Using Maps for Interlinking Geospatial Linked Data

Dieter Roosens[1], Kris McGlinn[2], Christophe Debruyne[1,2][0000-0003-4734-3847]

1 WISE Lab, Vrije Universiteit Brussel, Brussels, Belgium
{first.last}@vub.be
2 ADAPT Centre, Trinity College Dublin, Dublin, Ireland
{first.last}@adaptcentre.ie

Abstract. The creation of interlinks between Linked Data datasets is key to the creation of a global database. One can create such interlinks in various ways: manually, semi-automatically, and automatically. While quite a few tools exist to facilitate this process in a (semi-)automatic manner, often with support for geospatial data. It is not uncommon that interlinks need to be created manually, e.g., when interlinks need to be authoritative. In this study, we focus on the manual interlinking of geospatial data using maps. The State-of-the-Art uses maps to facilitate the search and visualization of such data. Our contribution is to investigate whether maps are useful for the creation of interlinks. We designed and developed such a tool and set up an experiment in which 16 participants used the tool to create links between different Linked Data datasets. We not only describe the tool but also analyze the data we have gathered. The data suggests the creation of these interlinks from these maps is a viable approach. The data also indicate that people had a harder time dealing with Linked Data principles (e.g., content negotiation) than with the creation of interlinks.

Keywords: Linked Data Interlinking, Maps, Geospatial Linked Data.

1   Introduction

Linked Data [1] is a set of best practices and guidelines to publish and interlink data on the Web. Those practices and guidelines prescribe how to cleverly combine several standardized technologies: the Resource Description Framework[1] (RDF) to describe things; URIs to identify those things and, in particular, HTTP URIs to retrieve these descriptions over the Web; and content-negotiation to retrieve the desired format (HTML for users and RDF for machines). We can observe the importance of geospatial data by the vast amount of geographic or geospatial datasets available on the Linked Data Web. The Linked Open Data cloud initiative tries to visualize the relations between such datasets that are both open and meet the inclusion criteria. While the image arguably only scratches the surface of the Linked Data available on the Web, it does indicate the importance of the datasets that contain geospatial information.

---

[1]   https://www.w3.org/RDF/

Key to the creation of a Linked Data Web is the creation of interlinks between resources both within and across Linked Data datasets. Moreover, as most datasets also have a geospatial dimension that is either explicit or implicit, it comes as no surprise that "location" is a convenient way for aligning and combining different datasets [21].

For the creation of interlinks, methods were proposed to create such Linked Data interlinks automatically. These methods often rely on subject matter experts to *validate* links. But what if interlinks have to be created manually by subject matter experts? Manual interlinking is essential when such interlinks need be authoritative. The study reported in [13] surveyed subject matter experts in the library field, and it indicated that there were challenges in tooling for creating authoritative interlinks. We believe that these challenges exist for other fields as well.

Geospatial information lends itself naturally to maps. Within the Semantic Web community, initiatives often limit the use of maps to displaying information or querying. To the best of our knowledge, no study has looked into the use of maps for the creation of links. As we were surprised by the lack of such tooling, we aim to answer the following question in this paper: "Can maps be used to create and manage interlinks of geospatial data on the Linked Data web?"

The remainder of this paper is organized as follows: Section 2 provides an overview of the related work; Section 3 introduces our requirements and elaborates how on the design and development of our tool; Section 4 describes the experiment we devised to evaluate our tool, the data we gathered, and an analysis of the data; and Section 5 concludes our paper.

2 Related Work

The State-of-the-Art often focusses on the creation of interlinks in a (semi-)automatic manner. In [20], the authors reported on declaring SILK rules [5] for generating interlinks between an Irish place names dataset and DBpedia [9]. Approaches such as SILK allow one to configure thresholds to classify pairs of entities as 'accepted', 'rejected', or 'to be validated'. The LIMES [15] interlinking platform, on the other hand, adopted EAGLE [16] and WOMBAT [23] to combine machine learning with geospatial functions for discovering interlinks. Others, such as [6], proposed semi-supervised approached in which users can guide the generation of rules that will generate the interlinks. While these approaches are valuable for the creation of a Linked Data Web, some interlinks need to be created manually [13]. While [13] reported on the challenges for authoritative Linked Data interlinking, we noticed that the state-of-the-art has not yet looked at the use of maps for creating interlinks between datasets that have a geospatial dimension. Within the Semantic Web community, however, the use of maps is often limited to displaying information or querying. Examples include:

— YASGUI [19] is a suite of tools that constitute a SPARQL editor and is entirely written in JavaScript. YASGUI has a plugin for displaying SPARQL query results on a map. Polygons and points on a map can be provided a label and a color if users follow a particular variable naming convention.

— Strabon [8] is a triplestore with support for GeoSPARQL[2] and allows the results of a GeoSPARQL query to be rendered on a map.
— FACETE [24] is a faceted browser for geospatial information. Unlike SPARQL in which one has to formulate a query, one is "guided" in searching for specific resources by constraining the selection criteria with so-called "facets". Each facet narrows down the results. Tools such as FACETE comprehend the facets by providing the right widgets depending on a facet's type –input validation for numbers, date and time pickers for `xsd:dateTime`,… Again, FACETE only uses map to display and engage with the results.
— GVIZ [11] is a tool that allows one to query GeoSPARQL triplestores by drawing search areas on a map. A user can achieve this by drawing the area as a polygon on a map. Their tool then translates that polygon into a GeoSPARQL query using Well-Known Text (WKT)[3] for representing that area.

The GeoKnow project was funded by the EU FP7 Programme and ran from 2013 to 2015. The project aimed to provide a suite of solutions to integrate and enrich external datasets with geospatial data using semantic technologies called the GeoKnow workbench [2]. The workbench provided support for the whole data lifecycle of projects involving geospatial datasets. This workbench integrated FACETE, which we have mentioned before, but also includes tools for transforming non-RDF into RDF with TripleGeo [18], automatic integration with LIMES [15], and discovery of implicit geospatial information with DEER [22]. The workbench does not, to the best of our knowledge, use maps to drive the interlinking process.

3   Design and Development

We observed from the State-of-the-Art that maps are mostly used to display results. Other studies have shown that the creation of Linked Data interlinks not always an automatic process and that appropriate tooling is lacking [13]. We wondered whether subject matter experts (e.g., historians, librarians, and archivists) could use maps for the creation of interlinks. Those subject matter experts will often seek resources to link and know what to look for. Our proposed solution should thus aid subject matter experts in creating Linked Data datasets. To design and develop a tool allowing one to create Linked Data interlinks on a map, we formulated the following requirements:

1. Search for geospatial resources in a Linked Data dataset;
2. Display the results on a map;
3. Engage with the resources on the map for the creation of interlinks;
4. Manage the created interlinks;

---

[2] https://www.opengeospatial.org/standards/geosparql, last accessed June 2019. GeoSPARQL is a standardized geospatial extension of the SPARQL query language. The extension consists of a vocabulary, geospatial functions, and query transformation rules to related predicates to geospatial functions.
[3] https://en.wikipedia.org/wiki/Well-known_text_representation_of_geometry, last accessed June 2019

5. Keep track of provenance information.

Requirements 2, 3, and 4 are key to our study. The first requirement is necessary for the creation of an experiment as we need to simulate a realistic workflow; subject matter experts will first look for the resources that they want to relate with other datasets. The use of SPARQL to search for resources is indeed a possibility but would require users to know the language. Instead, we will adopt a suitable technique for searching resources in a dataset.

In [25], the authors stated that provenance information "provides information about [a resource's] origin, such as who created it, when it was modified, or how it was created." This type of information is useful to assess the trustworthiness of an interlink (e.g., by determining who has created an interlink, and why). The last requirement is essential to consider the integration of our tool into broader interlinking or governance frameworks

As for the types of geospatial Linked Data datasets we want to support, we limit ourselves to such datasets that provide a SPARQL endpoint and use GeoSPARQL in combination with WKT to represent the geographic features and geometries. We can easily ingest a geospatial Linked Data dataset in their own (local) triplestore and endpoint if one wishes to avail of a dataset that has no (reliable) SPARQL endpoint. As for GeoSPARQL, we limit ourselves to GeoSPARQL to assess the viability of our approach. If viable, we will look into the other vocabularies for representing coordinates and geometries. GeoSPARQL also prescribes support for Geographic Markup Language (GML)[4] to represent geometries, but we will only consider incorporating GML once the approach is shown to be viable.

3.1 Searching for Geospatial Resources

While our study focuses on the use of maps for interlinking geospatial data on the Linked Data Web, a user would start by looking for the resources they want to interlink. A user can avail of GeoSPARQL and use tools such as YASGUI [19] or even adopt Faceted Search for RDF datasets with, for instance, Facete [24] and SemFacet [7]. We argue that it will be unlikely for subject matter experts to use SPARQL and deemed the adoption of faceted browsing more appropriate.

We initially wanted to reuse codebases of existing tools but were faced with a couple of challenges. Facete, for instance, was built for geospatial information, but we were unable to compile and run both versions of the tool.[5,6] Both versions required a convoluted setup using different programming languages and servers, and both – given the last commits were, at the time of writing, over four years ago – relied on deprecated libraries that were difficult to find or difficult to get working. SemFacet's codebase, referenced from their website, was unavailable.[7] We did find some branches hosted by others but were unable to run the project without problems.

Given that the means for searching was subordinate to the use of maps for creating links, we decided to design and develop a minimalistic faceted browser solely for the evaluation (see Fig. 1). In this browser, there are two facets: type-facets based on `rdf:type` and property-facets (based on the other properties of the Linked Data dataset.
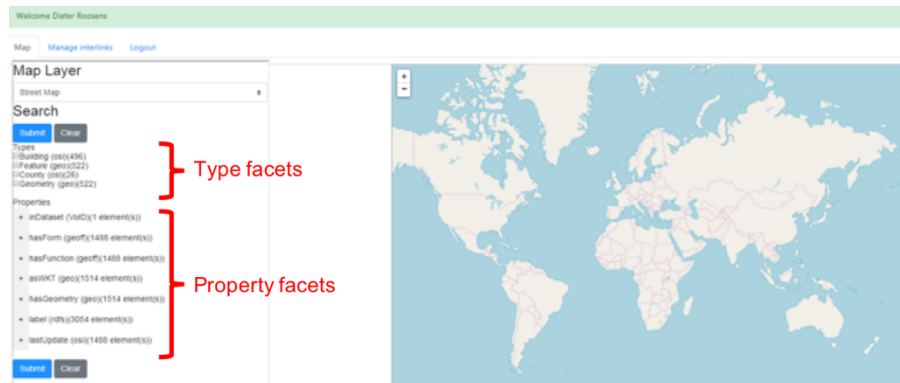

Fig. 1. Faceted browser built for the experiment

Property-facets are facets for property-value pairs, which are grouped by the property. For each selected facet, there is a new triple pattern added to the SPARQL query of the Linked Data dataset. Type-facets result in a triple pattern of the form `?this rdf:type *x*` where *x* is the URI of the selected type. Property-facets result in a triple pattern of the form `?this *y* *z*` where *y* is the URI of the predicate and *z* the value. The query furthermore retrieves:

1. An optional graph pattern, which looks for the geometries of `?this`:

   ```
   OPTIONAL { ?this geo:hasGeometry [ geo:asWKT ?wkt ] }
   ```

   We have chosen to make this graph pattern optional as the system would have otherwise only sought resources with a geometry. This approach furthermore allows us to inform of resources that meet all criteria but do not have a geometry. For this study, we have chosen to limit ourselves to GeoSPARQL.

2. Labels, by looking for triples using the `rdfs:label`, predicate, with a preference for English labels, followed by default and any other labels. We will use those labels for the interface elements of Requirements 2 and 3.

3.2 Displaying Information on a Map

Before one can engage with resources on a map, they need to be displayed. To this end, we drew inspiration from YASGUI [19]. YASGUI uses a particular variable naming convention to customize colors and add labels to the geometries (e.g., points and polygons). It is up to the person writing the query to manage those variable names. We did

appreciate the visuals provided by the Leaflet[8] that YASGUI adopted. Since we have adopted GeoSPARQL and focus on WKT representations of geometries, we need to use Wicket[9] to translate WKT into a JSON representation supported by Leaflet. Fig. 2 shows how multi-polygons and points are displayed on the map. Both are the result of a particular facet being chosen; counties in Ireland for the former and buildings (shops) in the latter.
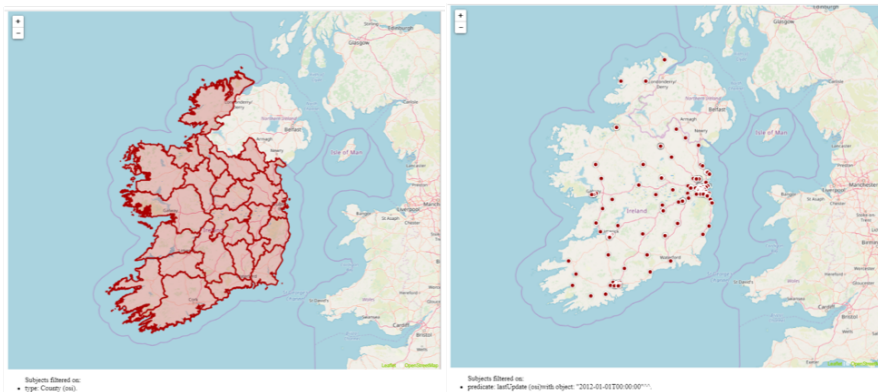
Fig. 2. Supporting the visualization of (multi)-polygons and point geometries.

3.3   Creating and Managing Interlinks

One creates interlinks by clicking on geometries, which result in a pop-up. The design of the popup was driven by ConcurTaskTrees [17]. ConcurTaskTrees allow one to model the tasks of users and systems. Those models can be used to drive the design of a user interface [14]. The algorithm outlined by [14] was used to create a mock-up, which was subsequently implemented (see Fig. 3). For managing and downloading the created interlinks, we created a separate tab displaying the information in a tabular manner. The table contains information on the triple constituting the interlink, the user, and the time it was created. Interlinks are stored in a local triplestore. Details on how we store the interlinks will be provided in Section 2.4.

---

8   https://leafletjs.com/, last accessed June 2019
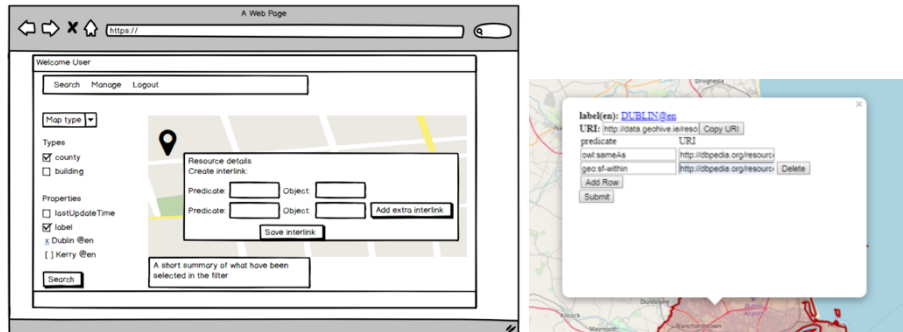9   http://arthur-e.github.io/Wicket/, last accessed June 2019

Fig. 3. From mock-up to implementation. Clicking on a polygon allows one to create interlinks

3.4 Knowledge Organization and Provenance

To facilitate storing interlinks and their metadata, we have chosen to save, for each interlink, the interlink as a reified RDF statement connected to its metadata in a separate named graph. Fig. 4 depicts an example of such a reified RDF statement in a graph. We also avail of the PROV-O[10] ontology, a W3C Recommendation for capturing provenance information. We thus have a resource that is both an RDF statement and an instance of `prov:Entity,` and represent its creation, creator, and so forth using PROV-O.
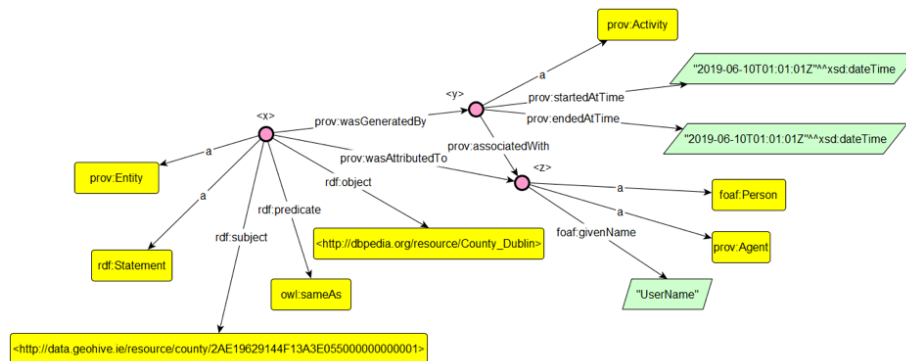
Fig. 4. Using reified RDF statements to treat interlinks as instances of `prov:Entity`. The RDF statement, the activity, and person are all identified by a URI, though we used x, y, and z for brevity.

While seemingly complicated, this approach allows our tool to be integrated into more complex environments. Our approach to knowledge-organization fits methods adopting niche-sourcing (such as DaCura [4]) where multiple subject matter experts are tasked with annotating resources, and one can analyze which annotations occur more often.

To download the set of interlinks that one has created, we use a simple SPARQL CONSTRUCT query (Listing 1).

```
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
CONSTRUCT { ?subject ?predicate ?object } WHERE {
  GRAPH ?g {
    ?statement prov:wasAttributedTo <__agent__> .
    ?statement rdf:subject ?subject ;
               rdf:object ?object ;
               rdf:predicate ?predicate . } }
```

Listing 1. A SPARQL CONSTRUCT query to create a set of interlinks created by a particular user. Note that "__agent__" is replaced by the URI of the user.

3.5   Implementation

The prototype is available on GitHub.[11] It is a Java Web application and uses Apache Jena ARQ for communicating with the SPARQL endpoint of the Linked Data dataset to be linked and uses Apache Jena Fuseki for storing the interlinks.[12]

4   Experiment

To evaluate our approach, we devised an experiment. We recruited 16 participants and asked them to:

1. sign an informed consent form;
2. fill in a pre-questionnaire where they assessed their prior knowledge in Linked Data technologies and geospatial information;
3. look at a YouTube video providing a quick introduction to RDF and Linked Data[13];
4. consult at any time slides summarizing important points of the video as well as annotated screenshots of the tool;
5. perform the tasks in a think-aloud manner; and
6. fill in a post-questionnaire to assess the tool.

We have chosen to adopt the GeoHive [3] dataset for the experiment, as this dataset has not yet been included in the Linked Open Data cloud and contains very few interlinks to other Linked Data. As GeoHive has decided not to provide a SPARQL endpoint, we created a local endpoint that included – for the sake of the experiment – boundary information of Ireland's counties and some metadata of shops (part of [12]).

---

[11] https://github.com/dieterroosens/LinkedDataApplication
[12] https://jena.apache.org/, last accessed June 2019
[13] https://www.youtube.com/watch?v=zeYfT1cNKQg, until 5'45"

The eight tasks related to the tool were (note we provide the written instructions given to the participants for only a couple of the tasks):

1. To state that County Dublin is `owl:sameAs` its resource in DBpedia [16] and that County Dublin is `geo:sf-within dbpedia:Ireland`. We provided the URIs for both Dublin and Ireland in the document. I.e., participants could copy the URIs. We gave participants almost step-by-step instructions on how to achieve this task. The task was written as follows:
   (1) Go to the screen "Map" and search for the county with label "Dublin"
   (2) Select the element on the map and interlink it with "geo:sf-within" to the URI http://dbpedia.org/resource/Dublin
   (3) Do the same, but now with the predicate "owl:sameAs" and the URI http://dbpedia.org/resource/County_Dublin

2. To state that County Donegal is `owl:sameAs` as its corresponding resource in DBpedia, which participants had to look up. Now, participants were tasked with looking up County Donegal in DBpedia and copying it URI.

3. Given the URI for `dbpedia:Wicklow_Mountains`, to identify in which county they reside and create that relationship. Here, the trick was to look for the appropriate relation `geo:sf-contains`. It was up to the participant to identify the correct relation. The interface provided a list of predicates, but participants were able to provide their own.

4. To create interlinks within our dataset. Identify the counties that border cork and create one `geo:sf-touches` relations. It was up to the participant to identify the correct relation. The interface provided a list of predicates, but participants were able to provide their own. The task was written as follows:
   (1) Select all resources of the type County
   (2) Identify the counties that touch the border of County Cork
   (3) Create a relationship (interlink) between Cork and those counties. There are 2 possibilities: From Cork to that county, or From that county to Cork. It is not necessary to declare both for a county x and Cork, but we encourage you to try both approaches for different counties.

5. To look for the highest mountain on the map, which requires switching layers. Then find the DBpedia resource representing that mountain and create the relationship. We provided a hint that the mountain was 1039m high.

6. To look for a specific shop and stating that its location is County Dublin using the representation of County Dublin in GeoNames[14]. Here, the participant had to use the facets to look for that shop. The task was written as follows:
   (1) Select the shop that has been last updated on "2014-01-16T13:14:52"
   (2) Interlink the shop with the county it is located in. Use the URI of the county that you can find in GeoName.

7. To look at the interlinks that the participant has created and correct any mistakes. We introduced one incorrect interlink which participants had to spot.

8. To download the interlinks.

---

[14] https://www.geonames.org/, last accessed July 2019

The difficulty of the tasks increases from 1 to 6. We started by giving participants clear instructions and even the URIs to copy. From the second task onwards, participants had to look for the appropriate resources themselves. From the third task onwards, participants had to choose the predicate. Tasks 1 to 6 were related to searching for resources, displaying those resources, and interlinking resources. They thus correspond with the first three requirements. Tasks 7 and 8 were related to the fourth requirement, managing the created interlinks. We did not formulate tasks for using the provenance information, as that would have required more in-depth knowledge of SPARQL and PROV-O and is outside the scope of this study –creating interlinks from a map. The provenance graphs were used to display information for Task 7.

4.1 Prior Knowledge

Before the experiment, participants were asked to assess their knowledge on the Web Ontology Language (OWL), Resource Description Framework (RDF), the Terse Triple Language (TTL), and geospatial information (GEO) in "general". Participants were able to indicate their level of familiarity with 'none' (0), 'low' (1), 'medium' (2), and 'high' (3). Table 1 not only provides an overview of the answers, but we also computed a total for future analysis. From this table, we can conclude that only 4 of the 16 participants deemed themselves knowledgeable in these technologies (a score of 2 or more for at least one of the four technologies).

Table 1. Summary of the pre-questionnaire in which participants assessed their knowledge of the Web Ontology Language, Resource Description Framework, the Terse Triple Language, and Geospatial Information. Values range from 0 (none) to 3 (high).

Participant	OWL	RDF	TTL	GEO	Total
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	1	1
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	2	3	3	1	9
11	2	3	3	2	10
12	3	3	3	2	11
13	2	1	1	1	5
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	0	0	0

4.2 On Task Performance

All participants managed to download the interlinks, which were used to compare their results with that of a gold standard –the expected output. There were multiple correct

answers for Task 4, so we manually checked those. When comparing the interlinks with the expected results, we noticed a couple of things:

— Many participants were unaware of the content-negotiation mechanism used by Linked Data datasets. Since participants were given the URI for the first interlink, many were able to execute the first task successfully. In other tasks where they had to look for DBpedia resources themselves, however, they looked for the resources in a browser and then copied the resource's page URI instead of the resource's URI.

— Only one participant was able to find the resource representing County Dublin in the GeoNames. The participant – in some capacity involved in the OSi projects – did state during the think-aloud experiment that the question was "tricky", and they knew how to find the resource. The other participants choose another URI, often referring to the city of Dublin.

— Some participants did not create interlinks within the dataset (Task 4) or an interlink with GeoNames (Task 6) and instead continued creating interlinks with DBpedia. We deem that this was due to haste, i.e., they read over those requirements.

Given the first two observations, we had decided to create a so-called "silver standard" in which we gave partial credit when the chosen URIs were either pointing to the page rather than the resource or when the city of Dublin was selected from GeoNames. From Table 2, it is clear that participants had trouble with Tasks 2 and 6. The silver standard does provide a more nuanced view of the results. Most participants had difficulty with DBpedia's content negotiation (Task 2), and some participants were able to choose and link to a resource with the label "Dublin" in GeoNames (Task 6). While none of the participants were able to complete all tasks correctly, we believe that these errors were not due to the tool, but a lack of experience with Linked Data principles and the datasets we have adopted.

Table 2. Results of the participants' performance.

| Participant | w.r.t. Gold Standard | | | | | | | | | w.r.t. Silver Standard | | | | | | | | |
	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Percentage	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Percentage
1	1.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	62.50	1.00	0.50	1.00	1.00	0.50	0.00	1.00	1.00	75.00
2	1.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	62.50	1.00	0.50	1.00	1.00	0.50	0.00	1.00	1.00	75.00
3	1.00	0.00	0.00	1.00	1.00	0.00	1.00	1.00	62.50	1.00	0.50	0.00	1.00	1.00	0.50	1.00	1.00	75.00
4	1.00	0.00	1.00	0.75	1.00	0.00	1.00	1.00	71.88	1.00	0.50	1.00	0.75	1.00	0.00	1.00	1.00	78.13
5	1.00	0.00	1.00	0.25	1.00	0.00	1.00	1.00	65.63	1.00	0.00	1.00	0.25	1.00	0.00	1.00	1.00	65.63
6	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	87.50	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	87.50
7	0.50	1.00	1.00	0.75	1.00	0.00	1.00	1.00	78.13	0.50	1.00	1.00	0.75	1.00	0.00	1.00	1.00	78.13
8	1.00	1.00	0.00	0.75	1.00	0.00	1.00	1.00	71.88	1.00	1.00	0.00	0.75	1.00	0.00	1.00	1.00	71.88
9	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00	75.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00	75.00
10	0.50	0.00	1.00	1.00	1.00	0.00	1.00	1.00	68.75	0.50	0.50	1.00	1.00	1.00	0.50	1.00	1.00	81.25
11	1.00	0.00	0.00	1.00	0.00	0.00	1.00	1.00	50.00	1.00	0.50	0.00	1.00	0.50	0.50	1.00	1.00	68.75
12	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	87.50	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	87.50
13	1.00	0.00	1.00	1.00	1.00	0.00	0.00	1.00	62.50	1.00	0.50	1.00	1.00	1.00	0.00	0.00	1.00	68.75
14	1.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	62.50	1.00	0.50	1.00	1.00	0.50	0.50	1.00	1.00	81.25
15	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	87.50	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	93.75
16	1.00	0.00	0.00	0.50	1.00	0.00	1.00	1.00	56.25	1.00	0.50	0.00	0.50	1.00	0.50	1.00	1.00	68.75
AVG:	0.94	0.25	0.75	0.81	0.75	0.13	0.94	1.00	69.53	0.94	0.56	0.75	0.81	0.88	0.28	0.94	1.00	76.95

4.3 On the Tool's Perceived Usability

Now we look into our tool's perceived usability. We have chosen to adopt the Post-study System Usability Questionnaire (PSSUQ) [10]. The survey consists of 19 questions using an invert Likert Scale from 1 (completely agree) to 7 (completely disagree). The questions are listed in Appendix A. Unlike other usability surveys, PSSUQ assesses three aspects of a system: the system's usefulness using questions 1 to 8; the information quality using questions 9 to 15; and the interface quality using questions 16 to 18. Question 19 asks a participant about their overall satisfaction and the average of questions 1 to 19 are used to compute the overall score. Participants were also able to indicate that a question was not applicable.

The results are shown in Table 3. A box plot of these values is shown in Fig. 5. Note that values closer to 1 are more positive and that a value of 4 is "neutral". With that in mind, we notice that there are four aspects worth investigating. Questions 5 and 6 – part of System Usefulness – were related to the tool's ability to help one efficiently complete the tasks and how comfortable it was doing so. Questions 9 and 10 were associated with Information Quality: error messages and instructions to recover from mistakes. Faceted browsing was not the focus of our study; we merely adopted the technique as the starting point for the interlinking tasks. Problems with the faceted browsing may have harmed the overall perceived usability and thus also the interaction with the map. We do recognize that the system lacked clear error messages and documentation.

Table 3. Results of the PSSUQ questionnaire

#	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	SysUse	InfoQual	IntQual	Overall
1	3	3	2	3	3	3	2	3	3	3	3	3	3	3	3	3	2	2	2	2.75	3.00	2.33	2.74
2	3	3	3	3	3	3	3	3	3	3	3	2	3	2	3	3	3	3	2	3.00	2.71	3.00	2.84
3	3	2	3	2	4	2	2	3	3	5	2	3	3	3	2	2	2	2	2	2.63	3.00	2.00	2.63
4	2	2	4	5	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	2.63	2.14	2.00	2.32
5	4	4	3	4	4	4	3	3	5	3	5	4	5	5	4	3	3	3	3	3.63	4.43	3.00	3.79
6	2	2	3	2	3	3	2	3	4	4	3	2	3	3	3	3	3	3	2	2.50	3.14	3.00	2.79
7	3	2	3	3	3	3	3	2	3	3	3	2	3	2	4	3	3	2	3	2.75	2.86	2.67	2.79
8	6	5	4	2	5	5	5	5	7	5	6	5	5	4	5	4	4	4	5	4.63	5.29	4.00	4.79
9	1	2	2	3	4	5	2	2	-	3	2	3	2	2	1	1	1	2	2	2.63	1.86	1.33	2.11
10	2	3	2	3	3	2	2	1	2	2	2	2	1	2	2	2	2	3	2	2.25	1.86	2.33	2.11
11	2	3	2	3	3	2	2	2	5	2	2	2	2	2	2	2	2	4	2	2.38	2.43	2.67	2.42
12	3	3	2	3	2	1	3	4	7	-	7	4	2	2	4	5	4	-	3	2.63	3.71	3.00	3.11
13	4	4	5	5	5	5	3	4	6	5	3	2	5	3	6	6	6	2	3	4.38	4.29	4.67	4.32
14	2	2	2	1	2	1	1	1	-	1	1	1	1	1	2	1	1	1	1	1.50	1.00	1.00	1.21
15	2	2	1	1	2	2	1	1	-	-	3	2	3	3	2	3	2	-	2	1.50	1.86	1.67	1.68
16	7	7	4	5	6	7	6	6	7	5	5	4	5	6	5	5	7	-	7	6.00	5.29	4.00	5.47
AVG	3.06	3.06	2.81	3.00	3.38	3.13	2.63	2.81	4.38	3.29	3.25	2.75	3.00	2.81	3.13	3.00	2.94	2.54	2.69	2.98	3.05	2.67	2.94
STDEV	1.57	1.39	1.05	1.26	1.20	1.67	1.31	1.42	2.45	1.67	1.65	1.06	1.37	1.28	1.41	1.41	1.65	1.29	1.45	1.16	1.26	0.99	1.13
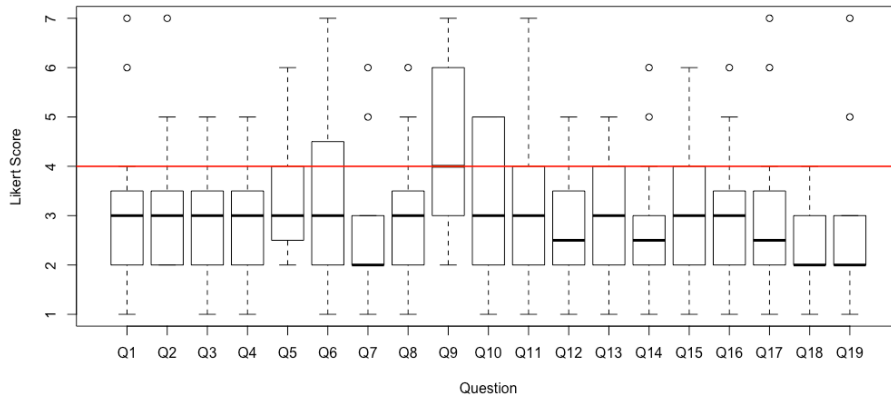
Fig. 5. Boxplot of the PSSUQ surveys

4.4 Analysis

There is a non-significant (p = 0.8) negative correlation (R=-0.068) between prior knowledge and the outcome (silver standard), from which we can conclude nothing. There were also no significant correlations between the four questions that stood out as having been problematic and task performance. If we analyze the correlations between the various PSSUQ dimensions and task performance, however, we see (from Fig. 6) that one correlation is significant (task performance vs. system usefulness with p = 0.0061) and one is almost significant (task performance vs. overall with p = 0.0019). The data thus seem to indicate that as task performance goes up, the perceived system's usefulness and overall usability are more appreciated.

If we look again at the PSSUQ in Table 3, we do notice that most values are close to 3, which is not terrific but leans towards a favorable impression. The only exception is Question 9, which has an average value of 4.38 (below 4). Three participants did not provide an opinion for that value, deeming it not applicable. Clear error messages and instructions on how to recover from errors are thus to be improved.

5 Conclusions and Future Work

In this paper, we presented an approach to interlink geospatial Linked Data datasets using maps, which to the best of our knowledge has not been tried before and therefore constitutes a contribution. A second contribution is the experiment's protocol that was devised to validate the tool. From the analysis of the experiment, we can conclude that using maps for driving manual interlinking processes is a viable approach. We also noticed that participants had more trouble with certain Linked Data principles (content negotiation), which led them to create an interlink to the wrong URIs and is in itself an interesting observation.
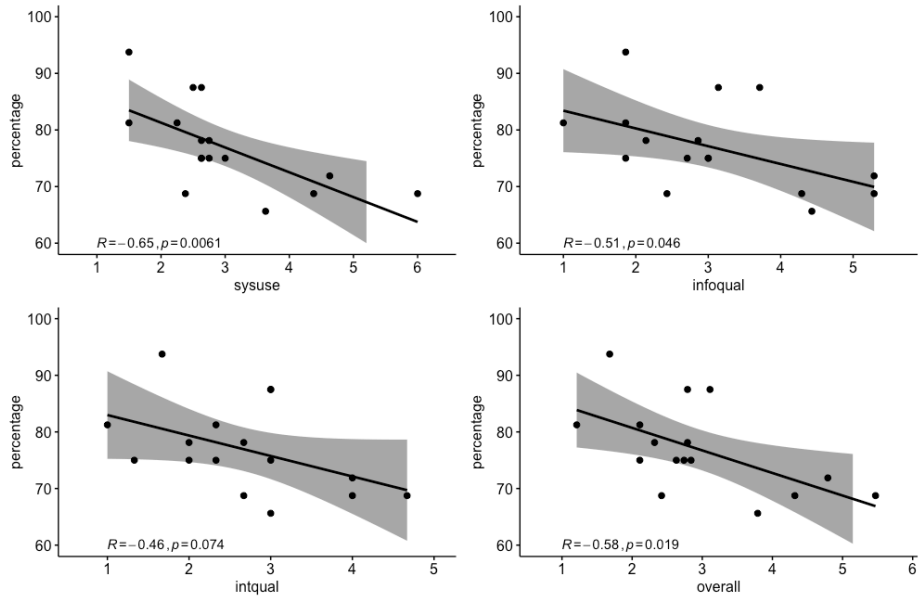
Fig. 6. Correlations between the PSSUQ dimensions and task performance according to the "silver" standard

Finally, we proposed an approach to organize and store the interlinks with provenance data using named graphs. While not validated in this study, this would provide an integration point for other initiatives in which such provenance data will be critical to assess the authoritativeness of an interlink. DaCura [4], for instance, is a platform that avails of niche-sourcing to collect annotations. DaCura keeps track of the annotations collected by annotators working on the same tasks to assess which interlinks are likely to be accurate. Our approach to knowledge organization would easily integrate with such methods.

There are a few limitations to this study. The first is the creation of our lightweight faceted browser, which may have had an impact on task performance and perceived usability. The second limitation is related to the geospatial vocabularies and types of interlinks our tool supports. The tool is only capable of processing GeoSPARQL data using WKT literals. Future work should thus look into the inclusion of GMLand other geospatial vocabularies. As for the links, the tool only supports the creation of interlinks, where the resource displayed on the map is the subject. We aim to extend the popup to support the inclusion of interlinks in which the resource on the map is the subject. That would leverage the problem of forcing users to think in terms of reverse relationships (e.g., `geo:sf-within` instead of `geo:sf-contains`).

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. Int. J. Semantic Web Inf. Syst. **5**(3), 1–22 (2009). https://doi.org/10.4018/jswis.2009081901

2. Both, A., Wauer, M., Garcìa-Rojas, A., Hladky, D., Lehmann, J.: The geoknow generator workbench - an integrated tool supporting the linked data lifecycle for enterprise usage. In: Filipowska, A., Verborgh, R., Polleres, A. (eds.) Joint Proceedings of the Posters and Demos Track of 11th International Conference on Semantic Systems - SEMANTiCS 2015 and 1st Workshop on Data Science: Methods, Technology and Applications (DSci15) 11th International Conference on Semantic Systems - SEMANTiCS 2015, Vienna, Austria, September 15-17, 2015. CEUR Workshop Proceedings, vol. 1481, pp. 92–95. CEUR-WS.org (2015)

3. Debruyne, C., Meehan, A., Clinton, E., McNerney, L., Nautiyal, A., Lavin, P., O'Sullivan, D.: Ireland's authoritative geospatial linked data. In: d'Amato, C., Fernández, M., Tamma, V.A.M., Lécué, F., Cudré-Mauroux, P., Sequeda, J.F., Lange, C., Heflin, J. (eds.) The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part II. Lecture Notes in Computer Science, vol. 10588, pp. 66–74. Springer (2017). https://doi.org/10.1007/978-3-319-68204-4_6

4. Feeney, K.C., O'Sullivan, D., Tai, W., Brennan, R.: Improving curated web-data quality with structured harvesting and assessment. Int. J. Semantic Web Inf. Syst. **10**(2), 35–62 (2014). https://doi.org/10.4018/ijswis.2014040103

5. Isele, R., Jentzsch, A., Bizer, C.: Silk server - adding missing links while consuming linked data. In: Hartig, O., Harth, A., Sequeda, J. (eds.) Proceedings of the First International Workshop on Consuming Linked Data, Shanghai, China, November 8, 2010. CEUR Workshop Proceedings, vol. 665. CEUR-WS.org (2010)

6. Isele, R., Bizer, C. Active learning of expressive linkage rules using genetic programming. Journal of Web Semantics. 23, 2-15 (2013) https://doi.org/10.1016/j.websem.2013.06.001

7. Kharlamov, E., Giacomelli, L., Sherkhonov, E., Grau, B.C., Kostylev, E.V., Horrocks, I.: Semfacet: Making hard faceted search easier. In: Lim, E., Winslett, M., Sanderson, M., Fu, A.W., Sun, J., Culpepper, J.S., Lo, E., Ho, J.C., Donato, D., Agrawal, R., Zheng, Y., Castillo, C., Sun, A., Tseng, V.S., Li, C. (eds.) Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017. pp. 2475–2478. ACM (2017). https://doi.org/10.1145/3132847.3133192

8. Kyzirakos, K., Karpathiotakis, M., Koubarakis, M.: Strabon: A semantic geospatial DBMS. In: Cudré-Mauroux, P., Heflin, J., Sirin, E., Tudorache, T., Euzenat, J., Hauswirth, M., Parreira, J.X., Hendler, J., Schreiber, G., Bernstein, A., Blomqvist, E. (eds.) The Semantic Web - ISWC 2012 - 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part I. Lecture Notes in Computer Science, vol. 7649, pp. 295–311. Springer (2012). https://doi.org/10.1007/978-3-642-35176-1_19

9. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. Semantic Web **6**(2), 167–195 (2015). https://doi.org/10.3233/SW-140134

10. Lewis, J.R.: Psychometric evaluation of the PSSUQ using data from five years of usability studies. International Journal of Human-Computer Interaction **14**(3-4), 463–488 (2002)

11. McGlinn, K., Blake, D., O'Sullivan, D.: Gviz - an interactive webapp to support geosparql over integrated building information. In: Amer-Yahia, S., Mahdian, M., Goel, A., Houben, G., Lerman, K., McAuley, J.J., Baeza-Yates, R.A., Zia, L. (eds.) Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019. pp. 904–912. ACM (2019). https://doi.org/10.1145/3308560.3316536

12. McGlinn, K., Debruyne, C., McNerney, L., O'Sullivan, D.: Integrating Ireland's geospatial information to provide authoritative building information models. In: Proceedings of the 13th International Conference on Semantic Systems, SEMANTICS 2017, Amsterdam, The Netherlands, September 11-14, 2017. pp. 57–64 (2017). https://doi.org/10.1145/3132218.3132223

13. McKenna, L., Debruyne, C., O'Sullivan, D.: Understanding the position of information professionals with regards to linked data: A survey of libraries, archives and museums. In: Chen, J., Gonçalves, M.A., Allen, J.M., Fox, E.A., Kan, M., Petras, V. (eds.) Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, JCDL 2018, Fort Worth, TX, USA, June 03-07, 2018. pp. 7–16. ACM (2018). https://doi.org/10.1145/3197026.3197041

14. Mori, G., Paternò, F., Santoro, C.: CTTE: support for developing and analyzing task models for interactive system design. IEEE Trans. Software Eng. **28**(8), 797–813 (2002). https://doi.org/10.1109/TSE.2002.1027801

15. Ngomo, A.N., Auer, S.: LIMES - A time-efficient approach for large-scale link discovery on the web of data. In: Walsh, T. (ed.) IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011. pp. 2312–2317. IJCAI/AAAI (2011). https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-385

16. Ngomo, A.N., Lyko, K.: EAGLE: efficient active learning of link specifications using genetic programming. In: Simperl, E., Cimiano, P., Polleres, A., Corcho, Ó., Presutti, V. (eds.) The Semantic Web: Research and Applications - 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012. Proceedings. Lecture Notes in Computer Science, vol. 7295, pp. 149–163. Springer (2012). https://doi.org/10.1007/978-3-642-30284-8_17

17. Paternò, F., Mancini, C., Meniconi, S.: Concurtasktrees: A diagrammatic notation for specifying task models. In: Howard, S., Hammond, J., Lindgaard, G. (eds.) Human-Computer Interaction, INTERACT '97, IFIP TC13 International Conference on Human-Computer Interaction, 14th-18th July 1997, Sydney, Australia. IFIP Conference Proceedings, vol. 96, pp. 362–369. Chapman & Hall (1997)

18. Patroumpas, K., Alexakis, M., Giannopoulos, G., Athanasiou, S.: Triplegeo: an ETL tool for transforming geospatial data into RDF triples. In: Proceedings of the Workshops of the EDBT/ICDT 2014 Joint Conference (EDBT/ICDT 2014), Athens, Greece, March 28, 2014. pp. 275–278 (2014)

19. Rietveld, L., Hoekstra, R.: The YASGUI family of SPARQL clients. Semantic Web **8**(3), 373–383 (2017). https://doi.org/10.3233/SW-150197

20. Ryan, C., Grant, R., Carragáin, E.Ó., Collins, S., Decker, S., Lopes, N.: Linked data authority records for Irish place names. Int. J. on Digital Libraries **15**(2-4), 73–85 (2015). https://doi.org/10.1007/s00799-014-0129-8

21. Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., Schraefel, M.C.: Linked open government data: Lessons from data.gov.uk. IEEE Intelligent Systems **27**(3), 16–24 (2012). https://doi.org/10.1109/MIS.2012.23

22. Sherif, M.A., Ngomo, A.N., Lehmann, J.: Automating RDF dataset transformation and enrichment. In: Gandon, F., Sabou, M., Sack, H., d'Amato, C., Cudré-Mauroux, P., Zimmermann, A. (eds.) The Semantic Web. Latest Advances and New Domains - 12th European Semantic Web Conference, ESWC 2015, Portoroz, Slovenia, May 31 - June 4, 2015. Proceedings. Lecture Notes in Computer Science, vol. 9088, pp. 371–387. Springer (2015). https://doi.org/10.1007/978-3-319-18818-8_23

23. Sherif, M.A., Ngomo, A.N., Lehmann, J.: Wombat – A Generalization Approach for Automatic Link Discovery. In: Blomqvist, E., Maynard, D., Gangemi, A., Hoekstra, R., Hitzler,

P., Hartig, O. (eds.) The Semantic Web - 14th European Semantic Web Conference, ESWC 2017, Portorož, Slovenia, May 28 – June 1, 2017, Proceedings, Part I. Lecture Notes in Computer Science, vol. 10249, pp. 103–119. Springer (2017). https://doi.org/10.1007/978-3-319-58068-5_7

24. Stadler, C., Martin, M., Auer, S.: Exploring the web of spatial data with Facete. In: Chung, C., Broder, A.Z., Shim, K., Suel, T. (eds.) 23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014, Companion Volume. pp. 175–178. ACM (2014). https://doi.org/10.1145/2567948.2577022

25. Zhao, J., Hartig, O.: Towards interoperable provenance publication on the linked data web. In: Bizer, C., Heath, T., Berners-Lee, T., Hausenblas, M. (eds.) WWW2012 Workshop on Linked Data on the Web, Lyon, France, 16 April, 2012. CEUR Workshop Proceedings, vol. 937. CEUR-WS.org (2012)

A. PSSUQ Questionnaire

1. Overall, I am satisfied with how easy it is to use this system
2. It was simple to use this system
3. I could effectively complete the tasks and scenarios using this system
4. I was able to complete the tasks and scenarios quickly using this system
5. I was able to efficiently complete the tasks and scenarios using this system
6. I felt comfortable using this system
7. It was easy to learn to use this system
8. I believe I could become productive quickly using this system
9. The system gave error messages that clearly told me how to fix problems
10. Whenever I made a mistake using the system, I could recover easily and quickly
11. The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear
12. It was easy to find the information I needed
13. The information provided for the system was easy to understand
14. The information was effective in helping me complete the tasks and scenarios
15. The organization of information on the system screens was clear
16. The interface of this system was pleasant
17. I liked using the interface of this system
18. This system has all the functions and capabilities I expect it to have
19. Overall, I am satisfied with this system