SPARQLing geodesy for cultural heritage – New opportunities for publishing and analysing volunteered Linked (Geo-)Data

Florian THIERY, Timo HOMBURG, Sophie Charlotte SCHMIDT, Jakob VOSS, Germany and Martina TROGNITZ, Austria

Key words: SPARQL, Linked Data, cultural heritage, Wikidata, Ogham

SUMMARY

Geodesists work in Industry 4.0 and Spatial Information Management by using cross linked machines, people and data. Yet, one of the most popular technologies for interlinking data - Semantic Web technologies - have been largely absent from the geodesy community, because of the slow development of standards, a mandatory non-trivial conversion between geospatial features and graph data, and a lack of commonly available GIS tools to achieve this. This is slowly changing due to an increased awareness of the advantages of Linked Data technology in the GIS community and an improvement of standards in the Semantic Web community. Hence, the importance of open source software, open geodata and open access increases. A fundamental requirement for data sharing is the use of standardised data models. In this paper we compare two different modelling approaches for Irish Ogham Stones as a best practice example for linked open data management: One approach uses Wikidata, and the other a custom ontology. While Wikidata offers direct integration into the Linked Open Data Cloud and needs less technological expertise, using a custom ontology enables the creation of best-fitting data models. Both approaches facilitate the use of new information sources for the geodesy community. We aim to demonstrate how Linked Open Geodata can be re-used and further enriched with information from other open sources such as spatial data from OpenStreetMap. For this purpose, we also present a QGIS plugin and a modified geospatial web service, as well as a geo-optimised Linked Data browser, as solutions for bridging the gap between geospatial features and Linked Open Data triples.

SPARQLing geodesy for cultural heritage – New opportunities for publishing and analysing volunteered Linked (Geo-)Data

Florian THIERY, Timo HOMBURG, Sophie Charlotte SCHMIDT, Jakob VOSS, Germany and Martina TROGNITZ, Austria

1. INTRODUCTION

This paper is an update of *SPARQLing Geodesy for Cultural Heritage – New Opportunities for Publishing and Analysing Volunteered Linked (Geo-)Data* (Thiery et al., 2020a) presented at *FIG 2020*[1]. There, we provided an overview on geospatial data modelling and standards, Semantic Web and Linked Open Data (LOD), the Linked Open (Geo-)Data cloud and community-driven projects making use of available data resources such as Wikidata. A use case presented was the *Ogi Ogham Project*, which aims at the integration of information about Ogham stones in Wikidata. Wikidata is an open knowledge hub, in which structured data can be linked to other LOD resources.

For years, using LOD repositories and working with geospatial LOD was not a priority for geodesists, although the geospatial Semantic Web represents a large resource of geospatial information. Various reasons prevented a broad adoption of geospatial Linked Data for geodesy: the LOD data models and the geographical feature model of typical GIS software require a conversion from a graph representation into a relational data representation. Until recently, no tools for this task had been adapted to the expectations of the geospatial community. Although the definition of vector data in Semantic Web standards is complete, the adoption of GIS data into Semantic Web software lags behind even to this day (Jovanovik et al., 2021). The GIS community did not see the need to migrate to a Linked Data approach, yet. A reason is that OGC web services seem to suffice for the daily work of the community and the capabilities of Semantic Web query languages did not support querying for all kinds of geospatial data. With an increased awareness of the advantages of Linked Data technologies, a shift towards using LOD is in progress and new standardisation approaches, as well as vocabularies to represent more types and characteristics of spatial data, are being developed.

In this article, we want to compare two fundamental strategies of dealing with Linked Data, the aforementioned approach of hosting data with Wikidata to one with a custom ontology model where data is hosted on a dedicated triple store. We focus on the modelling of spatial information and the possibilities of interlinking the resulting LOD with other open geodata sources. Comparability of the two approaches is enabled by again using the data on Ogham stones. After a description of Linked Open (Geo-)Data (section 2) we will give a general introduction on Wikidata as a data source and linking hub (section 3), and present the ogham.link platform as a use case example for publishing and analysing volunteered Linked (Geo-)Data (section 4). Solutions for bridging the gap between geospatial features and triples

[1] It was awarded *article of the month* in October 2020.

are suggested in section 5. In the concluding part (section 6) further developments on Linked Data for geodesy as well as cultural heritage and new opportunities for publishing and analysing volunteered Linked (Geo-)Data are summarised.

2. LINKED OPEN (GEO-)DATA

Linked Open Geodata addresses an information gathering problem faced by geodesists and researchers: On the one hand, the world wide web gives geodesists and researchers the possibility of sharing their (geo-)data and thus enable them to participate in scientific discourse. On the other hand, data is often not findable or accessible because it is not hosted in suitable repositories or linked to relevant websites. These unknown data sources can be called 'data dragons', in analogy to depictions of unknown territories in old maps (Thiery et al., 2019). Due to poor interoperability, they are not useful and (re-)usable. A set of techniques, standards, and recommendations can help in overcoming these shortcomings and enable science to open up to anyone: Semantic Web and Linked (Open) Data (Berners-Lee et al., 2001), Linked Open Usable Data (LOUD) (Sanderson, 2019) and the FAIR data principles (Wilkinson et al., 2016), which require that data and its metadata are Findable, Accessible, Interoperable and Re-usable.

The *Semantic Web* is a Web of Data, introduced by Sir Tim Berners-Lee. In the Semantic Web, resources and links are described semantically and can then be accessed using machine-readable interfaces and applications to create a Giant Global Graph: "*The Semantic Web isn't just about putting data on the web. It is about making links so that a person or machine can explore the web of data.*" (Berners-Lee, 2006). The links consist of statements where one object (data, information) is put in a relationship with another one. These three, the subject, the predicate, and the object, are called a triple, which is described using the Resource Description Framework (RDF). In this way, a web of connected data points is created: the Linked Data Cloud. A five-star rating system of openness (Hausenblas and Boram Kim, 2015) was introduced to rate the openness of Linked Data, which then becomes Linked Open Data. "*Linked Open Data (LOD) is Linked Data which is released under an open licence*" (Berners-Lee, 2006). To be usable for scientists and programmers, LOD needs to be accessible via an API and adopt a well-documented ontology that describes the semantic data model for the user community. The acronym to describe this is LOUD - Linked Open Usable Data and has its own set of principles (Sanderson, 2019). The query language used to query for data from the Linked Open Data Cloud (LOD Cloud) is SPARQL (SPARQL Protocol And RDF Query Language). Linked Open Geodata represents a subset of the LOD Cloud (an example from Wikidata is given in Figure 1) and comprises an ever-growing semantically interpreted wealth of geospatial information. Within the LOD world, geospatial data may be described using different vocabularies such as GeoSPARQL (see Figure 2), the Basic Geo (WGS84 lat/long) Vocabulary or custom vocabularies used in several ontologies such as Wikidata, GeoNames or Pleiades. Querying the geospatial Semantic Web is enabled by the GeoSPARQL query language, along with custom extensions provided by different triple store implementations such as Jena Spatial, Blazegraph or Strabon. Currently, any of these standards allows the serialisation of vector data into Well-Known-Text (WKT), GML or into single geocoordinates (double values assigned to lat/long properties). Query languages consequently only consider vector data representations and are not able to query other spatial representations such as rasters and 3D meshes, for which there currently are no standardised serialisations in RDF. Several resources on geospatial data exist, but they have to be consulted and queried individually, which could be streamlined if there was a central information hub. Recently, Wikidata has become a strong candidate to act as such a linking hub.

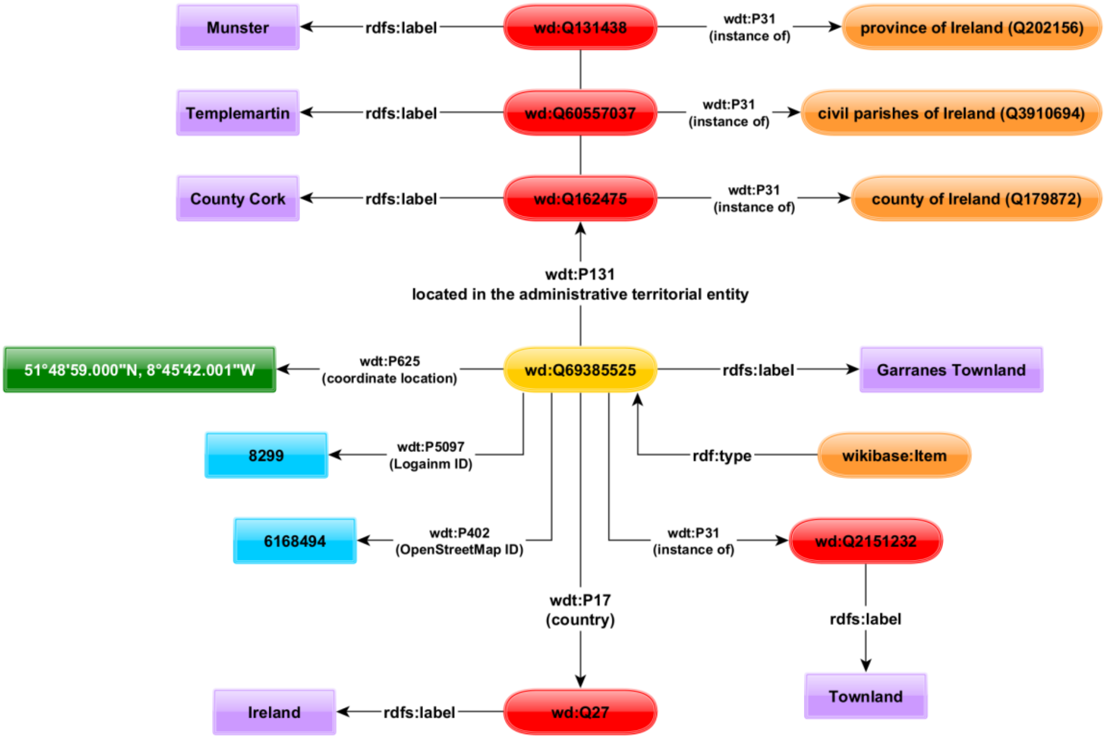

Figure 1. *Linked Open Geodata modelled using Wikidata on the example of the Garranes Townland. (yellow: Townland; orange: classes; red: instances; purple: labels; blue: external ids; green: geo).* **[**Florian Thiery and Timo Homburg, CC BY 4.0, via Wikimedia Commons] More see section 4.

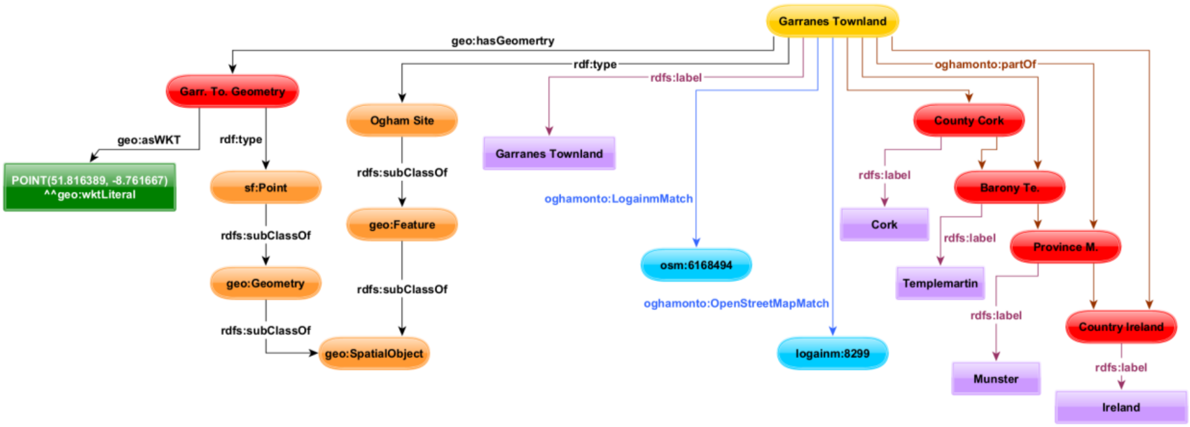

Figure 2. *Linked Open Geodata modeled using the GeoSPARQL vocabulary on the example of the Garranes Townland(yellow: Townland; orange: classes; red: instances; purple: labels; blue: external ids; green: geo).* **[**Florian Thiery and Timo Homburg, CC BY 4.0, via Wikimedia Commons] More see section 4.

3. WIKIDATA AS A DATA SOURCE AND LINKING HUB

The open knowledge base *Wikidata* (Vrandečić and Krötzsch, 2014) is a secondary database and data source for shared knowledge and structured data, established in 2012. It is the central storage for structured data used in the Wikimedia projects, e. g. Wikipedia or Wikimedia Commons. The multilingual data within Wikidata, which can be edited and added by anybody, is available under a free licence (CC0), is accessible to humans and machines, can be exported in standardised formats (e.g. JSON, RDF, CSV), and is interlinked to other open data sets in the LOD Cloud. Wikidata's data model consists of items further described with statements based on properties that are similar to the Semantic Web triples. Additionally, each item in Wikidata can contain links to external identifiers such as ORCID for researchers (Haak et al., 2012), DOI for publications (Paskin, 2010), GeoNames IDs for places and much more. These external identifiers enable the integration of data in Wikidata with other data sources, hence allowing for the use of Wikidata as a linking hub (Neubert, 2017). Section 3.1. provides an overview of geodata and archeological data in Wikidata. Section 3.2 introduces a toolkit aiming at using data in Wikidata without the need to be familiar with SPARQL using a QGIS Plugin for Linked Open (Geo-)Data as an example. Section 3.3 discusses ways of creating and integrating LOD in Wikidata.

3.1 Geodata in Wikidata

Geodata is included in Wikidata using the properties P625 (coordinate location) and P3896 (geoshape). Additionally, Wikidata provides links to other repositories of geospatial data; for example Wikidata QIDs are used in the *Open Street Map* project to uniquely identify points of interests and their attributes. Wikidata also relates properties and classes to Open Street Map by applying the property OpenStreetmap tag or key (P1282).

Geospatial questions are usually linked to a use case. Our example here deals with an archaeological application. Archaeological geodata has already been integrated and interlinked in Wikidata: The Wikidata class archaeological site (Q839954) describes a place (or group of physical sites) in which evidence of past activity is preserved. Ancient places in Wikidata are connected to Pleiades (Simon et al., 2016) – a community-built gazetteer of ancient places – with P1584 (Pleiades ID). Additionally, nearly 4000 spatial data sets from the Samian Research database are available in Wikidata (Thiery et al., 2020b).

Geospatial features within a bounding box or a bounding circle can be queried from Wikidata with a query language extension to SPARQL. While this enables querying for locations, more sophisticated query operations available with GeoSPARQL are not possible.

3.2 The SPARQL Unicorn and the SPARQLing Unicorn QGIS Plugin

Very few databases offer free and openly available and accessible data. Even fewer link into the LOD Cloud. This poses a challenge for comparative analyses of records across multiple datasets. As mentioned above, Wikidata could be an alternative data source and linking hub based on the SPARQL query language. But there is a lack of user-friendly, easy to use, [Free/Libre Open Source Software (FLOSS)] tools, especially for LOD technologies and repositories, including Wikidata itself. The idea of the *SPARQL Unicorn* is to close this gap by providing a set of tools for researchers who work with Wikidata and other related Linked Data repositories. The SPARQL Unicorn aims to help researchers from the humanities or geospatial domain to use the community driven data from Wikidata and to make it accessible to those without expertise in LOD or SPARQL. The *Research Squirrel Engineers* network brings this idea to life and proposes the SPARQL Unicorn principles (Thiery et al., 2020b) to overcome the aforementioned shortcomings.

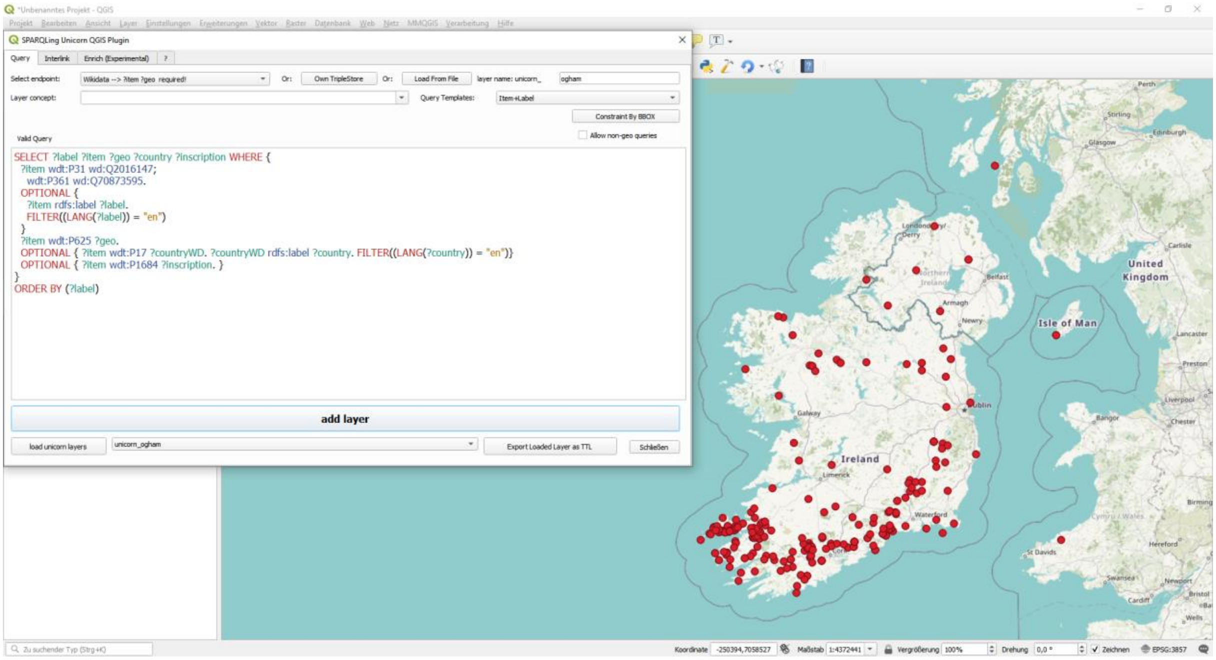

Figure 3. *SPARQLing QGIS Plugin with Wikidata Ogham stones.*
[Florian Thiery and Timo Homburg, CC BY 4.0, via Wikimedia Commons]

The *SPARQLing Unicorn QGIS Plugin* addresses the lack of available tools for working with Linked Open (Geo-)Data. The software is Open Source and accessible via GitHub (Homburg and Thiery, 2020). The plugin allows querying selected triplestores and geo-enabled SPARQL endpoints for data with (Geo)SPARQL. The results can then be further processed in QGIS. The plugin currently offers three functions: (a) simplified querying of LOD data sources, (b) enrichment of geodata and (c) transformation of QGIS vector layers to RDF data (Thiery and Homburg, 2020). Function (a) allows assisted querying of several geo-related triplestores e. g. Wikidata (Figure 3) and Pleiades. In order to assist the user, example queries, a concept search and query templates are given. Function (b) allows the enrichment of a given geodataset using LOD resources, in particular Wikidata (e.g. the elevation level of towns). Function (c) converts geospatial information, e.g. from GeoJSON into RDF, so that this information may be represented for the purpose of publishing geodata as Linked Data. We hope that this tool will enable geodesists to introduce data easily from Wikidata into their own workflows and research.

3.3 Strategies of data integration with Wikidata

The use of Wikidata in general, and in combination with archeological data in particular, can benefit researchers in a variety of ways, especially because data from different knowledge domains may be combined easily. However, should new Linked Data sets be created and included, there are two strategies to store and manage the newly created Linked Data. The first option is described in our previous paper: Create and model the corresponding Linked Data directly in Wikidata (Thiery et al., 2020a). A second option is to create your own database and ontology which is linked to entities in Wikidata. The first option requires the user to publish their data statements under CC0, and Wikidata community rules have to be abided by. When data is primarily managed in Wikidata it can be subject to changes by others. The second option requires knowledge in creating an ontology, the infrastructure to host a linkable database with the given data, and insights into which interlinks should be created. In turn, the user acts as a data curator, i. e. data cannot be modified without consent, and steps have to be taken for appropriate mappings to Wikidata entities and for exposure in search engines and other relevant places. In section 4 we describe how the second option has been implemented using the same data used in our first approach, detailed in the 2020 paper (Thiery et al., 2020a), and discuss the advantages and disadvantages of both approaches.

4. USE CASE: IRISH OGHAM STONES

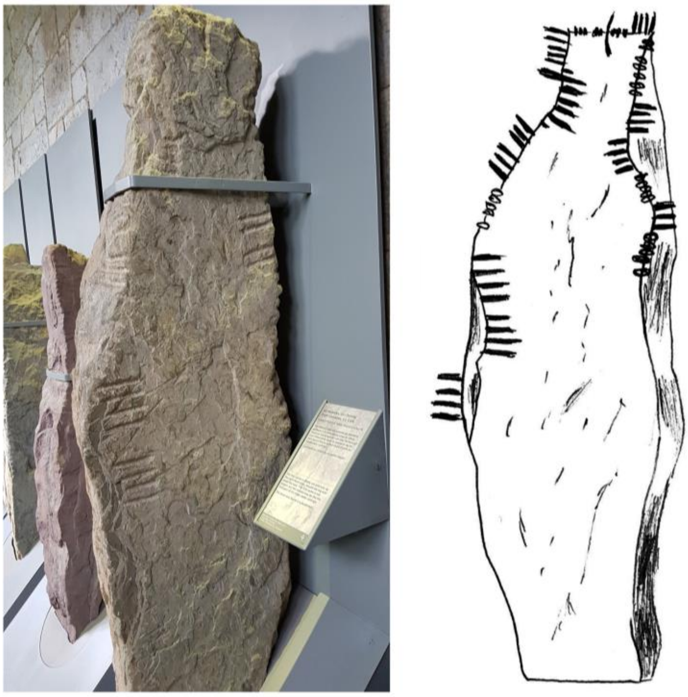

Figure 4. *Ogham Stone CIIC 81, left: stone in the Stone Corridor at University College Cork (UCC); right: drawing of CIIC 81.* **[**Florian Thiery, CC BY 4.0, via Wikimedia Commons]

Cultural Heritage is inherently spatial, and as such well suited to exemplify best practices, which can be transferred to other domains. Thus, we will use the *Ogi Ogham Project* (Thiery and Schmidt, 2020) as a use case for successful LOD modelling and publishing approaches for Linked Open (Geo-)Data. The project's ogham.link platform serves as a Linked Open Data hub for Ogham research and can function as a central node for online Ogham data sources. The project provides LOD information about Ogham stones (as depicted in Figure 4), including their site location, their content, the relationships of the people noted on stones, like tribal affiliations, and other metadata. The main goal is to enable semantic research processing by the scientific community. This can be done by using Wikidata, open Ogham data sources and different semantic modelling strategies. On top of this, the project *Irish ᚑᚌᚆᚐᚋ Stones in the Wikimedia Universe* funded by *Wikimedia Deutschland e.V.* within the *Open Science Fellows Program* will create a semantically described, transparent LOD collection of Irish Ogham stones which are accessible in the spirit of Open Science and knowledge equity. This collection will build on existing published research and can thus serve as an important research tool in the field of early medieval inscriptions. The semantic modelling will be done in two ways. First, as presented in our previous paper (Thiery et al., 2020a), Ogham data was stored in Wikidata in order to include the data in the LOD Cloud and to offer the community the opportunity to participate in free knowledge in the field of Ogham inscriptions. On the other hand, the data is modelled in more detail by using a custom ontology, stored in a custom triplestore available via a SPARQL endpoint and linked to the stones already entered in Wikidata. In addition, the search for Ogham stones on Wikidata and in other Triplestores will be enabled on a community-friendly web platform. Furthermore, integration into free GIS software is to be enabled so that scientists can carry out further analyses in their own software environment (Thiery, 2020). After a discussion on (spatial) data sources about Ogham stones (section 4.1) we will describe the two main concepts in which the data can be modelled as LOD (section 4.2) and compare the two approaches (section 4.3).

4.1 Ogham Data sources and their geospatial information

Several sets of spatial information in analogue and digital sources on Ogham Stones are available. In this section we will detail which sources offer which geospatial information and how we can connect them. The main analogue source – a book – is the catalogue *Corpus Inscriptionum Insularum Celticarum* (CIIC) by (Macálister, 1945). It is the most important reference work for Ogham inscriptions and its numbering schema CIIC has been used to reference stones in other sources as well. Macálister (1945) always gives spatial information about county, barony and townland as headlines in the catalogue. Furthermore, for some of the stones detailed information about the site is mentioned in descriptive texts. The most precise spatial information on the location that is given for all stones in the Macálister catalogue are townlands, the smallest spatial unit in Ireland, which can be compared to the German *Gemarkung*. In our workflow we used this information to link the data to OpenStreetMap with the help of townlands.ie and Logainm – the Placenames Database of Ireland – which also offers LOD interfaces (e. g. for CIIC 81 depicted in Figure 4: An Garrán Townland, Co. Contae Chorcaí; An Garrán/Garranes). Most Ogham stones that were recovered over the centuries can now be found as lintels in huts, or as part of walls in churches, etc., which means their findspot is not the location of their original placement (MacManus, 1997). The townland detail level is therefore sufficient for our purpose and also does not imply that a higher precision is possible with the CIIC data (Figure 5). A main digital source for information about Ogham Stones is the database of the *Celtic Inscribed Stones Project* (CISP) maintained by the University College London. CISP contains numerous inscriptions (not only Ogham) and published its database as an online resource in 2001 without a machine readable access. Fortunately, CISP made their Access Database available to us. In CISP, CIIC 81 is named GARES/1 and refers to the site GARES. That site is encoded by the Irish National Grid Reference as W 472 629 (IR), which can be transformed via available open online tools like the *Irish Grid Reference Finder* to 51.816217 -8.7664640. Another example of open online data is the *Ogham in 3D Project*, which creates 3D scans of accessible Ogham Stones in Ireland. Here, the location is described with a findspot (description and GPS coordinates), the original site and the last known location. Another available open source is the *Historic Environment Viewer* WebGIS, which provides access to the records of the *National Monuments Service Sites and Monuments Record* (SMR) and the *National Inventory of Architectural Heritage*. The Ogham in 3D Project contains National Monuments Service Record Numbers (e. g. CIIC 180: KE053-040----), which are also available in the WebGIS and enable interlinking.

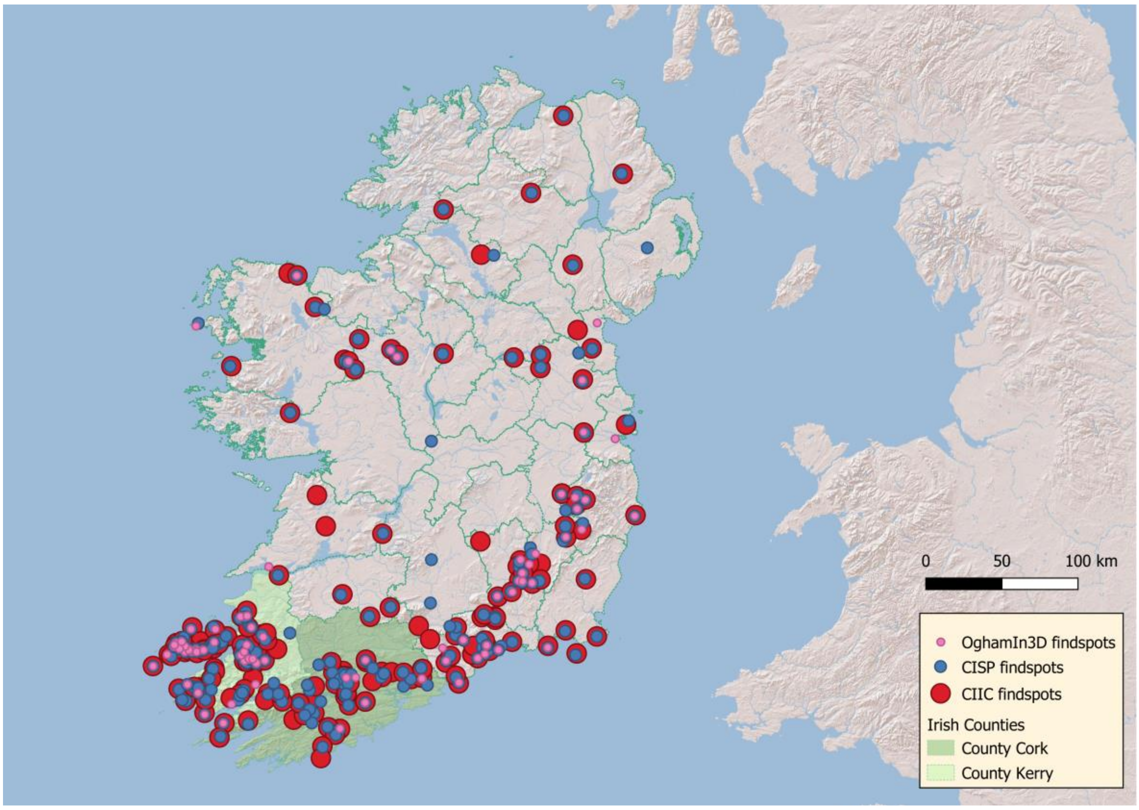

Figure 5. *Find Spots of Ogham Stones in Ireland recorded in the data sources of CIIC, CISP and Ogham in 3D.* **[**Florian Thiery, CC BY 4.0, via Wikimedia Commons]

4.2 Modelling Linked Open Ogham Data

Semantic modelling of the information about Ogham stones as LOD can be done in different ways. One approach presented by (Thiery et al., 2020a) modelled Ogham stones in Wikidata. Another approach presented here uses a custom ontology. Table 1 lists selected attributes of the two data models, which are also visualised in Figures 1 and 2. Wikidata already provides classes for modelling archaeological artefacts such as Ogham stones as well as spatial entities. The community can also create items as subclasses of existing classes, e. g. Ogham Site (Q72617071) as a subclass of archaeological site (Q839954) and geographic location (Q2221906). But properties cannot be added or changed without a discussion within the Wikidata community, whereas, creating a custom ontology allows the creation of classes and relations that best describes one's own data. A custom ontology also allows the creation of classes and relations from scratch, based on well-known ontologies and vocabularies in the Semantic Web to enable interoperability. The current Ogham Ontology in version 0.4 consists of a set of classes which are based on the Web Ontology Language (OWL), RDF Schema (RDFS), CIDOC Conceptual Reference Model (CIDOC-CRM) and GeoSPARQL among others (Thiery, 2021). For provenance information the PROV-O ontology is used. The modelling of vagueness (e. g. certainty of findspots) is done with the Academic Meta Tool (AMT) Ontology (Unold et al., 2019). Each geospatial entity (e. g. a townland or a findspot) is a GeoSPARQL Spatial Object, in order to establish a connection to other spatial LOD repositories or gazetteers and to provide the possibility of creating standardised GeoSPARQL features.

Table 1. *Spatial related classes and properties in Wikidata and the custom ontology*

description	***Wikidata***	***custom ontology***
Ogham Stone	ogham stone (Q2016147)	oghamonto:Stone
Ogham Site	Ogham Site (Q72617071)	oghamonto:OghamSite
Townland	townland (Q2151232)	oghamonto:Townland
individual	instance of (P31)	rdf:type
country	country (P17)	oghamonto:belongsToSpatialThing
location	coordinate location (P625)	geosparql:hasGeometry
findspot	location of discovery (P189)	oghamonto:foundAt
spatially part of	located in the administrative territorial entity (P131)	oghamonto:belongsToSpatialThing
Open Street Map ID	OpenStreetMap relation ID (P402)	oghamonto:OpenStreetMapMatch
Logainm ID	Logainm ID (P5097)	oghamonto:LogainmMatch
bidirectional link between Wikidata and the custom data	exact match (P2888)	oghamonto:exactMatch
node object for modelling vagueness	---	oghamonto:hasAMT

4.3 Comparison of LOD modelling and hosting strategies

When comparing the approaches to model Ogham stones as LOD presented above, strengths and limitations become apparent for both of them. An obvious advantage of creating a custom ontology is that more freedom is given to the ontology creator to semantically model the classes and their relations, as well as the usage of different well-known (de-facto) standard ontologies and vocabularies, e. g. CIDOC CRM, GeoSPARQL, and RDFS. In order to meet a self-set standard of data quality, a custom-made approach may be required since in this way data curation can be allowed solely for a certain user group, e. g. selected domain specialists. However, a shortcoming might be that Linked Data that is too specific might not be suitable to be integrated into existing SPARQL endpoints.

Adding data to Wikidata means data is published openly, using a stable provider with good coverage in the LOD cloud under a CC0 license, while also passing on data sovereignty to the public, hence enabling citizen science and fostering community building and data access via common APIs. As a result, data and semantic modelling sovereignty no longer exist, because anybody can edit the data and the semantic data modelling structure is tied to the community will (e. g. everybody can create new items in Wikidata, but creation of new properties requires a property proposal discussion).

When it comes to modelling provenance information, the Wikidata approach allows adding a reference to the Wikidata statement, while in the custom ontology the PROV-O Provenance Ontology can be used, for example. How (geospatial) data is to be modelled has to be a decision each researcher or geodesist needs to make based on the project's needs, and the advantages and disadvantages of the different approaches touched upon in this discussion.

In conclusion, for our use case we want to use the best of both worlds, which results in a third approach: Using Wikidata and a custom triple store, and adding bidirectional links between them. This allows the addition of updated, enriched and corrected data by Citizen Scientists into the original data set after quality control by the host. At the same time, the use of a custom ontology in the local triple store enables reasoning and extended query capabilities. Figure 6 delineates this hybrid workflow for Ogham stones. The modelling of Linked Open Ogham Data starts with the extraction of data from several sources and storing it as CSV files. After semantic modelling, transformation and importing, the LOD can be further analysed with Open Source Software. In our Ogham use case, we are able to use the CC0 license for the information stored in Wikidata. Data under a different licence (e. g. Ogham in 3D Project) needs to be stored in the self-hosted triple store.

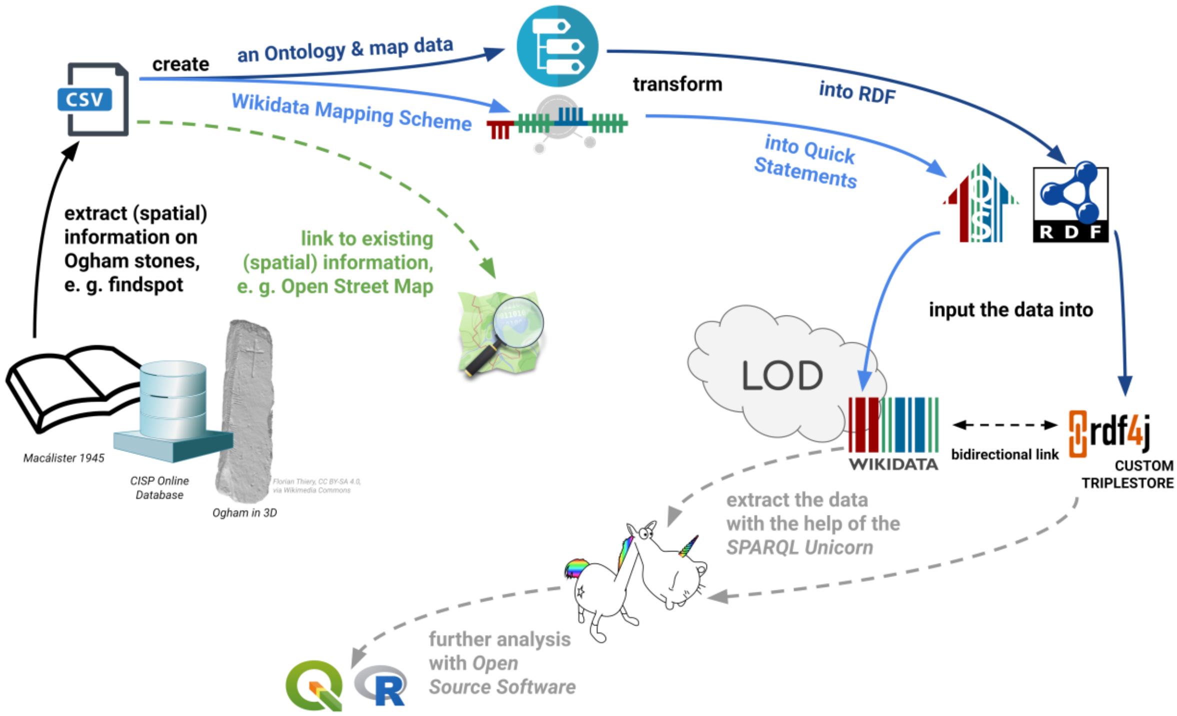

Figure 6. *Linked Open Ogham Data Workflow with Wikidata and a custom triplestore.* **[**Florian Thiery, Timo Homburg, Sophie C. Schmidt and Martina Trognitz, CC BY 4.0, via Wikimedia Commons]

5. BRIDGING THE GAP BETWEEN GEOSPATIAL FEATURES AND TRIPLES

Even if Linked Open Geodata can be provided using (GeoSPARQL-enabled) SPARQL endpoints and may subsequently be accessed by plugins such as the *SPARQLing Unicorn QGIS Plugin*, users are still often required to acquire rudimentary skills in the SPARQL query language to issue more complex queries. In addition, an RDF graph as opposed to a feature collection does not explicitly define what the components of a geospatial feature should be. This definition has to be provided by a specific SPARQL query. This constitutes a barrier for users more familiar with feature-based access of geospatial information and also prevents SPARQL endpoints from becoming established in the geospatial community. Geographic authorities currently need ways of defining geospatial features and feature collections, and exposing them using OGC Web Services that can be well-integrated into every available GIS software. Therefore, in this section we introduce the needed adapter tools to define and to convert predefined parts of RDF graphs to FeatureCollections (so called semantic downlifting).

5.1 Semantic WFS

A core component of every spatial data infrastructure is the provision of *OGC web services* as either Web Feature Services (WFS), Web Map Services (WMS) and Web Coverage Services (WCS) (Vretanos, 2005), or their successors in the OGC API feature family. Current implementations of these web services do not allow the inclusion of SPARQL endpoints as datasource backends. Thus, a new semantics-aware web feature service, the [SemanticWFS](), became necessary. Just as a spatial data infrastructure defines possible SQL queries in order to relate to FeatureCollections represented as SQL tables, SPARQL queries are defined in a SemanticWFS configuration to query a single SPARQL endpoint, or - in the case of federated queries - many at once, in order to combine data from various data sources. The *Ogi Ogham Project* exposes the Ogham stones as a [FeatureCollection]() as shown in Figure 7.

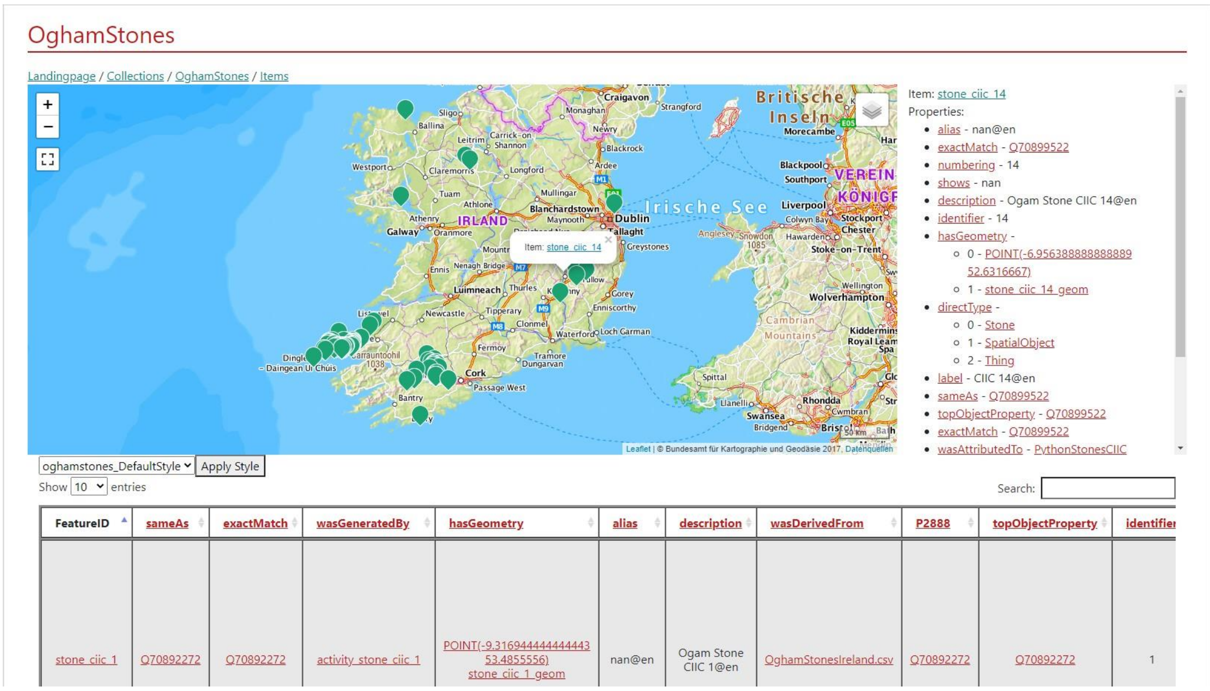

Figure 7. *SemanticWFS FeatureCollection of Ogham Stones.*
[Florian Thiery and Timo Homburg, CC BY 4.0, via [Wikimedia Commons]()]

Because the FeatureCollection is linked to Wikidata instances further information, for example about the location of discovery of the given Ogham stone, is available. Information about Ogham stones could be enriched by modifying the SPARQL query into a federated query which allows querying the local triple store and combines the results with query results from Wikidata. In this way, the SemanticWFS allows geodesists to gather important information from different knowledge bases and to expose them as a FeatureCollection.

5.2 GeoPubby

To make Linked Data accessible to less experienced users and in a way that URIs can be resolved, it is common practice to expose Linked Data using HTML renderers, so-called Linked Data frontends such as *[Pubby]()*. Current Linked Data frontends rarely meet the needs of the geospatial community. In particular, the conversion of geodata into different coordinate reference systems and the export, more precisely the semantic downlift of RDF instances, into geodata formats is a missing feature. *[GeoPubby]()* is a Linked Data frontend derived from [Pubby]() that implements these missing features, and has been implemented by one of the authors of this paper (Bucher et al., 2021, 10). It is openly available on GitHub. In the Ogham use case presented in this publication, GeoPubby helps to resolve and display Ogham stones, including their locations, and provides options to download them in more than 15 different geospatial formats (Figure 8).


CIIC 88 at Ogham Freies Wissen

Property	Value
?:P2888	• <http://www.wikidata.org/entity/Q70885720>
oghamonto:alias	• Ogam Stone 19 (Stone Corridor at UCC) (rdf:langString) (en)
geosparql:asWKT	• <http://www.opengis.net/def/crs/EPSG/0/4326> POINT(-8.071111111111112 52.0269444) (geosparql:wktLiteral)
dc:description	• Ogam Stone CIIC 88 (rdf:langString) (en)
sesame:directType	• oghamonto:Stone • geosparql:SpatialObject • owl:Thing
oghamonto:exactMatch	• <http://www.wikidata.org/entity/Q70885720>
skos:exactMatch	• <http://www.wikidata.org/entity/Q70885720>
geosparql:hasGeometry	• <https://digits.mainzed.org/ogham/pubby/stone_ciic_88_geom>
dc:identifier	• 88 (xsd:string)
rdfs:label	• CIIC 88 (rdf:langString) (en)
oghamonto:numbering	• 88 (xsd:string)
owl:sameAs	• <http://www.wikidata.org/entity/Q70885720>
oghamonto:shows	• BRANAN MAQI OQOLI (xsd:string)
owl:topObjectProperty	• <http://www.wikidata.org/entity/Q70885720>
rdf:type	[14 values]

Figure 8. *GeoPubby Linked Data Browser with information of Ogham stone CIIC 81.*
[Florian Thiery and Timo Homburg, CC BY 4.0, via *Wikimedia Commons*]

6. SUMMARY AND OUTLOOK

In our previous paper (Thiery et al., 2020a) we explored possibilities for publishing and analysing Linked (Geo-)Data in order to SPARQLify geodesy for cultural heritage by using Wikidata as a data repository for Ogham stones. We used the *SPARQLing Unicorn QGIS Plugin* to access and process the Linked Data resources created in this way in well-known geo-related software tools. In this updated paper, we show that the same data can be published as Linked (Geo-)Data with a different strategy: By creating a custom ontology and hosting the connected data on a self-owned server. We compared this approach to the previous one that used Wikidata as a data repository. Additionally, we explored methods of sharing FeatureCollections generated from Linked Data resources using OGC web services that are known and supported by a majority of GIS software, as well as how to make Linked Data browsable, accessible and downloadable in browsers in a variety of GIS formats. This allows geodesists for the first time to conveniently create, publish, modify and re-use Linked Data using GIS tools for geodetic applications. All of the mentioned geospatial software applications are currently under development and have yet to be integrated into common workflows. Furthermore, geospatial standards (e.g. GeoSPARQL, OpenGIS Web Services) need to be extended as prerequisites for the development of new interoperable geo applications based on semantic technologies. The OGC and its *GeoSemanticsDWG Special Interest Group* is currently working towards the creation of these geospatial standards and elaborating current use cases of geospatial Linked Data in various application fields (Abhayaratna et al., 2020). Since September 2020, the *GeoSPARQL Standards Working Group* has been meeting regularly to define GeoSPARQL 2.0, the next version of the standard, which is likely to include further geometry serialisations, a possible integration of coverage data, and support for spatial aggregate functions. The extension of the standard paves the way for more kinds of geospatial data to be supported in the

Semantic Web, and therefore also for the creation of further geospatial web services with Linked Data backends, e. g. OGC Web Coverage Services for raster data or the extension of the *SPARQLing Unicorn QGIS Plugin* and *GeoPubby* to also process coverage formats. Recent publications (Homburg et al., 2020) hint at the feasibility of such implementations. Referring back to the Ogham use case, digital elevation models from the geographic authority of Ireland could then be used to relate geospatial positions to the given Ogham stones, thus enriching geospatial vector data with information gained from raster data.

7. ACKNOWLEDGEMENTS

We would like to thank Dr. Kris Lockyear who made the CISP database available to us. We are also grateful to *Dr. David Wigg-Wolf* for English language corrections. We would like to also thank *Dipl.-Ing. Monika Przybilla* for her support as chair of the DVW working group 1 Profession/Education.

REFERENCES

Abhayaratna, J., van den Brink, L., Car, N., Atkinson, R., Homburg, T., Knibbe, F., McGlinn, K., Wagner, A., Bonduel, M., Holten Rasmussen, M., Thiery, F., 2020. OGC Benefits of Representing Spatial Data Using Semantic and Graph Technologies, OGC White Paper. OGC. http://docs.ogc.org/wp/19-078r1/19-078r1.html

Berners-Lee, T., 2006. Linked Data - Design Issues. URL https://www.w3.org/DesignIssues/LinkedData.html (accessed 31.03.2021).

Berners-Lee, T., Hendler, J., Lassila, O., 2001. The Semantic Web. Sci. Am. 284, 34–43. https://doi.org/10.1038/scientificamerican0501-34

Bucher, B., Folmer, E., Brennan, R., Beek, W., Hbeich, E., Würriehausen, F., Rowland, L., Maturana, R.A., Alvarado, E., Buyle, R., Di Donato, P., 2021. Spatial Linked Data in Europe: Report from Spatial Linked Data Session at Knowledge Graph in Action. October 6th 2020 - On-Line Conference, European Spatial Data Research Official Publication. European Spatial Data Research.

Haak, L.L., Fenner, M., Paglione, L., Pentz, E., Ratner, H., 2012. ORCID: a system to uniquely identify researchers. Learn. Publ. 25, 259–264.

Hausenblas, M., Boram Kim, J.G., 2015. 5-star Open Data. 5 ★ OPEN DATA. URL http://5stardata.info/en/ (accessed 31.03.2021).

Homburg, T., Staab, S., Janke, D., 2020a. GeoSPARQL+: Syntax, Semantics and System for Integrated Querying of Graph, Raster and Vector Data, ISWC. https://doi.org/10.1007/978-3-030-62419-4_15

Homburg, T., Thiery, F., 2020b. SPARQLing Unicorn QGIS Plugin. Zenodo. https://doi.org/10.5281/zenodo.3786814

Jovanovik, M., Homburg, T., Spasić, M., 2021. A GeoSPARQL Compliance Benchmark. ArXiv210206139 Cs.

Macálister, R.A.S., 1949. Corpus inscriptionum insularum Celticarum. Stationery Office, Dublin.

MacManus, D., 1997. A guide to Ogam, Maynooth monographs. An Sagart, Maynooth.

Neubert, J., 2017. Wikidata as a linking hub for knowledge organization systems? Integrating an authority mapping into Wikidata and learning lessons for KOS mappings.

Paskin, N., 2010. Digital object identifier (DOI®) system. Encycl. Libr. Inf. Sci. 3, 1586–1592.

Sanderson, R., 2019. LOUD: Linked Open Usable Data. URL https://linked.art/loud/ (accessed 31.03.2021).

Simon, R., Isaksen, L., Barker, E., de Soto Cañamares, P., 2016. The Pleiades Gazetteer and the Pelagios Project, in: Berman, M.L., Mostern, R., Southall, H. (Eds.), Placing Names: Enriching and Integrating Gazetteers, The Spatial Humanities. Indiana University Press, Bloomington, pp. 97–109.

Thiery, F., 2021. Linked Ogham Ontology. Zenodo. https://doi.org/10.5281/zenodo.4407344

Thiery, F., 2020. Irische ᚑᚌᚆᚐᚋ Steine im Wikimedia Universum. Open Science Framework. https://doi.org/10.17605/OSF.IO/TN5B7

Thiery, F., Homburg, T., 2020. QGIS - A SPARQLing Unicorn? Eine Einführung in Linked Open Geodata zur Integration von RDF in QGIS Plugins. Zenodo. https://doi.org/10.5281/zenodo.3719127

Thiery, F., Homburg, T., Schmidt, S.C., Trognitz, M., Przybilla, M., 2020a. SPARQLing Geodesy for Cultural Heritage – New Opportunities for Publishing and Analysing Volunteered Linked (Geo-)Data, in: FIG Working Week 2020 Proceedings. Fédération Internationale des Géomètres, Kopenhagen, Denmark. https://doi.org/10.5281/zenodo.3751770

Thiery, F., Mees, A., Gottwald, D., 2020b. Linked Open Samian Ware [RGZM/samian-lod: 2020-12-10]. Zenodo. https://doi.org/10.5281/zenodo.4305708

Thiery, F., Schmidt, S.C., 2020. SPARQLing ᚑᚌᚆᚐᚋ Stones: New options for analysing analog editions by digitisation in Wikidata. Zenodo. https://doi.org/10.5281/zenodo.3675284

Thiery, F., Schmidt, S.C., Homburg, T., Trognitz, M., 2020c. The SPARQL Unicorn: An introduction. Zenodo. https://doi.org/10.5281/zenodo.3742185

Thiery, F., Trognitz, M., Gruber, E., Wigg-Wolf, D., 2019. Hic sunt dracones! the modern unknown Data Dragons. Zenodo. https://doi.org/10.5281/zenodo.3345711

Unold, M., Thiery, F., Mees, A., 2019. Academic Meta Tool. Ein Web-Tool zur Modellierung von Vagheit., in: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Eds. Kuczera, A., Wübbena, T., Kollatz, T., Zeitschrift für digitale Geisteswissenschaften / Sonderbände, 4, Wolfenbüttel 2019. https://doi.org/10.17175/sb004_004

Vrandečić, D., Krötzsch, M., 2014. Wikidata: a free collaborative knowledgebase. Commun. ACM 57, 78–85. https://doi.org/10.1145/2629489

Vretanos, P., 2005. Web Feature Service Implementation Specification, Version 1.1.0.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018.

BIOGRAPHICAL NOTES

Florian Thiery is a geodesist and Research Software Engineer working in the Cultural Heritage domain at the Scientific IT department of the *Römisch-Germanisches Zentralmuseum* in Mainz, Germany. Florian is a member of the *DVW* working group 1 *Profession/Education* as well as of the Scientific Committee of the *Computer Applications and Quantitative Methods in Archaeology* (CAA) and fellow of the *Wikimedia Germany Open Science Fellows Program*. *Sophie C. Schmidt* is a computational archaeologist specialised in statistical data analysis using R and currently working at the NFDI4Objects consortial initiative. Sophie is a member of the advisory board of the German speaking CAA chapter and fellow of the Wikimedia Germany *Open Science Fellows Program*. *Timo Homburg* studied Computer Science with emphasis on Computational Linguistics, Semantic Web and Chinese studies and in the last years worked in the GIS field to integrate geospatial data with Semantic Web technologies. Timo's PhD thesis deals with semantic geospatial data integration and the quality of geospatial data in this Semantic Web context. *Jakob Voß* is an information scientist working in research and development at the German library network GBV. *Martina Trognitz* studied Computational Linguistics and Classical Archaeology. She is a member of the advisory board of the German speaking CAA chapter and alumna of the *Open Science Fellows Program*.

CONTACTS

Florian Thiery M.Sc., RGZM, Mainz, GERMANY
ORCID: 0000-0002-3246-3531, Email: mail@fthiery.de, Website: fthiery.de

Timo Homburg M.Sc., i3mainz, Mainz, GERMANY,
ORCID: 0000-0002-9499-5840, Email: timo.homburg@hs-mainz.de

Sophie C. Schmidt M.A., DAI, Cologne, GERMANY
ORCID: 0000-0003-4696-2101, Email: s.c.schmidt@uni-koeln.de

Dr. Jakob Voß, VZG, Göttingen, GERMANY
ORCID: 0000-0002-7613-4123, Email: jakob.voss@gbv.de

Mag. Martina Trognitz, ACDH-CH, Vienna, AUSTRIA
ORCID: 0000-0003-0485-6861, Email: martina.trognitz@oeaw.ac.at