries in linked data using DBpedia and GeoNames in Switzerland and Scotland. Transactions in GIS, 21(1), 114–133.

Guo, H., Liu, Z., Jiang, H., Wang, C., Liu, J., & Liang, D. (2017). Big Earth data: A new challenge and opportunity for Digital Earth's development. International Journal of Digital Earth, 10(1), 1–12.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. San Rafael, CA: Morgan & Claypool.

Herring, J. R. (Ed.) (2010). Implementation specification for geographic information – Simple feature access. Retrieved from https://portal.opengeospatial.org/files/?artifact_id=25354

Hietanen, E., Lehto, L., & Latvala, P. (2016). Providing geographic datasets as linked data in SDI. ISPRS-International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 41(B2), 583–586.

Jain, P., Hitzler, P., Yeh, P. Z., Verma, K., & Sheth, A. P. (2010). Linked data is merely more data. In Proceedings of the AAAI Spring Symposium on Linked Data Meets Artificial Intelligence. Stanford, CA: AAAI.

Janowicz, K., Bröning, A., Stasch, C., Schade, S., Everding, T., & Llaves, A. (2013). A restful proxy and data model for linked sensor data. International Journal of Digital Earth, 6(3), 233–254.

Janowicz, K., & Hitzler, P. (2012). The Digital Earth as knowledge engine. Semantic Web, 3(3), 213–221.

Janowicz, K., Schade, S., Bröning, A., Keßler, C., Maué, P., & Stasch, C. (2010). Semantic enablement for spatial data infrastructures. Transactions in GIS, 14(2), 111–129.

Jones, J., Kuhn, W., Keßler, C., & Scheider, S. (2014). Making the web of data available via web feature services. In J. Huerta, S. Schade, & C. Granell (Eds.), Connecting a digital Europe through location and place (Lecture Notes in Geoinformation and Cartography, pp. 341–361). Berlin, Germany: Springer.
Kolas, D., Emmons, I., & Dean, M. (2009). Efficient linked-list RDF indexing in Parliament. In A. Fokoue, Y. Guo, & T. Liebig (Eds.), Proceedings of the 5th International Workshop on Scalable Semantic Web Knowledge Base Systems (Vol. 517, pp. 17–32). Washington, D.C.: CEUR.

Koubarakis, M., Kyzirakos, K., Nikolau, C., Garbis, G., Bereta, K., Smeros, P., ..., & Vlachopoulos, I. (2014). Linked earth observation data: The projects TELEIOS and LEO. In Proceedings of the Linking Geospatial Data Conference. London, UK.

Kuhn, W. (2005). Geospatial semantics: Why, of what, and how? Journal on Data Semantics, 3, 1–24.

Kyzirakos, K., Vlachopoulos, I., Savva, D., Manegold, S., & Koubarakis, M. (2014). GeoTriples: A tool for publishing geospatial data as RDF graphs using R2RML mappings. In Proceedings of the 13th International Semantic Web Conference (Vol. 1272, pp. 393–396). Riva del Garda, Italy: CEUR.

Lehmann, J., Athanasiou, S., Both, A., Rojas, A. G., Giannopoulos, G., Hladky, D., ..., & Zaslawski, V. (2015). Managing geoLopez–Pellicer, F. J., Renteria–Agualimpia, W., Nogueras–Iso, J., ZarazagaSoria, F. J., & Muro–Medrano, P. R. (2012). Towards Llaves, A., Corcho, O., & Fernandez–Carrera, A. (2014). Map4rdf–ios: A tool for exploring linked geospatial data. In Kyzirakos, K., Vlachopoulos, I., Savva, D., Manegold, S., & Koubarakis, M. (2014). GeoTriples: A tool for publishing geospatial data as RDF graphs using R2RML mappings. In Proceedings of the 13th International Semantic Web Conference (Vol. 1272, pp. 393–396). Riva del Garda, Italy: CEUR.

Lehmann, J., Athanasiou, S., Both, A., Rojas, A. G., Giannopoulos, G., Hladky, D., ..., & Zaslawski, V. (2015). Managing geoLopez–Pellicer, F. J., Renteria–Agualimpia, W., Nogueras–Iso, J., ZarazagaSoria, F. J., & Muro–Medrano, P. R. (2012). Towards Llaves, A., Corcho, O., & Fernandez–Carrera, A. (2014). Map4rdf–ios: A tool for exploring linked geospatial data. In Kyzirakos, K., Vlachopoulos, I., Savva, D., Manegold, S., & Koubarakis, M. (2014). GeoTriples: A tool for publishing geospatial data as RDF graphs using R2RML mappings. In Proceedings of the 13th International Semantic Web Conference (Vol. 1272, pp. 393–396). Riva del Garda, Italy: CEUR.
Salas, J., & Harth, A. (2011). Finding spatial equivalences across multiple RDF datasets. In *Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web* (pp. 114–126). Bonn, Germany: CEUR.

Salk, C., Sturn, T., See, L., Fritz, S., & Perger, C. (2016). Assessing quality of volunteer crowdsourcing contributions: Lessons from the Cropland Capture game. *International Journal of Digital Earth*, 9(4), 410–426.

Schade, S., & Cox, S. (2010). Linked data in SDI or how GML is not about trees. In *Proceedings of the 13th AGILE Conference on Geographic Information Science*. Guimarães, Portugal: AGILE.

Sheriff, M. A., Dreßler, K., Smeros, P., & Ngomoa, A. C. N. (2017). Radon: Rapid discovery of topological relations. In Smeros, P., & Koubarakis, M. (Eds.). *Discovering spatial and temporal links among RDF data*. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. San Francisco, CA: AAAI.

Sinha, A. (2008). Web Feature Service (WFS). In S. Shekhar & H. Xiong (Eds.), *Encyclopedia of GIS* (pp. 1256–1259). Boston, MA: Springer.

Smeros, P., & Koubarakis, M. (2016). Discovering spatial and temporal links among RDF data. In *Proceedings of the 9th Workshop on Linked Data on the Web*. Montreal, Canada.

Stevens, R. (2002). Ontology-based knowledge representation for bioinformatics. *Briefings in Bioinformatics*, 1, 398–414.

Stucky, B. J., Deck, J., Conlin, T., Ziemba, L., Cellinese, N., & Guralnick, R. (2014). The BiSciCol Triplifier: Bringing biodiversity data to the Semantic Web. *BMC Bioinformatics*, 15(1), 257.

Tandy, J., van den Brink, L., & Barnaghi, P. (Eds.) (2017). *Spatial data on the web: Best practices*. Retrieved from https://www.w3.org/TR/sdw-bp/

Usery, E. L., & Varanka, D. (2012). Design and development of linked data from the national map. *Semantic Web*, 3(4), 371–384.

van den Brink, L., Janssen, P., Quak, W., Stoter, J., & Kadaster, J. (2014). Linking spatial data: Semi-automated conversion of geo-information models and GML data to RDF. *International Journal of Spatial Data Infrastructures Research*, 9, 59–85.

Varanka, D. (2008). National topographic modeling, ontology-driven geographic information in the context of the national map. In *Proceedings of the First International Workshop on Information Semantics and its Implications for Geographical Analysis*. Park City, UT.

Vilches-Blázquez, L. M., Saquicela, V., & Corcho, O. (2012). Interlinking geospatial information in the Web of Data. In J. Gensel, D. Josselin, & D. Vandenbergroucke (Eds.), *Bridging the geographic information sciences* (Lecture Notes in Geoinformation & Cartography, pp. 119–139). Berlin, Germany: Springer.

Vilches-Blázquez, L. M., Villázón-Terrazas, B., Corcho, O., & Gómez-Pérez, A. (2014). Integrating geographical information in the linked digital Earth. *International Journal of Digital Earth*, 7(7), 554–575.

Vilches-Blázquez, L. M., Villázón-Terrazas, B., Saquicela, V., de León, A., Corcho, O., & Gómez-Pérez, A. (2010). GeoLinked data and INSPIRE through an application case. In *Proceedings of the 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 446–449). San Jose, CA: ACM.

Vretanos, P. A. (Ed.) (2010). *OpenGIS® Web Feature Service 2.0 interface standard*. Retrieved from https://portal.opengeospatial.org/files/?artifact_id=39967

Vretanos, P. A. (Ed.) (2014). *OGC® Web Feature Service 2.0 interface standard – with corrigendum*. Retrieved from https://docs.opengeospatial.org/is/09-025r2/09-025r2.html

Walls, R. L., Deck, J., Guralnick, R., Baskau, S., Beaman, R., Blum, S., ..., & Woolry, J. (2014). Semantics in support of biodiversity knowledge discovery: An introduction to the biological collections ontology and related ontologies. *PLoS ONE*, 9(3), e89606.

Weiss, C., Karras, P., & Bernstein, A. (2008). Hexastore: Sextuple indexing for semantic web data management. *Proceedings of the VLDB Endowment*, 1, 1008–1019.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., ..., & Vieglais, D. (2012). Darwin Core: An evolving community developed biodiversity data standard. *PLoS ONE*, 7, e29715.

Wiemann, S., & Bernard, L. (2016). Spatial data fusion in spatial data infrastructures using linked data. *International Journal of Geographical Information Science*, 30(4), 613–636.

Yuan, J., Yue, P., Gong, J., & Zhang, M. (2013). A linked data approach for geospatial data provenance. *IEEE Transactions on Geosciences & Remote Sensing*, 51(11), 5105–5112.

Yue, P., Gong, J., Di, L., He, L., & Wei, Y. (2011). Integrating semantic web technologies and geospatial catalog services for geospatial information discovery and processing in cyberinfrastructure. *GeoInformatica*, 15(2), 273–303.

Yue, P., Guo, X., Zhang, M., Jiang, L., & Zhai, X. (2016). Linked data and SDI: The case on Web geoprocessing workflows. *ISPRS Journal of Photogrammetry & Remote Sensing*, 114, 245–257.

Zhao, P., Foerster, T., & Yue, P. (2012). The geoprocessing web. *Computers & Geosciences*, 47, 3–12.

How to cite this article: Vilches-Blázquez LM, Saavedra J. A framework for connecting two interoperability universes: OGC Web Feature Services and Linked Data. *Transactions in GIS*. 2019;23:22–47. [https://doi.org/10.1111/tgis.12496]