

# Reviews of GeoLD 2018 accepted papers

Part of the joint proceedings of GeoLD-QuWeDa-2018 published at CEUR-WS

## 1. Ali Khalili, Peter van den Besselaar and Klaas Andries de Graaf. Using Linked Open Geo Boundaries for Adaptive Delineation of Functional Urban Areas

Review 1	
PC member:	anonymous
Reviewer's confidence:	4: (high)
Relevance:	5: (excellent)
Impact of ideas and results:	3: (fair)
Clarity and quality of writing:	2: (poor)
Related work:	1: (very poor)
Implementation and soundness:	4: (good)
Evaluation:	3: (fair)
Overall evaluation:	-2: (reject) <p>With 15 pages the paper exceeds the official maximum page count for long papers (12 pages) in the CFP. This should lead to the rejection of this paper. However, I read the paper anyways and want to provide my feedback on the work.</p> <p>In their work, the authors address the delineation of functional urban areas (FUAs) based on open data. In contrast to the rigid definition of FUAs provided by the OECD and EC, they aim to provide means to adaptively define FUAs by weighting relevant factors differently. In the motivation, the authors claim that governments need a way to dynamically redefine different types of urban areas. However, the citation provided does not support that claim. The authors should either elaborate on that statement or provide a source supporting the claim.</p> <p>A contribution mentioned by the authors is the reconstruction of FUAs defined by the OECD based on open data sources. However, a presentation of how this reconstruction is achieved is missing as well as an evaluation of how well the OECD FUAs can be reconstructed by open data sources. It would be interesting to see, how these areas differ and why they differ. Furthermore, the authors argue that the access to the OECD shapefiles is limited and requires negotiation, however, the data seems to be openly available at: <a href="http://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm">http://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm</a> .</p> <p>Unfortunately, the paper does not present related work which does not allow for the comparison of the work to previous research endeavors. A good starting point for a literature review in this area would be "The semantics of populations: A city indicator perspective".</p>

The authors present the integration of different open geo dataset and the linking between the data sets in detail. The datasets provide data on administrative boundaries on several levels which should be used to reconstruct FUAs. Furthermore, they present applications (SPARQL endpoint and an API) to access the integrated data. The data is integrated by transforming all datasets into GeoJSON and using a Mapping and Enrichment Function to transform the data to RDF. Details on the Mapping and Ontology used for the enrichment are not provided.

The authors present two case studies. The first one evaluates the usability of a tool which allows for geocoding addresses with the different administrative boundaries from the datasets in a spreadsheet. The result of 10 participants filling out the questionnaire revealed a decent usability of the tool, however, the relation to how this allows for dynamically defining FUAs is not obvious.

The second case study applies the tools to study the number of projects funded by the Netherlands Enterprise Agency in different FUAs. For this purpose, they use the OpenStreetMap data in combination with statistical data provided by the Netherlands to define different adaptive FUAs (based on population, businesses and both) and also use the FUAs defined by the OECD. For these FUAs they compare the number of projects and show that the OECD FUAs do not capture all relevant areas. The process how the FUAs in the second use case is derived from the OSM and combined with the statistics data is not presented thoroughly enough.

Overall, the authors present a pipeline for integrating geodata from different data sources and provide means to query the data. The data is used to define FUAs based on the combination of different datasets on administrative areas and statistical data. A detailed description of methodologies to define these areas and for which use cases which methodology may be applied is not provided. Furthermore, they present two use cases of the implementation and investigate different research questions with the help of these tools. The implementations and demos of them are provided online. Unfortunately, some demos of the implementations are merely available for registered user or with a valid API key, which somehow is in contrast to making the data openly available.

#### Layout/Writing:

- Page count (15 pages) exceeds the Long Paper page limit of 12 pages
- Missing page count in the PDF
- Footnote spacing inconsistent (sometimes there is a space before the footnote)
- Typos:
  - \* p.3: shpapefiles
  - \* p.6: Virtuosos
  - \* p.10: Fig. 4 An screenshot
  - \* p.11: seem --> seems

## Review 2

PC member:	Kleanthi Georgala
Reviewer's confidence:	4: (high)
Relevance:	5: (excellent)
Impact of ideas and results:	5: (excellent)
Clarity and quality of writing:	5: (excellent)
Related work:	5: (excellent)
Implementation and soundness:	5: (excellent)
Evaluation:	5: (excellent)
Overall evaluation:	<p>3: (strong accept)</p> <p>This paper introduces a dynamic approach of classifying urban areas using openly available spatial and non-spatial resources on the Web. The authors argue the current static factors of the Functional Urban Areas (FUAs) notion and propose a dynamically defining FUAs based on linked open data.</p> <p>The paper is clear and reads well. The use of English is advanced and highly scientific. The structure of the paper is well defined and helps the reader go through the paper, even if he lacks of previous knowledge in the domain.</p> <p>Their data discovery and collection step includes a brief but compact overview of the datasets in use with the right citations. Their Data Extraction &amp; Conversion step is very well written and to the point. Their Data Storage &amp; Querying is also brief but well explained. Regarding the 4th step of Data Linking, I have some concerns. Even though they describe well the query they used for linking, querying an endpoint in order to perform linking is not the most suitable method when it comes to link discovery. There are plenty of LD frameworks such as LIMES and SILK that can perform geospatial data linking with high accuracy and completeness. The remaining steps are also well described.</p> <p>Finally, their In-Use Case Study and Evaluation is well written and explained. I have no major comments regarding their evaluation procedure and how they explained the results.</p> <p>This paper needs no further improvement from my side. The authors know the problem at hand very well and their work is high quality. However, since the upper limit of the paper is 12, I would suggest that they should try to minimize the size of the paper.</p>

## Review 3

PC member:	anonymous
Reviewer's confidence:	4: (high)
Relevance:	5: (excellent)
Impact of ideas and	4: (good)

<i>results:</i>	
<i>Clarity and quality of writing:</i>	<b>3:</b> (fair)
<i>Related work:</i>	<b>2:</b> (poor)
<i>Implementation and soundness:</i>	<b>4:</b> (good)
<i>Evaluation:</i>	<b>3:</b> (fair)
<i>Overall evaluation:</i>	<p><b>1:</b> (weak accept)</p> <p>The paper proposes an implementation for managing and combining different linked open geo-data sources together with non-spatial resources on the Web and an approach (and implementation) for identifying and delineating Functional Urban Areas (FUAs). The framework for the management of linked geo-data covers all the phases of the LOD Lifecycle (Auer et al.) and the paper is structured in a way that follows these phases (or steps) too.</p> <p>Interesting contribution is the proposed approach for dynamically defining FUAs based on open data such as Flickr, OSM, GADM, etc. transformed into Linked Data and interlinked with LOD data sources such as DBpedia and Wikidata.</p> <p>The applications built on top of this infrastructure are also quite interesting and well designed, as shown in the screenshots provided and in the link to the online demos.</p> <p>These are validated with 2 use case studies which are clearly described at the end of the paper.</p> <p>The paper is well written and clear, despite some typos and grammatical errors.</p> <p>The paper goes well beyond the pages limitation for this workshop, and should be penalized for that. However, it is a very interesting contribution for the workshop and its audience and should be accepted.</p> <p>A short section for comparison with SOTA approaches and projects is missing and would be beneficial.</p> <p>In the usability evaluation section it is written that 20 users took part in the study but only 10 completed the survey, why? Also, in Fig 5., the answers to questions such as the 3rd one and the 9th one highlight some negative results, this should be explained.</p> <p>Section 9 could be shortened, as well as the datasets descriptions in section 2.</p> <p>Minor issues:</p> <p>Add a space before some reference numbers in the text.</p> <p>In Section 1: "OpenStreetMap3 is already interlinked" -&gt; "OpenStreetMap3 are already interlinked" "cloud and provides the opportunity" -&gt; "cloud and provide the opportunity"</p> <p>Section 2:</p>

"shpapefiles" -> "shapefiles"

Section 2.1: add spaces before the "(" characters.

Section 3: "spacial" -> "spatial"

Section 4: "Virtuosós" -> "Virtuoso"

"specific addres"->"specific address"

Section 8: "20 particiapants" -> "20 participants"

Footnote 43 is a duplicate.

## 2. Alan Meehan, Kaniz Fatema, Rob Brennan, Eamonn Clinton, Lorraine McNerney and Declan O'Sullivan. License and Template Access Control for Geospatial Linked Data

Review 1	
PC member:	Ivan Ermilov
Reviewer's confidence:	4: (high)
Relevance:	5: (excellent)
Impact of ideas and results:	3: (fair)
Clarity and quality of writing:	4: (good)
Related work:	2: (poor)
Implementation and soundness:	4: (good)
Evaluation:	1: (very poor)
Overall evaluation:	<p>0: (borderline paper)</p> <p>This paper provides the description of an access control implementation for Ordnance Survey Ireland (OSi). It describes an engineering solution for the access control problem from OSi.</p> <p>There is no comparison with existing access control models nor critical analysis of the implemented access control model. The source code is also not made available for the general public. The main contribution of the paper is the technical description of the use case, which the authors have implemented for OSi. It is relevant to the topic of workshop and is a borderline paper in my opinion.</p> <p>The paper describes an authorization model for RDF data based on SPARQL query templates. The templates have to be created by an administrator depending on a license model. The access control solution proposed in this work masquerades SPARQL interface with a custom RESTful API. This takes away the benefits of using RDF from the users in favor of providing fine-granular security model for the data provider.</p> <p>The motivation simply comes from the necessity of OSi to have an authorization model for its' users. The benefits of such system for other use cases are not outlined. Could it be used in general for geospatial data providers? If yes, then what are the benefits and drawbacks of such a solution?</p> <p>The authors provide a related work section, but do not compare their work with the existing access control models. Moreover, they claim that "Our access control approach is specialized for retrieving fine grain geospatial instance data, which can exploit GeoSPARQL functions (if an administrator creates a query to do so), while existing approaches do not offer such specific access control to data. For example, the security mechanisms in Apache Rya [1] allows fine granular authorization on per triple basis. In this case, the user can be restricted access to only a part of the graph by labeling necessary set of the triples. Then for the same SPARQL query, different users will have</p>

different responses based on their rights. It would be good to see why existing access control models were not used for this use case.

The questions, which are not answered in the paper:

- How much time does it take to create templates based on a license or a user?
- What is the impact of such system on the users in comparison of using SPARQL?

Other remarks:

- Figure 2 should be visualized (the same as figure 1)
- on page 8: Figure 7 shows a high architecture... --> should be Figure 5
- on page 8: example of a status call... --> do not include protocol and domain in the status call, i.e. /acon/status/{userID}

[1] Punnoose, Roshan, Adina Crainiceanu, and David Rapp. "SPARQL in the cloud using Rya." Information Systems 48 (2015): 181-195.

## Review 2

PC member:	Matthias Wauer
Reviewer's confidence:	4: (high)
Relevance:	4: (good)
Impact of ideas and results:	4: (good)
Clarity and quality of writing:	3: (fair)
Related work:	3: (fair)
Implementation and soundness:	3: (fair)
Evaluation:	2: (poor)
Overall evaluation:	1: (weak accept) What I would expect after the abstract: The work provides three contributions to the problem of access control to geospatial Linked Data: 1) an access control model including a vocabulary for describing a "licence" defining what data can be accessed and a "template" for actually accessing data, 2) an architecture for implementing this access control method using a proxy service, and 3) a case study of a prototypical implementation  What I expect after the introduction: The introduction clearly presents the research question. However, it fails to motivate 1) why there is a need for fine grained data access control, and 2) how a geospatial data retrieval scenario differs from generic data access. Further, while motivating LD with enabling linking to other datasets etc.) it is uncertain at this point if customers using a RESTful API will still have this benefit.

Summary (following the procedure described in <https://violentmetaphors.com/2013/08/25/how-to-read-and-understand-a-scientific-paper-2>):

Identify the BIG QUESTION.

→ How to enable fine-grained access control for geospatial Linked Data.

Summarize the background in five sentences or less.

→ The related work section presents many general access control approaches and methods specific to RDF data. While not particularly experienced in this specific field of research, many of the cited works appear to be in an early state (several workshop and “Towards...” papers).

Identify the SPECIFIC QUESTION(S).

→ How to model access rights to geospatial Linked Data. How to enforce access restrictions to geospatial Linked Data.

Identify the approach.

→ A separate licence and access model. Apparently the access model is a combination of a “facade” (the template) in addition to licence based checks.

Read the methods section.

→ Section 4 provides a comprehensive explanation of the proposed approach. Unfortunately the link to the vocabulary is not accessible. A few details are left out, e.g., whether the geographical feature classes also support using other geospatial data types, such as multipolygons.

Read the results section. Write one or more paragraphs to summarize the results.

→ The authors claim that the proposed architecture and vocabulary are able to represent the access restrictions given by the requirements. The method by which these requirements were gathered are not explained. A closed-source prototypical implementation was done to realize the proposed concept. While the authors stress the benefits of Linked Data in the introduction, they propose access only through a REST API in their architecture. While the result set to such a request may still include URIs that can be dereferenced, I’m wondering if an approach like GraphQL might be more reasonable here.

Do the results answer the SPECIFIC QUESTION(S)? What do you think they mean?

→ Question 1 is likely solved by the given approach within certain limitations. For example, if a scenario requires a single query with two distinct values of a certain type (e.g., radius) the licence model can’t distinguish between those. Question 2 was already clarified in Section 2, and appears to be answered sufficiently by the proposed concept.

Now, go back to the beginning and read the abstract. Does it match what the authors said in the paper? Does it fit with your interpretation of the paper?

→ Yes it does.

Reasons to accept:

- novel approach to access control for geospatial linked data, relevant to the workshop
- real use case and application
- comprehensive description of vocabulary and architecture
- well-structured paper

Reasons to reject:

- vocabulary URL not accessible
- very limited evaluation
- some implementation details omitted (e.g., any information about the result format/content of the status call)
- paper clearly hasn't been re-read before submission (see suggestions)

Improvement suggestions:

Clarify that the licence AND template are both necessary to enforce access control (in the example URI above Fig. 4 I first assumed the template ID might be sufficient because the system could resolve itself which licence would fit, but this is not necessarily the case).

Discussion of potential alternative approaches and their drawbacks compared to this licence+template approach would be interesting.

Writing:

abstract:

"an implementation architecture ... which implements..."

introduction:

"ISU is ... and \_are\_ tasked"

next line: "...geospatial data -- ..." (remove long dash?)

"..., and hence \_ an access control..." (missing word?)

"...and flexible enough to \_be\_ meet the (potential)..." (remove)

"The remained of this paper..." (remainder?)

"Section 2 presents \_some\_ geospatial access control ... from \_a\_ geospatial data organization" (imprecise, better: be more specific)

requirements:

"For example, OSI's building... in the country." (this is not a sentence.)

"...construction companies \_ utility companies \_ etc." (missing commas)

related work:

"Role based access control [7]\_,\_ is..." (remove comma)

"Steyskal and Polleres [9] \_examines\_ ..." (plural)

"Such policies \_specify specific\_ ... to \_specific\_ datasets..." (repetition)

"Our approach models ... Our approach..." (repetition)

approach:

"... and then present \_ proposed ..." (missing word?)

4.1:

URL not dereferencable, and should be provided as footnote instead of inline

"...a user will be able \_ access." (missing word)

"...a floating point number of \_what\_ the permitted radius." (remove)

Fig. 2 line 16: "geohiveb:Building" looks like a very unusual URI with the given prefix

4.2:  
 "...and \_the\_ a Query Processor..." (remove)  
 case study:  
 "...would correctly reject a query call \_then\_ when values... allowed \_to\_ according to..." (remove)  
 "...did not contain any \_did\_ data that..." (remove)  
 conclusion:  
 "...also allows \_provides\_ the specification..." (remove)  
 "...how well it would \_fair\_ in situations..." (fare?)  
 references:  
 "Ireland? s Authoritative..."  
 "21-Februrary-..."

### Review 3

PC member:	anonymous
Reviewer's confidence:	4: (high)
Relevance:	4: (good)
Impact of ideas and results:	3: (fair)
Clarity and quality of writing:	3: (fair)
Related work:	4: (good)
Implementation and soundness:	2: (poor)
Evaluation:	1: (very poor)
Overall evaluation:	<p>0: (borderline paper)</p> <p>The paper presents an approach for providing different levels of access control to geographical data. In particular, the approach propose a vocabulary which represents different levels of licensing to indicate different access control to the users. In addition, they provide a template to indicate how the access control can be provided to the users.</p> <p>I find this work interesting and quite relevant to the workshop although I have been used to see licensing as a process of associating one access control to the data for all users. The idea of providing different types of licenses to the same dataset for different users is a different way of exploiting licensing.</p> <p>Except for the above observation, I have a concern that is related to the experimentation. Although this is an initial work I would have expected some initial results that can be relevant to be presented and discussed in this workshop which to me are missing.</p> <p>correct link <a href="http://ontologies.geohive.ie/">http://ontologies.geohive.ie/</a> -&gt; <a href="http://ontologies.geohive.ie/osi/index.html">http://ontologies.geohive.ie/osi/index.html</a>. why will it be hosted in the future? why is it not yet hosted in this domain?</p> <p>Although the paper is relevant to the workshop, I believe that the paper is not well written and English need to be revised. In the following, I will provide a few examples of minor comments which do not represent an</p>

exhaustive list of errors. So please revise the whole paper according to similar examples.

straight forward -> straightforward

valid up until -> valid up to

retrieve data, but -> retrieve data but

will be able access -> will be able to access

for a templates variables -> for templates variables

### 3. Finn Årup Nielsen, Daniel Mietchen and Egon Willighagen. Geospatial Data and Scholia

Review 1	
PC member:	anonymous
Reviewer's confidence:	3: (medium)
Relevance:	5: (excellent)
Impact of ideas and results:	2: (poor)
Clarity and quality of writing:	2: (poor)
Related work:	3: (fair)
Implementation and soundness:	3: (fair)
Evaluation:	2: (poor)
Overall evaluation:	<p><b>0: (borderline paper)</b>            This paper presents an extension to Scholia, the authors previous work presented in [5]. I like the topic of the paper. However, I have the following concerns needs to be addressed in the final camera-ready paper.</p> <ol style="list-style-type: none"> <li>1. Motivation and story: The paper missing a clear connecting story and motivation.</li> <li>2. Contributions: I was not able to detect what are the exact contributions of this paper.</li> <li>3. User stories are presented in section 4. I think they can be handled by original Scholia as well? As such what are the stories that can only be handled due to the new extension?</li> <li>4. The paper needs a through proof reading.</li> </ol> <p>I encourage authors to continue their good in this direction.</p>
Confidential remarks for the program committee:	I am not expert in this topic and happy to change my score after discussion.

Review 2	
PC member:	anonymous
Reviewer's confidence:	3: (medium)
Relevance:	4: (good)
Impact of ideas and results:	3: (fair)
Clarity and quality of writing:	4: (good)
Related work:	3: (fair)
Implementatio	4: (good)

<i>n and soundness:</i>	
<i>Evaluation:</i>	<b>1:</b> (very poor)
<i>Overall evaluation:</i>	<p><b>1:</b> (weak accept)</p> <p>The paper presents Scholia, a web site that visualizes bibliographic and scientific datasets contained in WikiCite/Wikidata. A part of the visualization instruments in Scholia regards geospatial aspects of the data and, thus, are map-based. The authors provide a good overview of what type of visualizations Scholia can support. However, they could elaborate more on how these visualizations can be achieved, i.e. how the underlying data can be queried or browsed.</p> <p>Specifically, the authors just mention the WDQS functionality, but not elaborate on how this can be exploited (e.g. through <a href="https://query.wikidata.org/">https://query.wikidata.org/</a>) to build all the examples presented in the latter parts of the paper; all these examples are just provided as fixed links, resembling pre-calculated and stored views for selected queries, on the database contents. The respective SPARQL queries are not present in the paper (of course expectedly, due to lack of space) but they are also not visible at the respective web pages that comprise the examples of the paper (at least after a quick look). Eventually, it is not clear for the reader how easily a user can browse-construct the exemplary views in the paper, from scratch.</p> <p>The authors should further elaborate a bit more on the architecture of the system and its scalability potential. The paper (and the contained examples) leave the impression that Wikidata (and all its affiliated initiatives) are still ongoing efforts, currently including a small amount of data compared to other initiatives (e.g. DBpedia). Even so, some of the contained links in the paper require considerable time to fully load, despite the very small amount of data they eventually visualize. The authors should elaborate on what are the efficiency/scalability plans for the time when Wikidata gathers orders of magnitude more data.</p> <p>Finally, I would like to have seen, in the related work, a small discussion as to what new is contributed by Wikidata (and the affiliated sites/software) compared to existing/established initiatives.</p>

### Review 3

<i>PC member:</i>	anonymous
<i>Reviewer's confidence:</i>	<b>5:</b> (expert)
<i>Relevance:</i>	<b>5:</b> (excellent)
<i>Impact of ideas and results:</i>	<b>3:</b> (fair)
<i>Clarity and quality of writing:</i>	<b>2:</b> (poor)
<i>Related work:</i>	<b>5:</b> (excellent)
<i>Implementation and soundness:</i>	<b>2:</b> (poor)
<i>Evaluation:</i>	<b>3:</b> (fair)
<i>Overall evaluation:</i>	<b>0:</b> (borderline paper) *General comments and Summary*

Scholia caters information about scientific works from Wikidata by querying the Wikidata Query Service. The paper mostly describes a set of user stories to illustrate the applicability of Scholia for the consumption of Geospatial data using Wikidata. While the paper does not produce any technical contribution per se, it presents interesting user stories illustrating the question answering capability of Scholia on geospatial data.

The system -- Scholia is built mostly (or dominantly) around handling “nearby” queries using the geof:distance function. While the application of the system is of great value, the coverage of the system in terms of GeoSPARQL query features is a concern.

The presented system -- Scholia is novel, however, the method/technique using which these queries (representing users interests or information need) are generated is not mentioned clearly. Whether this process is automatic or manual is also of interest to a reader. At this point the system appears to be too static or in an extremely early stage to be of benefit for an interested user (the user is not an expert in SPARQL or GeoSPARQL for that matter).

#### \*Introduction\*

Pg 1, Paragraph 2, Last line -> “whereas, in GeoSPARQL, the function with the same name takes a third argument for the unit.” => which one? And how is it different from geof:distance?

#### \*User stories\*

While the user stories are narrative, the GeoSPARQL queries used to generate the answers are not mentioned. This is not mission critical but would be very interesting to see/have. Though one can find the query from the mentioned links (from below the results page -> edit this query). I wonder whether these queries were handcrafted or generated automatically. In the case of handcrafted queries, it would be near impossible for users who are not experts in GeoSPARQL to be able to use the Scholia with any ease. If these queries are automatically generated based on the user interests, it should be mentioned how and where.

No map is generated for story 1 -

<https://tools.wmflabs.org/scholia/country/Q33/topic/Q2539> which would be ideal for the geospatial aspect of knowing from which city/area is a particular researcher, or at least the location of the research lab/group. Also, the reason for co-citation graph is reported to be empty but a clear reason is not mentioned.

No map is generated for the story 2 -

<https://tools.wmflabs.org/scholia/location/Q1748/topic/Q2539>. Also, no additional graphs or tables are reported apart from the list of authors in the descending order of their scores. The scoring function, is to be assumed to be based on the distance from the center of the Copenhagen? In general, this

case inclines more towards a regular (non-GeoSPARQL) user story than geospatial.

No map is generated for the story 3 -

<https://tools.wmflabs.org/scholia/location/Q3806/topic/Q52> . Why is this? Also similar case as of user story 3. Furthermore, “links do not work” is reported in stories 2 and 3, which is worth investigating.

User story 4 is quite interesting, however, appears to be flawed. While the story description states “relevant scientific meetings”, the results wrt to WWW conference visitors report rather irrelevant points of interests in terms of people and publications, such as (1) Alessandro Vespignani Italian physicist, (2) Luciano Floridi Italian philosopher and (3) Combining Participatory Influenza Surveillance with Modeling and Forecasting: Three Alternative Approaches, etc. Also, no map is generated highlighting in which part of the city or nearby suburbs are these events taking place.

For story 5, the administrator would be more interested in observing the patterns from Denmark to South Korea and not the reverse.

#### 4. Peru Bhardwaj, Christophe Debruyne and Declan O'Sullivan. On the Overlooked Challenges of Link Discovery

Review 1	
PC member:	Mohamed Sherif
Reviewer's confidence:	5: (expert)
Relevance:	4: (good)
Impact of ideas and results:	3: (fair)
Clarity and quality of writing:	5: (excellent)
Related work:	2: (poor)
Implementation and soundness:	3: (fair)
Evaluation:	1: (very poor)
Overall evaluation:	<p><b>0:</b> (borderline paper)</p> <p>The paper highlight four lessons learned during the process of interlinking Ordnance Survey Ireland (OSi) and DBpedia. The first lesson was concerning the difficulty of ontology matching of the datasets to be interlinked. The second lesson was regarding the incomplete results of SPARQL endpoints. The third lesson was concerning finding proper measures for comparing resources. Finally, the fourth lesson is pertaining to the difficulty of finding proper link specification.</p> <p>The paper is well structured and written in good English, which ease the understanding of the paper. I am not completely agree with the authors pertaining to some points they claim in the paper. Here, I list my points that could help in extending this paper (may be a to a complete resource paper or a dataset paper for SWJ):</p> <p>Lesson 1: The problem here is well known in the literature as the ontology matching problem. There is a massive amount of work to deal with his problem. For example see the DL-Learner paper [a]</p> <p>Lesson 2: I think the problem here was that you tried to issue 1 SPARQL query to get the whole result set in one go, which is only possible if your result set is less than the internal endpoint limit (let us assume it is n). In case your result set is greater than the endpoint internal limit n, you get only n random resources. To overcome such a limit, you normally ask the endpoint programmatically using some paging techniques (using the LIMIT and OFFSET in your SELECT query). See (<a href="https://github.com/SmartDataAnalytics/jena-sparql-api">https://github.com/SmartDataAnalytics/jena-sparql-api</a>).</p> <p>Lesson 3 and 4: I do not agree with your claim in Section 4.2 “Techniques to learn a link specification [7] proved to be unusable in our case study due to the non-overlapping nature of the two datasets to be interlinked.”. Have you tried to use machine learning in LIMES [c] to do your task? WOMBAT (the most recent machine learning approach in LIMES) provide supervised, active and unsupervised versions to do your task. I think even the unsupervised version of WOMBAT should give you good results. If not, try</p>

the supervised version of WOMBAT.

Table 1: The number of returned links have no meaning without evaluating them. I know, that evaluating all the returned links would be a tedious task. Instead, you could pick a random 100 links to evaluate and compute accuracy for each of the measure. The same hold if you use WOMBAT and you want to evaluate its results.

Other minor comments:

- In Section 2, it would be good to reference [b] together with the original LIMES paper to emphasis the superiority of LIMES performance
- In Section 4.2, you reference [8], while no such a reference exists

[a] DL-Learner - A framework for inductive learning on the Semantic Web by Lorenz Bühmann, Jens Lehmann, and Patrick Westphal in Web Semantics: Science, Services and Agents on the World Wide Web

[b] RADON - Rapid Discovery of Topological Relations by Mohamed Ahmed Sherif, Kevin Dreßler, Panayiotis Smeros, and Axel-Cyrille Ngonga Ngomo in Proceedings of The Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

[c] WOMBAT - A Generalization Approach for Automatic Link Discovery by Mohamed Ahmed Sherif, Axel-Cyrille Ngonga Ngomo, and Jens Lehmann in 14th Extended Semantic Web Conference, Portoroz, Slovenia, 28th May - 1st June 2017

## Review 2

PC member:	Kleanthi Georgala
Reviewer's confidence:	5: (expert)
Relevance:	4: (good)
Impact of ideas and results:	2: (poor)
Clarity and quality of writing:	3: (fair)
Related work:	2: (poor)
Implementation and soundness:	3: (fair)
Evaluation:	2: (poor)
Overall evaluation:	<b>0:</b> (borderline paper) This paper gives an overview of the existing challenges of linking geospatial datasets. The authors begin by explaining the problem at hand, focusing on the pre-processing phase of Link Discovery (LD), thus the title of the paper itself does not describes well the content of their work. They continue by describing the issues involved with linking the OSi and DBpedia geospatial datasets. Then, they give a detailed overview of how they identified the needed resources from the two KBs and how they derived the datasets from the endpoints,

along with the challenges they faced. Finally, they describe how they selected properties for matching and similarities measures to perform linking, along with the challenges they faced when using a declarative LD framework.

Comments for each section:

Abstract

a case study in -> a case study of

The word 'experience' is not a suitable word for research. Please replace it throughout the whole document.

The abstract reads well. However, the pre-processing phase of LD is far from being overlooked. In order to identify matching properties between two datasets, there has been plenty of related work in the domain of ontology matching. Please take a look at

<http://www.ontologymatching.org/relwork.html>.

Regarding the identification of LSs in order to link datasets, the framework you used (LIMES) incorporates 3 machine learning algorithms (wombat simple and complete, eagle) that can produce appropriate LSs.

Introduction

Provide citation in the first sentence.

described in two parts -> divided in two parts

The introduction reads well. For further comments, re-read my comments for your abstract.

Section 2

I have no comments for Section 2. The preliminaries are well described and so are the challenges the authors faced.

Section 3.1

Please try to avoid phrases like 'picking up' and 'figured out'. They are quite informal for a scientific document.

The 'Counties\_of\_the\_Republic\_of\_Ireland' is not present in DBpedia. Its equivalent is 'Counties\_towns\_in\_the\_Republic\_of\_Ireland'.

You mention 5 different ways, however I read only 4. Also using Q\_i is not appropriate, since you are not asking questions, but making remarks.

Q\_1: How did you deal with townlands that do not have the pattern in their article category?

Q\_2 and Q\_4: How do you know that some resources are or aren't townlands?

My general comments for this section is that you tried to perform ontology matching between predicates from your KBs. The whole idea seems to be quite arbitrary to me and based solely on assumptions and insights. As I mentioned above, you could have used an ontology mapping tool to carry out this work for you.

Section 3.2

I completely disagree with Lesson 2.

1) SPARQL endpoints are not affected by which browser you use.

2) The reason you were getting different results sets has to do with how the query plan is executed internally.

3) There is no ordering in the retrieved result set of a SPARQL query unless you use the ORDER BY tag. This is not a problem of the Virtuoso or any

other endpoint. This is how SPARQL queries operate.

4) It is not possible to return more than 1M of rows in SPARQL query result set over HTTP when using Virtuoso. There is a setting `ResultSetMaxRows` in `virtuoso.ini`, but if you specify something huge, you won't have more than 1048576 rows, which is  $2^{20}$ . I believe this number is very high compared to the size of your retrieved datasets. If you run the same query from `isql`, you will get all the rows. So it is not a bug, it is a feature that has a very specific purpose.

5) If you wanted to retrieve all results with one query, you could set up Virtuoso locally and change the `ResultSetMaxRows` setting in the INI file.

#### Section 4.1

The first paragraph contradicts Lesson 3.

#### Section 4.2

You used LINES for linking. LINES incorporates 3 machine learning algorithms for learning LSs. Their unsupervised version does not require any training dataset and it is able to provide you with `owl:sameAs` links that you were targeting.

Additionally, a very quick transformation of the WKT polygon to/from WKT points could have aided your work.

#### Section 4.3

LINES have a detailed manual of how a user can use the pre-processing functions ([http://dice-](http://dice-group.github.io/LINES/user_manual/configuration_file/data_sources.html)

[group.github.io/LINES/user\\_manual/configuration\\_file/data\\_sources.html](http://dice-group.github.io/LINES/user_manual/configuration_file/data_sources.html)).

As you can see, there are plenty of examples of how to use the pre-processing functions. So your comment about not knowing that the quotes should be omitted is not reasonable if you have checked the manual.

Additionally, each atomic LS which consists of one similarity function will retrieve a similarity between 0 and 1.

so:

$0 \leq \text{similarity}_1 \leq 1$  (+)

$0 \leq \text{similarity}_2 \leq 1$  (=)

$0 \leq \text{similarity}_1 + \text{similarity}_2 \leq 2$

if you multiply the first two lines with 0.5, you get:

$0 \leq 0.5 * \text{similarity}_1 \leq 0.5$  (+)

$0 \leq 0.5 * \text{similarity}_2 \leq 0.5$  (=)

$0 \leq 0.5 * \text{similarity}_1 + 0.5 * \text{similarity}_2 \leq 1$

thus, if you have used the AND function in the correct way ([http://dice-](http://dice-group.github.io/LINES/user_manual/configuration_file/metric/metric_operations.html)  
[group.github.io/LINES/user\\_manual/configuration\\_file/metric/metric\\_operations.html](http://dice-group.github.io/LINES/user_manual/configuration_file/metric/metric_operations.html)), then your similarity will be bounded to [0,1]

My final comment on the paper is that the authors tried to assist a very interesting problem in linking geospatial datasets, however their work has not considered a lot of main points described in my comments above. It was a good initial work but it lacks insight at the problem at hand.

### Review 3

PC member:	anonymous
Reviewer's	4: (high)

<i>confidence:</i>	
<i>Relevance:</i>	4: (good)
<i>Impact of ideas and results:</i>	2: (poor)
<i>Clarity and quality of writing:</i>	4: (good)
<i>Related work:</i>	3: (fair)
<i>Implementation and soundness:</i>	3: (fair)
<i>Evaluation:</i>	3: (fair)
	1: (weak accept) The paper describes a practical case of interlinking a subset of geospatial data from Ordnance Survey Ireland (OSi) with a reference public LOD dataset (DBpedia). The subset included entities of two types: counties and townlands. The authors describe the challenges arising with the selection of relevant data subsets to match, appropriate similarity measures, and special configuration parameters of the chosen link generation engine (LIMES).
<i>Overall evaluation:</i>	Overall, the paper does not provide some novel research contribution. However, it is interesting from the point of view of a practitioner's experience with methods and tools developed by the research community. One aspect that is not fully clear to me is why the authors did not try the active learning extensions available for data interlinking tools (both LIMES and SILK as the better known ones). That could potentially save the effort of picking an appropriate string similarity measure. Overall, however, the paper shows well the issues arising with the usage of existing tools originated in the Semantic Web research community and highlights the need for the improved usability of tools.

## 5. Matthias Wauer and Axel-Cyrille Ngonga Ngomo. Towards a Semantic Message-driven Microservice Platform for Geospatial and Sensor Data

Review 1	
PC member:	anonymous
Reviewer's confidence:	4: (high)
Relevance:	5: (excellent)
Impact of ideas and results:	4: (good)
Clarity and quality of writing:	5: (excellent)
Related work:	4: (good)
Implementation and soundness:	4: (good)
Evaluation:	4: (good)
Overall evaluation:	<p>2: (accept)</p> <p>Profound description of the approach with high relevance for geo linked data. Evaluation of the approach still weak due to the early stage of the project.</p> <p>There are some remarks on the paper:</p> <ol style="list-style-type: none"> <li>1. Keywords are missing.</li> <li>2. Last paragraph of section 2 the mention of section 3 is missing</li> </ol> <p>The code snippets are not self-explaining. Can you reference some 3rd party publication or give some short explanation?</p> <p>Section 3: Do you have prove that it works? Did you already test the chain of services as explained in sec. 2.4?</p> <p>Section Acknowledgement - Please check regulations on publications for the funding program (Kommunikations-Toolbox)</p> <p>Spelling:            Page 6: "value AMQP value"            Page 9, section 4.2, 2nd paragraph: readonable -&gt; reasonable</p>

Review 2	
PC member:	anonymous
Reviewer's confidence:	3: (medium)
Relevance:	4: (good)
Impact of ideas and results:	4: (good)
Clarity and quality of writing:	4: (good)
Related work:	3: (fair)
Implementation and	2: (poor)

<i>soundness:</i>	
<i>Evaluation:</i>	<p><b>2:</b> (poor)</p> <p><b>1:</b> (weak accept)</p> <p>The paper presents a funded research project called GEISER that aims at developing a flexible and scalable platform for managing geospatial and sensor data.</p> <p>The proposed platform is based on a microservices architecture and integrates data using semantic technologies. It is based on the RabbitMQ implementation of the AMQP standard.</p> <p>The project is still at an early stage of development and the software implementation is still not complete. Hence, a proper evaluation of the platform is still not available.</p> <p>However, the paper and the project are definitely relevant and interesting for the workshop and its audience.</p> <p>The architecture of the platform is clearly described and some details of its current implementation are provided. The paper is well written and clear.</p>
<i>Overall evaluation:</i>	<p>An issue of the paper is that some sentences and claims are not properly backed up by a reference. E.g. in Introduction claims in the 2nd and 4th paragraphs should be supported by references. Similarly, in Section 2.3, 3rd paragraph, the choice of RabbitMQ as compared to Apache Kafka and other approaches is not fully explained or supported by objective evidence. As reg. the writing style, parts of Section 2 in particular, feel closer to project proposal writing rather than scientific paper writing. This could be easily tuned by adding more references or justifications for the design choices. Requirements for the platform have been collected by authors and project partners but they are not clearly discussed in the paper.</p> <p>The evaluation section is very short and only provides some pointers to papers where adopted tools have been evaluated independently. This section should be rather titled "Validation" and could be either extended a little or even included in Section 2 as a subsection.</p> <p>I am not sure the readers would be very interested in the Docker Compose configuration in Listing 1.3. I would say that probably they would be more interested in getting to know about the detailed requirements collected for the platform and how they are reflected into the design decisions. This could be interesting for the presentation at the workshop.</p> <p>Minor issues:</p> <ul style="list-style-type: none"> <li>- UnifiedViews has been replaced by Linked Pipes ETL, so its reference and comparison in Section 4.2 could be updated.</li> <li>- End of Introduction should also mention about the "Evaluation" and section 3.</li> <li>- The numbering of the Listings should be revised, not 1.x.</li> </ul>

### Review 3

<i>PC member:</i>	anonymous
<i>Reviewer's confidence:</i>	<b>4:</b> (high)
<i>Relevance:</i>	<b>5:</b> (excellent)
<i>Impact of ideas</i>	<b>4:</b> (good)

<i>and results:</i>	
<i>Clarity and quality of writing:</i>	4: (good)
<i>Related work:</i>	3: (fair)
<i>Implementation and soundness:</i>	3: (fair)
<i>Evaluation:</i>	2: (poor)
<i>Overall evaluation:</i>	<p>1: (weak accept)</p> <p>The paper addresses the need for a flexible and scalable platform for handling integration of geospatial and sensor data. This work is an ongoing research project called GEISER for the creation of a platform for extraction, transformation, interlinking and fusion of such data. I understand that the project is in its initial stage but still I would appreciate some details on how the authors intend to extract, transform and link geospatial and sensor data as well as how will they evaluate such platform.</p> <p>Strong points:</p> <ol style="list-style-type: none"> <li>1. The goal of the project is very sound and important to the community</li> <li>2. The authors present three important use cases.</li> </ol> <p>Weak Points:</p> <ol style="list-style-type: none"> <li>1. I would like to see more information about use cases such as; how do they intend to integrate geographic information. What kind of data will they integrate for the geomarketing? What kind of industrial data do they have / are willing to have and how?</li> <li>2. Authors state that due to the early stage of the project, have focused only on functional evaluation according to the requirements gathered from the use cases. However, as there is still space a reader would like to know what are the requirements.</li> <li>3. What was missing in the paper is a general workflow on how the authors intend to integrate and process the data on the fly. Maybe with an example its easier for the reader to understand the workflow. Are they willing to use Google data or Geoname / Dbpedia? What kind of social networking data will they use?</li> <li>4. In the Related Work, you may also cite the EW-SHOPP project which has a similar objective as GEISER project. You can find more information on <a href="http://www.ew-shop.eu">www.ew-shop.eu</a></li> <li>5. Authors do not provide any additional information about how scalable this platform will be and how are they willing to evaluate it.</li> </ol>