SPARQLing Geodesy for Cultural Heritage – New Opportunities for Publishing and Analysing Volunteered Linked (Geo-)Data

Florian THIERY, Timo HOMBURG, Sophie Charlotte SCHMIDT, Germany, Martina TROGNITZ, Austria and Monika PRZYBILLA, Germany

Key words: sparql, linked data, cultural heritage, wikidata, ogham

SUMMARY

Geodesists are working in Industry 4.0 and Spatial Information Management by using cross linked machines, people and data. Moreover, open source software, open geodata and open access are becoming increasingly important. As part of the *Semantic Web*, *Linked Open Data* (LOD) must be created and published in order to provide free open geodata in interoperable formats. With this semantically structured and standardised data it is easy to implement tools for GIS applications e.g. QGIS. In these days, the world's *Cultural Heritage* (CH) is being destroyed as a result of wars, sea-level rise, floods and other natural disasters by climate change. Several transnational initiatives try to preserve our CH via digitisation initiatives. As best practice for preserving CH data serves the *Ogi Ogam Project* with the aim to show an easy volunteered approach to modelling Irish `Ogam Stones` containing Ogham inscriptions in Wikidata and interlinking them with spatial information in OpenStreetMap and (geo)resources on the web.

ZUSAMMENFASSUNG

Geodäten arbeiten in der Industrie 4.0 mit vernetzten Maschinen, Menschen und Daten. Zudem werden die Themen Open Source Software, offene Geodaten und Open Access immer wichtiger. *Linked Open Data* (LOD) müssen erstellt und veröffentlicht werden, um so freie und offene Geodaten in interoperablen Formaten als Teil des *Semantic Web* zur Verfügung zu stellen. Mit diesen semantisch strukturierten und standardisierten Daten ist es einfach Tools für GIS Applikationen, z.B. QGIS, zu erstellen. Zurzeit werden *Kulturgüter* (CH) der Welt durch Kriege, Meeresspiegelanstieg, Überflutungen und andere Naturkatastrophen, die durch den Klimawandel verursacht wurden, zerstört. Einige transnationale Initiativen versuchen daher die Kulturgüter durch Digitalisierungs-Initiativen zu erhalten. Als Beispiel dient hier die Digitalisierung im *Ogi Ogham Projekt*, das zum Ziel hat, einen einfachen auf Freiwilligenbasis fußenden Ansatz zu verfolgen, der irische `Ogham Steine` und deren Ogham Inschriften in Wikidata modelliert und diese mit Geodaten aus OpenStreetMap und deren Geodaten im Web referenziert.

SPARQLing Geodesy for Cultural Heritage – New Opportunities for Publishing and Analysing Volunteered Linked (Geo-)Data

Florian THIERY, Timo HOMBURG, Sophie Charlotte SCHMIDT, Germany, Martina TROGNITZ, Austria and Monika PRZYBILLA, Germany

1.   INTRODUCTION

Geodetic methods are constantly developing: Traditionally geodetic methods were based on analogue measurements (*Geodesy 1.0*); moving into the digital era, in which digitisation and data publishing in standards on the web speeded up using web mapping platforms like geoserver, leaflet or open layers (*Geodesy 2.0*); into the semantic era, where semantic modelling and publication of Linked (Geo)Data prevail (*Geodesy 3.0*). Today, geodesists record, save and process machine-readable data via the *World Wide Web* (WWW). We are now at the threshold of what we may call the *knowledge era*, in which the machine analyses and creates new knowledge through *Artificial Intelligence* (AI), *Machine Learning* (ML) or semantic reasoning (Wahlster 2017). To fully achieve *Geodesy 4.0*, several challenges must be tackled. Geodesists experience the *Industry 4.0* (Lasi et al. 2014) and Spatial Information Management in everyday work by using linked machines, people and data. Geodata is an important fuel of the digital society, as about 80% of the generated data has spatial contexts (Hahmann & Burghardt 2012). The *OOO model* consisting of *Open Source Software*, *Open (Geo)data* and *Open Access* (Mayer 2016) leads to geodesists working digitally in the cloud using generally accepted standards. The development of the cloud and the web to a *Web 4.0* (Aghaei 2012) includes the publishing of *Linked Data* (LD), *Linked Open Data* (LOD) and *Linked Open Usable Data* (LOUD). Providing free open geodata in interoperable formats creates parts of the *Semantic Web* (Berners-Lee, Hendler and Lassila 2001). Some administrative agencies, as well as community driven volunteered databases, provide geodata as LOD, interlinked with several resources on the web. They are continuously growing as they become enriched with source information and linked to related material in other official databases. A combination of these repositories, as well as databases of different domains, such as natural sciences or *Cultural Heritage* (CH), form a Linked Open Data Cloud creating Geospatial Big Data (Kashyap 2019). If this geodata is semantically structured and standardised it can therefore be easily implemented in tools such as QGIS. The world's CH is constantly in danger of wars, sea-level rise, natural disasters and the impacts of climate change. Therefore, several transnational initiatives try to preserve as much information as possible on CH objects via digitalisation and geospatial analysis, e.g. the Syrian Heritage Archive project (Pütt 2018), the Ogham in 3D project (Bennett, Devlin and Harrington 2016) or the documentation of the Rock Art in Alta (Tansem and Johansen 2008). In addition, volunteered databases allow everyone to easily assert digital CH items as LOD. Thereby, the digital CH cloud will grow in the upcoming years, which makes it crucial for the geospatial community to take an active role in the technical (geodetic) evolution in order to reach *Geodesy 4.0*.

In this paper we consider steps and workflows needed to reach one prerequisite for *Geodesy 4.0*: Implement Linked Open Data standards of the semantic era (*Geodesy 3.0*) for CH data.

Starting with geospatial data modelling standards (cf. section 2) we will give a general introduction into the concept of Linked Data (cf. section 3), followed by the most common Linked Geodata ontologies (cf. section 4), an insight into LOD (geo)datasets (cf. section 5) and the idea of Wikidata (cf. section 6). Next, we introduce the famous SPARQL unicorn (cf. section 7), give examples of Linked Geodata in action (cf. section 8) and showcase a best practice example using Wikidata and Linked Data in the *Ogi Ogham Project* (cf. section 9).

2. GEOSPATIAL DATA MODELLING AND STANDARDS

Geodesists invented a lot of standards to exchange their geospatial data, e.g. EPSG codes for projections. Starting in the *Geodesy 2.0* era, standards for digital data modelling were created and applied to enable interoperability, data exchange and reusability, e.g. *GML* (Portele 2007), *OGC web services*, *GeoJSON*, *GeoSPARQL* and *Neo4J spatial functions* (Agoub, Kunde and Kada 2016). The *OGC* provides a variety of web service definitions, in order to provide, process and display geospatial data to the Geospatial community. WFS Services (Vretanos 2005) give access to vector datasets, WCS Services (Benedict 2005) enable the download of raster data, WMS Services (Wenjue, Yumin and Jianya 2004) offer pre-rendered maps and CSW Services (Nogueras-Iso et al. 2005) provide an overview of different available aforementioned web service types. A common problem in the geospatial sciences is that those services are not interlinked among each other, except for the CSW services. In particular, links inside datasets to other datasets are not possible, only external links to an entire dataset can be provided. *GeoJSON* is a community-driven data format that displays vector data which emerged in 2008 from the need to create a simple JSON-based (Severance 2012) format for sharing geospatial data on the web (Butler et al. 2016). *GeoJSON* became a de-facto web standard which is today often used as a means of geospatial data provision for web applications such as Leaflet or JavaScript-based frameworks, or as a common return type in OGC web services. This standard defines geospatial features and Feature Collections whereas a feature is comprised out of a geometrical part which includes geo-coordinates, the geometry type and a list of key/value pairs describing the properties of the respective feature. Several extensions have been proposed for GeoJSON, such as GeoJSON-LD for linked data and GeoJSON-T for temporal aspects. Recently, the CoverageJSON format has been standardised to represent coverages and their annotations in JSON. The *GeoSPARQL* standard (Battle and Kolas 2012) defines both a vocabulary to encode geospatial features and a query extension to the SPARQL query language (Prud'hommeaux and Seaborne 2008) allowing the definition of geospatial relations.

3. LINKED (OPEN USABLE) DATA

Wuttke (2019) shows that areas which are unknown to the map creator were described in ancient times by the phrase `*Hic sunt dracones*` (engl. *here be dragons*). Today the web gives geodesists the possibility of sharing their geodata and enables them to participate in the scientific and political discourse. However, much of this shared data is not findable or accessible, thus resulting in modern *unknown data dragons*. Often these *data dragons* lack connections to other datasets, i.e. they are not interoperable and can therefore lack usefulness, reusability or usability. To overcome these shortcomings, a set of techniques, standards and recommendations can be used: Semantic Web and Linked (Open) Data, the FAIR principles

(Wilkinson et al. 2016) and LOUD data. Tim Berners-Lee introduced the concept of *Semantic Web*, by using the ideas of Open Data, semantically described resources and links, as well as usable (machine readable) interfaces and applications for creating a *Giant Global Graph* (Thiery et al. 2019). "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). A five star rating system of openness (Hausenblas and Boram Kim 2015) was introduced to rate Linked Data, i. e. "Linked Open Data (LOD) is Linked Data which is released under an open licence." (Berners-Lee 2006). Furthermore, LOD must be *usable* for scientists and programmers in order to take full advantage of all the LOD power. Following the LOUD principles (Sanderson 2019) will make LOD even more FAIR.

4. LINKED OPEN GEODATA ONTOLOGIES

The Linked Open Data principles mentioned in section 3 are applied in several Linked Open Data projects across all domains, e.g. geodesy, humanities and natural sciences. The *WGS84 Geo Positioning RDF vocabulary* (GEO) is a lightweight common used LOD vocabulary representing latitude, longitude and altitude information in the WGS84 geodetic reference datum (Atemezing et al. 2013). The GEO vocabulary is used e.g. in the nomisma project as a Linked Data hub for ancient coins (Gruber 2018). The *GeoSPARQL ontology* (cf. section 2) defines the concepts of a spatial object which is broken down into a feature part describing its semantic meaning and a geometric part. The geometric part includes serialisations of the respective geometry as literal descriptions in either WKT or GML, providing a class hierarchy of GML and WKT Geometry concepts respectively. Properties of the respective geospatial entity are annotated at instances of the Feature class which is linked to the geometrical representation. GeoSPARQL is used in projects such as LinkedGeoData and the SemGIS project (section 8.1).

5. LINKED OPEN GEODATA

The Linked Open Data Cloud offers large data repositories which can be used by different communities, for various purposes. The strength of *Linked Open Data* (LOD) is the linking of information from a wide variety of decentrally hosted knowledge domains. For the geoinformatics domain, community-based data repositories published their data. Moreover, gazetteer repositories and administrative providers also offer their geodata as LOD. *GeoNames* (Hahmann and Burghardt 2010; Khayari and Banzet 2019) aims to be the first geospatial Linked Data gazetteer by linking geographical names to geo coordinates to facilitate geocoding and the usage of geographical places in other Semantic Web contexts. In a joint project between *Ordnance Survey Ireland (OSi)* and ADAPT research centre at Trinity College Dublin, Ireland's geospatial information has been made available as Linked Data on a dedicated portal (Debruyne et al. 2017). The Placenames Database of Ireland, Bunachar Logainmneacha na hÉireann (*Logainm*), is a management system for research conducted by the State. It was made publicly available as Linked Open Data for Irish people at home and abroad, and for all those who appreciate the rich heritage of Irish placenames (Lopes et al. 2014). The *Ordnance Survey (OS)* offers several British datasets as geospatial data (Goodwin, Dolbear and Hart 2008; Shadbolt et al. 2012). OS has published the 1:50 000 Scale Gazetteer, Code-Point Open and the

administrative geography for Great Britain, taken from Boundary Lines. *LinkedGeodata.org* (Stadler et al. 2012) created an ontological model for OpenStreetMap geospatial concepts, which in OpenStreetMap may be defined as tags or key/value combinations of tags. This allowed the Linked Data community to access a repository of geospatial data, while at the same time opening the semantic concepts of a Volunteered Geographic Information (VGI) world map to the Semantic Web community for further analysis. *Pleiades* (Simon et al. 2016), similar to GeoNames, created a gazetteer of geographical names for ancient places to allow historical researchers to link their findings to a unique identifier, indicating a place in time of historical significance.

6. WIKIDATA

Wikidata (Vrandečić and Krötzsch 2014) is a secondary database for structured data, established in 2012. It is a free and open knowledge base where anybody can add and edit data. It is the central storage for structured data of Wikimedia projects, e.g. Wikipedia and Wiktionary. Data held within Wikidata is available under a free licence (CC0), it is multilingual, accessible to humans and machines (GUI, API, SPARQL), exportable using standard formats (e.g. JSON, RDF, SPARQL) and interlinked to other open data sets in the Linked Data Cloud. Wikidata's data model contains items (e.g. label, description, alias, identifier) and statements (e.g. property, value, qualifier, reference), cf. Trognitz and Thiery (2019). The Open Science Fellows Program is aimed at researchers who want to promote their research in an open manner, an example being Martina Trognitz in *A Linked and Open Bibliography for Aegean Glyptic in the Bronze Age*.

7. SPARQL UNICORN

In humanities and geospatial related research documentation, databases and their analyses play a central role. Some of these databases are available as online resources. However, very few are made openly available and accessible and even less are linked into the *Linked Open Data Cloud*. This hinders comparative analyses of records across multiple datasets. Nevertheless, there is one database that has been around since 2012 and recently gained momentum: Wikidata (cf. section 6). We would like to propose the *SPARQL unicorn* as a friendly tool series for researchers working with *Wikidata*. The unicorn's aim is to help researchers in using the community driven data from Wikidata and make it accessible to them without expertise in LOD or SPARQL (Trognitz and Thiery 2019). One existing implementation of the *SPARQL unicorn* is the *SPARQL unicorn QGIS Plugin*, cf. section 8.2. Another implementation using the unicorn for combining SPARQL and R for statistical analysis is currently under development. First results are visible in section 9.2.

8. LINKED OPEN GEODATA IN ACTION

Linked Open Data and Linked Geodata are not only theoretical concepts. The data in the Linked Data Cloud as part of the Semantic Web is used in several projects to help the scientific and geo-community to address their challenges using Linked Data techniques. The following sections will describe two research projects dealing with applied Linked Data.
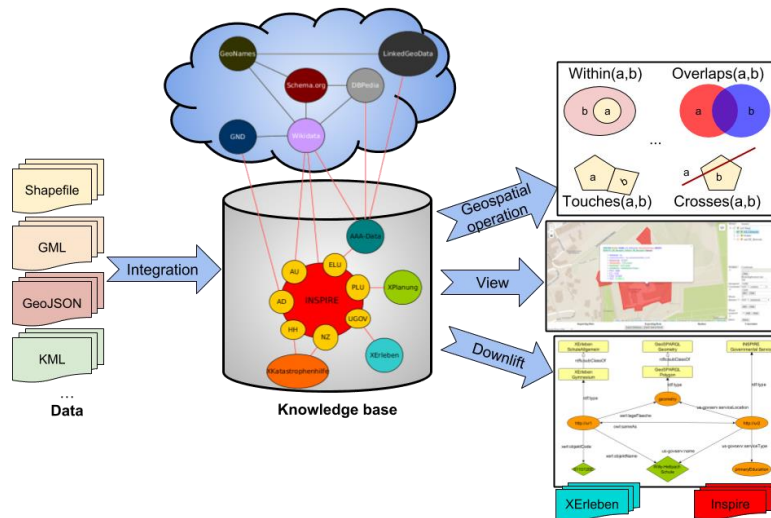
8.1.   SemGIS project

figure 1: Overview of a Semantic GIS System, heterogeneous geospatial data is integrated into an ontological structure, the so-called knowledge base which is in turn interlinked to the Linked Open Data Cloud. The integrated system allows for queries downlifts in other geospatial data formats and may provide views for parts of geospatial data. (CC BY 4.0 Timo Homburg, Claire Prudhomme)

The *SemGIS project* was a research project conducted by Mainz University of Applied Sciences which aimed at finding methods to integrate GIS data into a semantic context for the purpose of data integration, interlinking, reasoning and finally data application. To that end so-called Semantic Uplift and Downlift methods have been developed, which allow for the conversion and semantic enrichment of geospatial data in heterogeneous formats. Semantic Uplifts may be performed on data without a given schema description (Prudhomme et al. 2019), on common geospatial formats using pre-extracted ontologies and an automated converter (Würriehausen, Homburg and Müller 2016) using mapping schemas on databases. The *GeoSPARQL* query language has been thoroughly investigated to come up with proposals on how to extend the language to cope with coverage data, geometry manipulations in-query and the handling of further geospatial data formats. Such proposals are currently discussed in the *OGC GeoSemanticsDWG special interest group* for standardisation. Finally, Downlift approaches, which result in making SPARQL accessible by means of traditional GIS web services, are currently applied in a pilot study at the German Federal Agency for Cartography and Geodesy to pioneer a linked data powered spatial data infrastructure which is interlinked to other governmental and *VGI data sources*, cf. figure 1. Application cases tackled by the SemGIS project, include the assessment of disasters, specifically the simulation of floods and action response systems supporting crisis management. Here, information of different sources needs to be acquired, combined and finally evaluated which was accomplished using reasoning rules. For example: A rising flood level would trigger a change in the ontological model which would in turn trigger corresponding rescue units to respond in an appropriate manner. In this way, semantics support a real-world application case which could only be realized using considerable efforts using traditional GIS integration methods of *Geodesy 2.0*.

8.2.   SPARQL Unicorn QGIS Plugin

Sections 5 and 6 give an insight into community-based data repositories that may be used by geodata domain experts, such as *Wikipedia* or *LinkedGeoData*. Furthermore, gazetteer repositories e.g. *GeoNames* or *Pleiades*, are publishing their (ancient) spatial data as LOD. Moreover, administrative providers like the *OS* or *OSi* model provide geospatial data, containing linked information, into the Linked Open Data Cloud. Unfortunately, all these LOD resources have become of minor importance in the geo-community. This is due to a lack of support for GIS applications in processing LOD. *Triplestores* and SPARQL are currently not supported by GIS software, *GeoServer* implementations or OGC services. The Linked Data serialization *GeoJSON-LD* poses challenges due to some outstanding issues but is not often used in applications like its *unsemantic sister GeoJSON*. This is exactly where the *SPARQLing Unicorn QGIS plugin* comes into play. The plugin enables the execution of Linked Data requests in (Geo-)SPARQL to selected triplestores and geospatial capable SPARQL endpoints. The results are converted into GeoJSON layers, so that they can be used directly in QGIS. In the future, the *SPARQLing Unicorn plugin* will offer users the possibility to automatically generate simple queries - out of extracted concepts of selected ontologies - such as `Give me all cultural heritage sites in BOUNDINGBOX with directly connected relations` and thus make loading more dynamic content of data repositories possible. It is desired that the geo community takes an active part in the (further) development of the plugin, thus making the world of LOD known in the geo context. The source code is freely available for forking on GitHub.

9.   THE OGI OGHAM PROJECT

Stones carrying Ogham inscriptions are found in Ireland and the western part of Britain (Wales and Scotland). Ogham stones mainly served as memorials and/or boundary markers as well as indicators of land ownership and contained relationships as well as personal attributes. They date from the 4th century AD to the 9th century AD (MacManus 1997).

figure 2: left: Ogham Stones - CIIC 81 at University College Cork (UCC) (CC BY 4.0 Florian Thiery via Wikimedia Commons), middle: CIIC 180 as 3D view using MeshLab (CC BY-NC-SA 3.0 Ireland http://www.celt.dias.ie), right: CIIC 180 (Macalister 1945:173) carrying the inscription BRUSCCOS MAQQI CALIACI (ᚈᚂᚋᚑᚄᚉᚉᚒᚏᚁ — ᚋᚐᚊᚊᚔ — ᚉᚐᚂᚔᚐᚉᚔ)

One of the largest publicly available collections of Ogham stones is in the Stone Corridor at University College Cork (cf. figure 2, l.). Probably the most complete standard reference is found in Macalister (1945, 1949), who established the CIIC scheme. The *Ogham in 3D project* currently scans Irish Ogham stones and provides the data, metadata and 3D models (cf. figure 2, m.) for the community. Ogham inscriptions contain *formula words* like *MAQI* (�-ⅢⅢ⋯; son, e.g. figure 2, r.) or *MUCOI* (�; tribe/sept). The Irish personal name nomenclature reveals details of early Gaelic society, e.g. *CUNA* (ⅢⅢ⋯ⅢⅢ⁺; wolf/hound) or *CATTU* (ⅢⅢⁱⅢⅢ⋯; battle), details in Thiery (2020) and MacManus (1997).

The idea of the *Ogi Ogham Project* is to provide the Ogham stones, their content, the relationships of the people noted on stones, their tribal affiliations and other metadata as Linked Open Data; thus enabling semantic research processing by the scientific community. The project group creates a semantic dictionary for Ogham, which is done by a dynamical extraction from text sources using natural language processing methods of keyword extraction. The relevant keywords were collected from the literature Thiery (2020). Linked Ogham Stones allow the following research questions to be addressed by linking knowledge and enriching it: (i) classification of stones (e.g. family hierarchy) and (ii) visualisation of relationships in maps generated by LOD. As a fundament for the analyses, we rely on a Wikidata retro-digitisation of the CIIC Corpus by Macalister (1945, 1949), EPIDOC data of the Ogham in 3D project and on the Celtic Inscribed Stones Project (CISP) database (Lockyear 2000). Furthermore, we actively maintain missing and suitable elements in *Wikidata* (cf. section 9.1) to provide the data to the research community in the sense of the *SPARQL Unicorn* (cf. section 7).

9.1. Ogham data modelling in Wikidata

For inserting, publishing and maintaining Wikidata's data the software OpenRefine is recommended (Association of Research Libraries 2019). First, the data will be imported via CSV. Second, an open refine model for mapping the CSV import files has to be created. Third, a Wikidata mapping scheme model for maintaining the entities needs to be established. In the *Ogi Ogham Project* it is done in Thiery and Schmidt (2020a) for townlands and in Thiery and Schmidt (2020b) for the Ogham stones. In this paper, we would like to focus on the townland modelling from old textual resources, as well as from database entries which rely on outdated text sources. Drawing on Macálisters *Corpus Inscriptionum Insularum Celticarum* (1945, 1949) enabled a geospatial placement of the Ogham stones on the level of townlands. A townland (Irish *baile fearainn)* is a small geographical classificatory unit in Ireland and of Celtic origins, though their boundaries, names and locations may shift over time. Macálister's catalogue is ordered by county, barony and townland, therefore this information was used to identify the modern townland to which to link the Ogham stone. The first resources for comparison were: townlands.ie (based on OSM) and logainm.ie (cf. section 5). Several problems arose during this process. They can be classified as:

- locations unknown to Macálister
- mistakes made by Macálister (typographical errors, wrong place names)
- imprecise information given by Macálister
  - the occurrence of several townlands with the same name in this barony
  - not giving precise names, such as leaving out `upper` or `lower`
  - not providing a townland, but giving a town or electoral division
- a shift in the structure of baronies, electoral divisions and townlands between 1945 and 2020

In many cases, it was not possible to determine which of the above was the problem. Whether there was a shift in the naming of the townland or whether Macálister made such a grievous typographical error that one could not reconstruct the name led to the same result: The townland could not be identified. This is a common problem. The National Monuments Service of Ireland holds a [database of archaeological finds](#) uploaded by the Department of Culture, Heritage and the Gaeltacht, with which we could check our information and which also has a number of unknown locations registered. Nonetheless, in this database a few decisions had been made by local experts to which we adhere (16 times; e.g. for CIIC 204 we followed their advice that Macálister made an error in naming Curraghmore West instead of East). Logainm was also helpful, as a townland given there was linked to a monument, which was used as localisation by Macálister. Further information given by Macálister proved to be invaluable: In seven cases of imprecise place names given in the catalogue, we could use his additional description to improve the precision of the spatial data. For example, Macálister elaborated that the stone CIIC 54 was built into the cathedral of the town, or that a stone was found south of the village (CIIC 48), which enabled us to choose a very probable townland. On a few occasions, we resorted to using the larger of the two townlands with the same name, if they were located right next to each other, verifying this educated guess with the help of the Department of Culture, Heritage and the Gaeltacht. All in all, we managed to locate 185 of 196 townlands mentioned by Macálister. We enriched the data set with information such as: the name of the townland, the Gaelic name of the townland (alias), it's province, county, barony, civil parish, the electoral division it belongs to, a point coordinate, OSM ID, logainm ID, OSi GeoHive IDs, as well as the link to the townlands.ie, from which we derived most of the data. To be able to map the remaining 11 Ogham stones, we chose the centre of the barony given by Macálister.

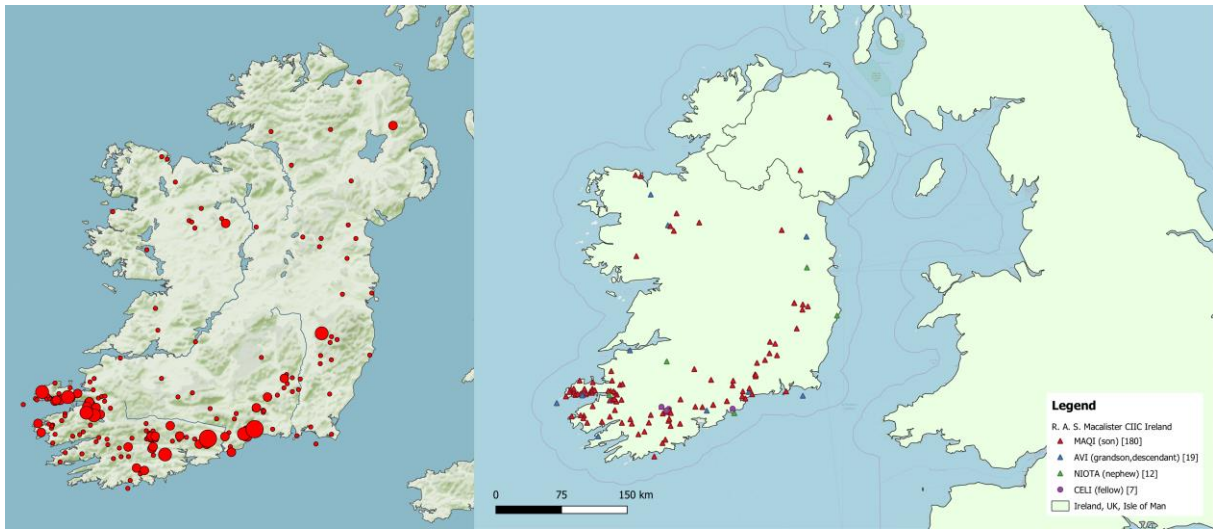## 9.2.   Analysis Ogham data using LOD Plugins

*figure 3: left: Ogham stones in Ireland (CC BY 4.0 Katja Hölzl, RGZM),
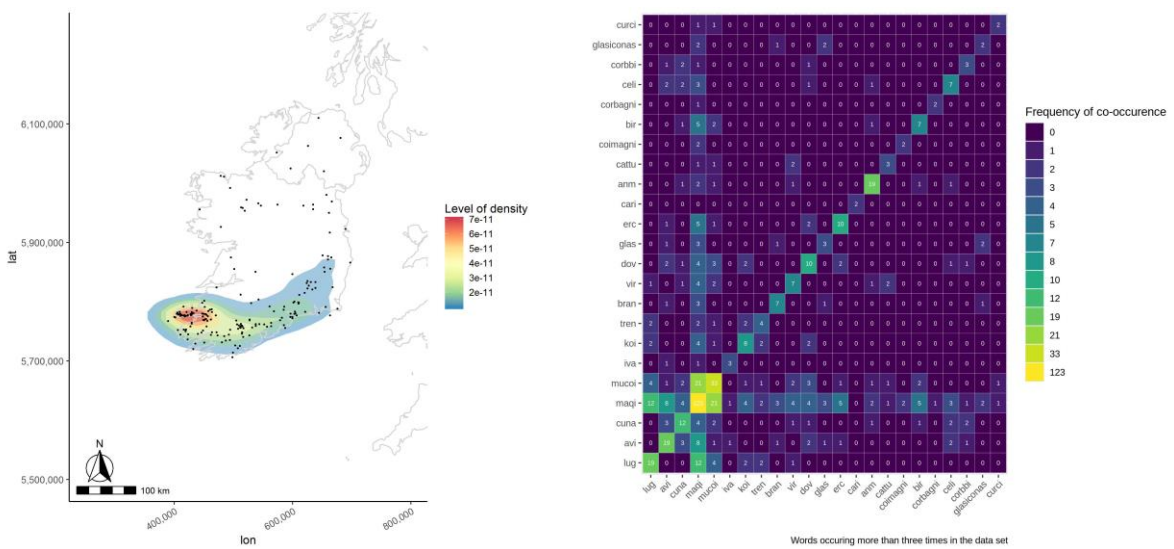right: distribution of family relation stones in QGIS (CC BY 4.0 Florian Thiery)*

figure 4:left: density plot created in R, right: co-occurrences of words in R (CC BY 4.0 Sophie C. Schmidt)

The Wikidata *SPARQL endpoint* enables the query of Ogham stones and their coordinates, to export the data and visualise the stone frequencies in third party software (cf. figure 3, l.) Using the *SPARQL Unicorn QGIS Plugin*, the Ogham stones can be queried and mapped in GIS software. For further research, GIS can be used to do geospatial analysis like analyse the distribution of stone in Ireland by certain family relations. Figure 3, r. indicates that most of the stones mention the word *MAQI* (son) and can be found in the province of Munster. Figure 4, l. shows a density plot of all Ogham Stones created within the *programming language R*. The main distribution of stones in the south of Ireland, especially the peninsula Dingle, is easily recognisable. An analysis of the contents of Ogham stones, i.e. a linguistic analysis has been

done using the [Ogham Extractor Tool](#). Usually, this involves an analysis of the texts' content using Natural Language Processing methods such as topic modelling (Murakami et al. 2017) for the purpose of categorising the input of a text. In this process, statistics about the texts' contents, e.g. their word frequencies or sentiment analysis, can be conducted. Usually, a dictionary of the available text corpus is created using vocabularies such as the Lexicon Model for Ontologies: Lemon (McCrae, Spohr and Cimiano 2011) and Ontologies of Linguistic Annotation (OLiA) (Chiarcos and Sukhareva 2015). The results can be annotated and shared as LOD or provide the basis for exports in GeoJSON, such as the ones shown in figure 3, r. As Ogham stones only provide limited text content, a simple keyword matching was sufficient to match meanings of names and to create a LOD dictionary out of the whole corpus of Ogham contents for further analysis in the linguistic or historical communities. Combined with spatial information, not only a spatial distribution of categorised names can be shown, but also the linguistic organization of the Ogham language in terms of words, phrases, characters and their interlinkage to concepts representing their meaning. As an example, an analysis showing how often two words co-occur on Ogham stones has been calculated (cf. figure 4, r.): The information *MAQI* (son) being supplemented with *MUCOI* (tribe) very often, shows the importance of the tribal affiliation and not just immediate family. On the other hand, it is interesting, that *ANM* (name) though occurring relatively often, coincides on only 4 stones together with *MAQI*.

10. SUMMARY AND OUTLOOK

This paper aimed to answer the questions: Is it possible to step into *Geodesy 3.0* doing SPARQLing geodesy for CH? Can publishing and analysing volunteered Linked (Geo-)Data in Wikidata preserve information on CH? We consider it possible and have exemplified a workflow using the *Ogi Ogham Project*. Some challenges remain, especially in the geospatial domain. Publishing strategies and applications for semantic data in order to integrate LOD in the common workflow are still needed. The *SPARQL Unicorn QGIS Plugin* is one step closer to achieving this. If the data is made accessible in a *Geodesy 3.0* approach, this data may be used in *AI* and *ML* projects to reach *Geodesy 4.0* to allow an excellent field of work in the future. In upcoming projects, the working group *Research Squirrel Engineers* will apply methods that preserve digital information on CH. We plan to use Linked Data techniques and the *SPARQL unicorn approach* to e.g. publish the rock art carvings in Alta, Norway, (Tansem and Johansen 2008) and make them semantically available. This World Heritage site is located next to the coast and is beginning to disappear as a result of erosion and rise in sea level from climate change. On the one hand it will be [conserved](#) by the [VAM](#) and on the other hand Linked Data will be [created](#) by the [Research Squirrel Engineers](#) to make the carvings available via [Wikidata](#).

11. ACKNOWLEDGEMENTS

REFERENCES [ONLINE: LAST ACCESS: 05.02.2020]

Aghaei, S 2012 Evolution of the World Wide Web : From Web 1.0 to Web 4.0, *International journal of Web & Semantic Technology*, 3(1): 1–10. DOI: https://doi.org/10.5121/ijwest.2012.3101.

Agoub, A, Kunde, F and Kada, M 2016 *Potential of Graph Databases in Representing and Enriching Standardized Geodata*. In: 06/2016.

Association of Research Libraries 2019 *ARL White Paper on Wikidata Opportunities and Recommendations*.

Atemezing, G, Corcho, O, Garijo, D, Mora, J, Poveda-Villalón, M, Rozas, P, Vila-Suero, D and Villazón-Terrazas, B 2013 Transforming meteorological data into Linked Data, *Semantic Web*, 4(3): 285–290. DOI: https://doi.org/10.3233/SW-120089.

Battle, R and Kolas, D 2012 Enabling the geospatial Semantic Web with Parliament and GeoSPARQL, *Semantic Web*, 3(4): 355–370. DOI: https://doi.org/10.3233/SW-2012-0065.

Benedict, K 2005 *The Open Geospatial Consortium Web Map, Web Feature and Web Coverage Service Standards ñ an Overview*.

Bennett, I, Devlin, G and Harrington, C 2016 CORCA DHUIBHNE 3D, *Archaeology Ireland*, 30(2): 17–20.

Berners-Lee, T 2006 *Linked Data - Design Issues*. 27 July 2006.

Berners-Lee, T, Hendler, J and Lassila, O 2001 The Semantic Web, *Scientific American*, 284(5): 34–43. DOI: https://doi.org/10.1038/scientificamerican0501-34.

Butler, H, Daly, M, Doyle, A, Gillies, S, Hagen, S and Schaub, T 2016 The GeoJSON Format, *Internet Engineering Task Force (IETF)*.

Chiarcos, C and Sukhareva, M 2015 OLiA – Ontologies of Linguistic Annotation, Hellmann, S., Moran, S., Brümmer, M., and McCrae, J.P. (eds.) *Semantic Web*, 6(4): 379–386. DOI: https://doi.org/10.3233/SW-140167.

Debruyne, C, Meehan, A, Clinton, É, McNerney, L, Nautiyal, A, Lavin, P and O'Sullivan, D 2017 Ireland?s Authoritative Geospatial Linked Data. In: d'Amato, C., Fernandez, M., Tamma, V., Lecue, F., Cudré-Mauroux, P., Sequeda, J., Lange, C., and Heflin, J. (eds.) *The Semantic Web – ISWC 2017*. Cham: Springer International Publishing. pp. 66–74. DOI: https://doi.org/10.1007/978-3-319-68204-4_6.

Goodwin, J, Dolbear, C and Hart, G 2008 Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web, *Transactions in GIS*, 12: 19–30. DOI: https://doi.org/10.1111/j.1467-9671.2008.01133.x.

Gruber, E 2018 Linked Open Data for Numismatic Library, Archive and Museum Integration. In: Matsumoto, M. and Uleberg, E. (eds.) *CAA2016. Oceans of Data. Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology*. 2018 Oxford: Archaeopress. pp. 55–62.

Hahmann, S and Burghardt, D 2010 *Connecting LinkedGeoData and Geonames in the Spatial Semantic Web*.

Hahmann, S and Burghardt, D 2012 Forschungsergebnisse zur Frage: Haben 80% aller Informationen einen Raumbezug?, *gis.SCIENCE - Die Zeitschrift für Geoinformatik*, 25(3): 101–108.

Hausenblas, M and Boram Kim, J G 2015 *5-star Open Data*. 5 ★ OPEN DATA, 31 August 2015.

Kashyap, R 2019 Geospatial Big Data, Analytics and IoT: Challenges, Applications and Potential. In: Das, H., Barik, R.K., Dubey, H., and Roy, D.S. (eds.) *Cloud Computing for Geospatial Big Data Analytics*. Cham: Springer International Publishing. pp. 191–213. DOI: https://doi.org/10.1007/978-3-030-03359-0_9.

Khayari, A and Banzet, G 2019 *Alignement d'un thésaurus sur GeoNames : retour d'expérience*.

Lasi, H, Fettke, P, Kemper, H-G, Feld, T and Hoffmann, M 2014 Industry 4.0, *Business & Information Systems Engineering*, 6(4): 239–242. DOI: https://doi.org/10.1007/s12599-014-0334-4.

Lockyear, K 2000, The Celtic Inscribed Stones Project Database: a guide. https://www.ucl.ac.uk/archaeology/cisp/database/manual/ , accessed 2020-01-27.

Lopes, N, Grant, R, Ó Raghallaigh, B, Ó Carragáin, E, Collins, S and Decker, S 2014 Linked Logainm: Enhancing Library Metadata Using Linked Data of Irish Place Names. In: Bolikowski, Ł., Casarosa, V., Goodale, P., Houssos, N., Manghi, P., and Schirrwagen, J. (eds.) *Theory and Practice of Digital Libraries -- TPDL 2013 Selected Workshops*. Cham: Springer International Publishing. pp. 65–76. DOI: https://doi.org/10.1007/978-3-319-08425-1_7.

Macalister, R A S 1945 *Corpus inscriptionum insularum Celticarum*. Vol. I. Dublin: Stationery Office.

Macalister, R A S 1949 *Corpus inscriptionum insularum Celticarum*. Vol II. Dublin: Stationery Office.

MacManus, D 1997 *A guide to Ogam*. Maynooth monographs 4. Maynooth: An Sagart.

McCrae, J, Spohr, D and Cimiano, P 2011 Linking Lexical Resources and Ontologies on the Semantic Web with Lemon. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., and Pan, J. (eds.) *The Semantic Web: Research and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 245–259. DOI: https://doi.org/10.1007/978-3-642-21034-1_17.

Mayer, C 2016 *Das OOO-Modell: Open Source - Open Data - Open Access. Wie GIS im Jahre 2016 funktionieren kann.*

Murakami, A, Thompson, P, Hunston, S and Vajn, D 2017 'What is this corpus about?': using topic modelling to explore a specialised corpus, *Corpora*, 12(2): 243–277. DOI: https://doi.org/10.3366/cor.2017.0118.

Nogueras-Iso, J, Zarazaga-Soria, F J, Béjar, R, Álvarez, P J and Muro-Medrano, P R 2005 OGC Catalog Services: a key element for the development of Spatial Data Infrastructures, *Computers & Geosciences*, 31(2): 199–209. DOI: https://doi.org/10.1016/j.cageo.2004.05.015.

Portele, C 2007 *OpenGIS® Geography Markup Language (GML) Encoding Standard. Version 3.2.1*.

Prudhomme, C, Homburg, T, Ponciano, J-J, Boochs, F, Cruz, C and Roxin, A-M 2019 Interpretation and automatic integration of geospatial data into the Semantic Web, *Computing*, 1–27.

Prud'hommeaux, E and Seaborne, A 2008 SPARQL Query Language for RDF. *W3C Recommendation*, 15 January, .

Pütt, K 2018 Documentation and Digital Preservation of Syrian Heritage: A German Archive Project for Syria, *The Public Historian*, 40(4): 107–128. DOI: https://doi.org/10.1525/tph.2018.40.4.107.

Sanderson, R 2019 *LOUD: Linked Open Usable Data*. 28 May 2019.

Severance, C 2012 Discovering JavaScript Object Notation, *Computer*, 45(4): 6–8. DOI: https://doi.org/10.1109/MC.2012.132.

Shadbolt, N, O'Hara, K, Berners-Lee, T, Gibbins, N, Glaser, H, Hall, W and schraefel, m. c. 2012 Linked Open Government Data: Lessons from Data.gov.uk, *IEEE Intelligent Systems*, 27(3): 16–24. DOI: https://doi.org/10.1109/MIS.2012.23.

Simon, R, Isaksen, L, Barker, E and de Soto Cañamares, P 2016 The Pleiades Gazetteer and the Pelagios Project. In: Berman, M.L., Mostern, R., and Southall, H. (eds.) *Placing names: enriching and integrating gazetteers*. The spatial humanities. Bloomington: Indiana University Press. pp. 97–109.

Stadler, C, Lehmann, J, Höffner, K and Auer, S 2012 LinkedGeoData: A core for a web of spatial open data, *Semantic Web*, 3(4): 333–354. DOI: https://doi.org/10.3233/SW-2011-0052.

Tansem, K and Johansen, H 2008 *The World Heritage Rock Art in Alta* 2008(Adoranten 2008): 1–20.

Thiery, F 2020 *Ogham Words* DOI: https://doi.org/10.5281/zenodo.3611029.

Thiery, F and Schmidt, S C 2020a *Ogham Stones Wikidata Import* DOI: https://doi.org/10.5281/zenodo.3612654.

Thiery, F and Schmidt, S C 2020b *Ogham Townlands Wikidata Import* DOI: https://doi.org/10.5281/zenodo.3612649.

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

Trognitz, M and Thiery, F 2019 *Wikidata - A SPARQL(ing) Unicorn?* DOI: https://doi.org/10.5281/zenodo.3471404.

Vrandečić, D and Krötzsch, M 2014 Wikidata: a free collaborative knowledgebase, *Communications of the ACM*, 57(10): 78–85. DOI: https://doi.org/10.1145/2629489.

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

Wahlster, W 2017 *Artificial Intelligence as the driver of the second wave of digitalization*. 26 September 2017.

Wenjue, J, Yumin, C and Jianya, G 2004 Implementation of OGC web map service based on web service, *Geo-spatial Information Science*, 7(2): 148–152. DOI: https://doi.org/10.1007/BF02826653.

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 ., 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 and Mons, B 2016 The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3: 160018.

Wuttke, U 2019 "Here be dragons": Open Access to Research Data in the Humanities. *ulrikewuttke*.

Würriehausen, F, Homburg, T and Müller, H 2016 Using an INSPIRE Ontology to Support Spatial Data Interoperability, *INSPIRE 2016, Barcelona, Spain*.

BIOGRAPHICAL NOTES

Florian Thiery is a geodesist and Research Software Engineer working in the Cultural Heritage domain. He 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). *Sophie Charlotte Schmidt* is a computational archaeologist specialised in statistical data analysis using R. Mrs. Schmidt is a member of the advisory board of the German speaking CAA chapter. *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. His PhD thesis deals with semantic geospatial data integration and the quality of geospatial data in this Semantic Web context. *Martina Trognitz* studied Computational Linguistics and Classical Archaeology at the University of Heidelberg and is currently working on a PhD. Mrs. Trognitz is a member of the advisory board of the German speaking CAA chapter. *Monika Przybilla* holds a university degree in geodesy and has long term activities in the DVW working group 1 "Profession/Education", chair since 2015.

CONTACTS

Florian Thiery M.Sc.
Research Squirrel Engineers, Mainz
Josef-Traxel-Weg 4
D – 55128 Mainz
GERMANY
Email: rse (at) fthiery (dot) de
Web site: http://fthiery.de

Timo Homburg M.Sc.
Institute for Spatial Information and Surveying Technology, Mainz, GERMANY
Email: timo.homburg(at)hs-mainz.de
Sophie Schmidt M.A.
University of Cologne, Institute of Archaeology, Cologne, GERMANY
Email: s.c.schmidt(at)uni-koeln.de
Mag. Martina Trognitz
Austrian Centre for Digital Humanities and Cultural Heritage, Vienna, AUSTRIA
Email: martina.trognitz(at)oeaw.ac.at
Dipl.-Ing. Monika Przybilla
Regionalverband Ruhr, Essen, GERMANY
Email: monika.przybilla(at)dvw.de