

Evaluating linked data location based services using the example of Stolpersteine

Timo Homburg^{a*}, Klaus Böhm^a, Nicole Bruhn^a, Gregor Hubrich^a

^a Mainz University Of Applied Sciences, Germany, {timo.homburg,klaus.boehm,nicole.bruhn}@hs-mainz.de

* Corresponding Author

Abstract: In this publication we introduce a linked data powered application which assists users to find so-called Stolpersteine, stones commemorating Jewish victims of the second world war. We show the feasibility of a dedicated location based service as an app using linked data resources and evaluate this approach against local data sources gathered by communities to find out if the current linked data environment can equally and/or sufficiently support an application in this knowledge domain.

Keywords: Stolperstein, linked data, app, location based service

1. Introduction

A common problem for the geospatial community is the evaluation of data sources in terms of fitness for use (Pôças et al., 2014), i.e. if the quality of the data sources is sufficient enough to support a certain use case. Data quality is commonly measured using data quality metrics customized for a specific use case (Homburg and Boochs, 2019) to quantify the fitness for use of the respective data set. We would like to test the fitness for use of (linked) open data sources representing Stolpersteine (Franke and Demmig, 2009), stones commemorating Jewish victims of the second world war in three similar-sized cities of Germany. We compare (linked) open data sources against our gold standard data: Data sources which have been collected by local communities or historical societies (official data) dedicated to preserving the history of the city in which they operate. For our usecases in this publication, we traced those back to Wikipedia lists for which we are sure the local communities maintain them. As we will describe in more detail in Section 2.3, many of the local communities provide means of accessing Stolperstein data such as maps or apps limited to their respective area of operation. The research question proposed in our publication is to which extent (linked) open data sources can provide an equivalent fitness for use for the use case of an interested tourist/interested person compared to the data provided by the local communities. The results of this analysis can provide indications of the usefulness for a variety of possible linked data powered applications such as city tour planning apps, geofencing apps or apps advising communities about possible maintenance work (e.g. polishing) of Stolpersteine and the methodology itself is useful to validate a variety of similar location based services in terms of their fitness for use. The paper is structured as follows: In Section 2 we discuss related work concerning geospatial linked open data, geospatial data quality assessment and existing Stolperstein applications. In Section 3 we explain how we gather (linked) open data resources for our comparison case which we describe in Section 4. The results of our experiment are presented in Section 5 and discussed in Section 6 before we conclude the paper in Section 7.

1.1 Motivation

Stolpersteine have a significance as a cultural heritage monument, as a destination for tourists interested in the Jewish history and for the local Jewish community to deal with the events of the past. Every local community has an interest to raise awareness of the bibliographies of the Jewish victims so that commemoration can be shared by a variety of people. A perfect way to raise awareness of Stolpersteine in the digital age is the development of digital tour guides and apps which may also be used for city marketing. Such location based applications can only be useful if the data they provide is accurate and the geocoordinates are precise. However, not every local community has the knowledge and/or skills to develop their own apps or websites and would supposedly like a unified and simple solution for Stolperstein data integration which could be used anywhere. With the emergence of OpenStreetMap (OSM) and linked data resources this possibility has now been established on the data provider side and poses the question if this linked data collection is sufficient to match the expectations of the local community.

2. Related Work

In this section we discuss related work concerning the geospatial semantic web Section 2.1, geospatial data quality Section 2.2 and existing similar implementations of Stolperstein applications Section 2.3.

2.1 The Geospatial Semantic Web

The Semantic Web community introduced 2001 by (Berners-Lee et al., 2001) has developed a greater interest in dealing with geospatial data. Geospatial data is an essential part of many semantic databases and ontologies as illustrated by the LinkedGeoData (Auer et al., 2009), GeoNames (Wick, 2006) and more recently also Wikidata (Vrandečić and Krötzsch, 2014) ontologies. However, the means of accessing geospatial data is – even though GeoSPARQL (Battle and Kolas, 2011, Prud et al., 2006), as an official W3C standard exists – currently not implemented in every major Semantic Web database. Instead, many semantic databases implement only a subset of the GeoSPARQL standard, often

also using their own vocabulary. For example: Wikidata implements a query to find elements in a bounding box or bounding circle, but does not implement Egenhofer topological relations (Egenhofer and Franzosa, 1991) to compare relations between surface geometries. The availability of geospatial operations in the Semantic Web is despite the emergence of GeoSPARQL not sufficient e.g. for manipulations of raster and geometry manipulations. Therefore, the geospatial semantic web is despite its advantages not often considered by the geospatial community. Obvious advantages of linked data include a unified classification access to interlinked datasets with potentially more useful information for the application case and the ability to do reasoning on the given linked data set to inference data previously not present in the data set. While reasoning is an interesting application and could, to an extent, be useful in the context of this publication, we focus and investigate if the advantage of data integration holds up against a manually created data set provided by supposed experts.

2.2 Geospatial Data Quality

Geospatial Data Quality can be broken down among others to Positional Accuracy (Guptill and Morrison, 2013), Completeness (Hecht et al., 2013) and Attribute Accuracy/Completeness which are necessary to understand the contents of this publication. Positional Accuracy refers to how far a position of geometry deviates from the gold standard. Completeness defines the availability of all geometries according to a reference data set (e.g a gold standard data set). In addition attribute completeness refers to the completeness of attributes and its values in relation to the reference data set. Attribute accuracy describes the correctness of attribute values according to a gold standard to compare against. It can be measured syntactically and semantically using various distance metrics.

2.3 Existing Implementations of Apps to highlight Stolpersteine

A variety of implementations and resources to show and find Stolpersteine can be found on the web. The city of Heilbronn¹ created open data resources to share Stolperstein information as GeoJSON (Butler et al., 2016). A JavaScript Map project shows the Stolpersteine of the city of Bonn² building up on open data resources stored in the OSM project. Some projects deal with the implementation of Android apps to view Stolpersteine of selected cities which are then integrated into the apps' resources folders, e.g. Stolpersteine-Android³. Other projects include and define their own locally shared database schemas like the stolperstein-server project⁴. All of the aforementioned approaches have in common, that they are only regional solutions and not scaling up to a worldwide representation of Stolpersteine. Users has to download apps of their local city in order to get the service to find Stolpersteine nearby - a in the authors opinion unnecessary obstacle.

3. Modelling

In this section we describe the implementation of our solution following fig. 1.

¹<https://github.com/opendata-heilbronn/stolpersteine>

²<https://github.com/penpendede/stolperstein-bonn>

³<https://github.com/Stolpersteine/stolpersteine-android>

⁴<https://github.com/Stolpersteine/stolpersteine-server>

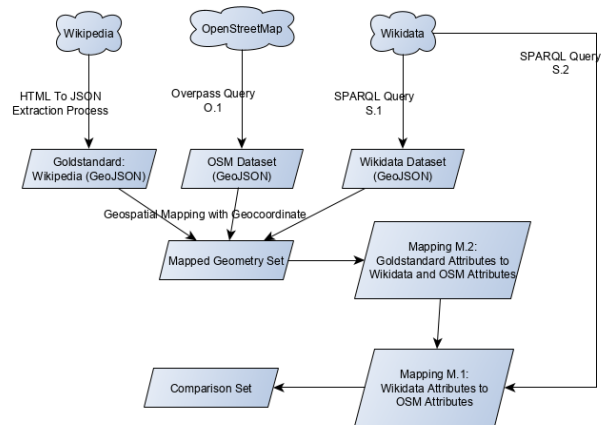


Figure 1. Dataflow of the experimental setup

3.1 Data Access

At first, we introduce the data sources we use to access Stolpersteine in a worldwide context. The first data source is OpenStreetMap (Ramm and Topf, 2010), which we access using the Overpass API (Raifer et al., 2018). A Stolperstein in OSM is modeled using the tag/value pair `memorial:type=stolperstein`, which can be used to query Stolpersteine within a defined bounding box. Usually, Stolpersteine are modelled as Point geometries (nodes in OSM), but in special occasions thresholds of Stolpersteine exist, so called "Stolperschwellen", in which case the geometry might be modeled as a way or relation in OSM. Therefore the query in Listing 1 queries all three geometry types in OSM. According to `TagInfo memorial:type=stolperstein`, OSM currently contains 20598 Stolpersteine worldwide.

Listing 1. Overpass API Query O.1: Querying Stolpersteine in a given bounding box

```
1 [out:json][timeout:15];(
2 node["memorial:type"="stolperstein"](BBOX);
3 way["memorial:type"="stolperstein"](BBOX);
4 relation["memorial:type"="stolperstein"](BBOX);
5 ;out geom;
```

The result of the Overpass API Query is a GeoJSON FeatureCollection of geometries and attributes which represents the OSM geometry. The second data source, Wikidata contains about 8000 Stolperstein representations⁵. Wikidata can be queried using a bounding circle query to match Stolpersteine in Wikidata in the same area. This query is highlighted in Listing 2

Listing 2. SPARQL Query S.1: Querying Stolpersteine (Q26703203) in a given bounding circle of 30m diameter (here with sample coordinates)

```
1 SELECT ?item ?itemLabel ?location ?rel ?relLabel
2       ?val ?valLabel ?pic ?monument
3 WHERE{ ?item wdt:P31 wd:Q26703203 .
4       ?item wdt:P18 ?pic . ?item wdt:P1684 ?monument .
5       SERVICE wikibase:around {
6         ?item wdt:P625 ?location .
7         bd:serviceParam wikibase:center "Point
8         (8.12,50)"^^geo:wktLiteral .
9         bd:serviceParam wikibase:radius "30" .}
10      ?item ?rel ?val . ?rel wikibase:directClaim ?
11      rel .
```

⁵<https://query.wikidata.org/>

```
9 SERVICE wikibase:label { bd:serviceParam
    wikibase:language "en" . }
```

The result of a Wikidata query can be given as a JSON object which we use for further processing.

3.2 Equivalent Geometry Matching

Obviously, two geometry sets have been queried, one OSM data set and one Wikidata data set. As both queries have been executed in the same area and (hopefully) depict the same objects, it is to be expected to find a lot of equivalent geometries which need to be matched in order to combine or compare their attributes. OSM geometries can be interlinked to Wikidata instances using the [OSM wikidata tag](#) and a Wikidata ID as a value. According to TagInfo⁶ this tag has not yet been added to a lot of geometries, yet the amount of geometries annotated with Wikidata IDs is increasing steadily, also, but not exclusively due to the availability of linking tools such as OSMWikidataLink⁷. As for our test areas, we found that an increasing amount of Stolpersteine depending on the city are interlinked to Wikidata - a potential result of the Wikidata Stolpersteine working group. Official data has been matched to the OSM and Wikidata representations using two matching criteria:

- The geolocation using a bounding box query of 30m radius
- The name of the person to which the Stolperstein is dedicated

The second step is especially important because Stolpersteine are often co-located in front a building, as they commonly depict a whole family which used to live at the respective place. As a result we end up with a maximum of three different representations per Stolperstein (OSM, Wikidata, official data)

3.3 Equivalent Attribute Matching

An OSM feature consists of a geometry and a list of tag/value combinations. The Wikidata community collects tag and tag/value equivalences of Wikidata classes and Wikidata instances using the Wikidata relation [wdt:P1282 \(OSM tag or key\)](#). Using the query in Listing 3, it is possible to get a list of equivalent tags and tag/value combinations which can be used to match equivalent attributes between OSM and Wikidata. This list cannot be considered as exhaustive, but is usually quite comprehensive, especially with common tags and tag/value pairs.

Listing 3. SPARQL Query S.2: Detecting Wikidata/OSM Equivalences as Mapping M.1

```
1 SELECT DISTINCT ?wdclass ?wdclassLabel ?osm
  WHERE { ?wdclass wdt:P1282 ?osm .
    SERVICE wikibase:label { bd:serviceParam
      wikibase:language "en" . }
```

For official data, the kinds of attributes are usually limited and equivalences in Wikidata and OSM can be manually mapped (cf. Listing 4).

Listing 4. Mapping M.2: Official - OSM - Wikidata

```
var equivs={
  "Adresse":{ "osm": "memorial:addr", "wd": "P969" },
  "Standort":{ "osm": "memorial:addr", "wd": "P969" },
  "Name":{ "osm": "name", "wd": "rdf:label" }}
```

After matching equivalent geometries and attributes the result can be displayed on a map⁸ as shown in Figure 2. Differences between the respective data sets can be highlighted in different colors to get an idea which equivalences and possible mismatches exists between the different data sources. The location based service developed in this re-

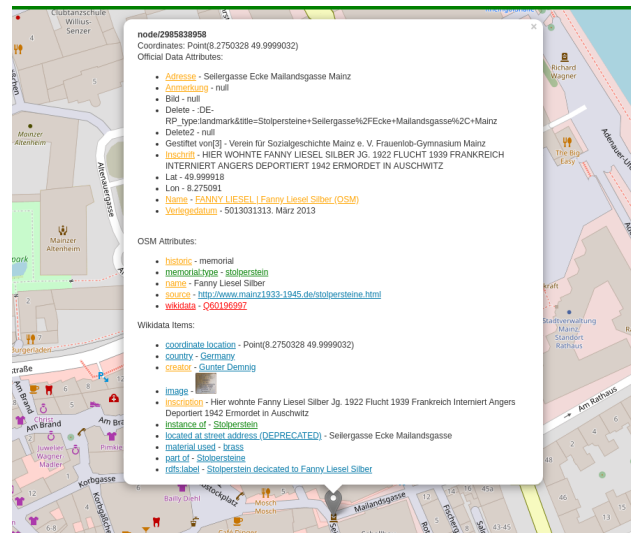


Figure 2. Comparison Screen build upon the mappings generated from Listing 3 and a custom mapping to official data shown in Listing 4: Green items represent a matching, orange item values represent a non-matching value, orange tag names represent that an equivalence is known in one of the other knowledge bases SPARQL Query to detect Wikidata/OSM Equivalences

search work - the progressive web app *StoneViewer*⁹ - is depicted in fig. 3. It allows users to navigate to Stolperstein monuments in their vicinity and to provide the information which has been provided to the app using the Wikidata and OSM Overpass API Queries. Yet our App can only be as good as the provided data feeding the application interface. The further part of this publication will therefore try to examine if our web application based on OSM and Wikidata data can achieve a comparable data quality score in comparison to the supposedly better manually created gold standard data set.

4. Experimental Setup

We investigate the quality of open data Stolperstein representations in three cities of Germany: Aschaffenburg, Mainz and Wiesbaden. The cities have been chosen because they are similar-sized, in direct proximity, state or regional capitals and therefore chances of an active mapper and Wikidata community are very high. Initial analysis yielded a considerably large presence of Wikidata objects and interlinkings in Wiesbaden and Mainz and a non-

⁶<https://taginfo.openstreetmap.org/search?q=wikidata>

⁷<https://osm.wikidata.link/>

⁸<https://i3mainz.github.io/stolperstein>

⁹<https://hsmainzgvp.pages.gitlab.rlp.net/students/stoneviewer/index.html>

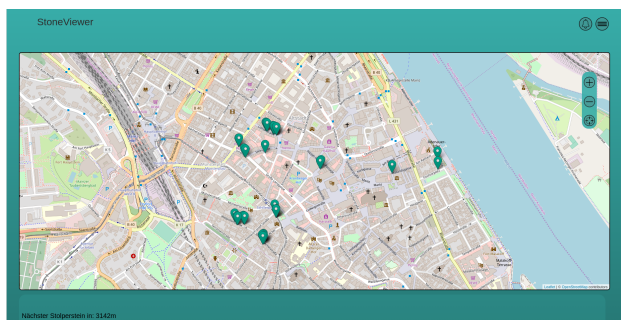


Figure 3. StoneViewer as a progressive web application to show Stolpersteine and to show information about Stolpersteine nearby

existent representation of Wikidata annotations in Aschaffenburg. In order to investigate quality of data we firstly need to define data quality criteria to investigate using data quality metrics. An obvious task to investigate is the completeness of the data set. One expectation could be that the local community would add Stolpersteine regularly, for example when they attend the deployment ceremonies of new Stolpersteine in their city. Ideally, their list should be the most complete one. The next criteria we would like to look at is the amount of facts which have been provided by the respective data sources. How many facts are provided by each data source? How many facts match the gold standard/official data? How many facts deviate and can they be categorized? How many and which facts can only be found in one of the three data sources?

While the previous questions may sound reasonable, the answers may be misleading in terms of the structure and aim of the data sets. Official data is commonly subject to only include data which is relevant to the description of the history depicted by the Stolperstein, i.e. information about people and not e.g. the material of the stone. One might argue that the material of the actual Stolperstein, the initiator of the project Gunter Demnig and the fact that it is a monument can be considered common knowledge among people familiar with the subject and add no additional value to the description of a Stolperstein. An image of the respective Stolperstein might instead equally add value to OSM, official data and Wikidata. Therefore a distinction of "valuable attributes" and common attributes will be made in the evaluation. The last criteria to investigate is the average as well as min/max distance between the given coordinates (or centroids in case of thresholds) between OSM, Wikidata and official data sets.

4.1 Requirement Profile

In the last section we have introduced the quality criteria we would like to evaluate. In this section we describe which outcomes of data quality are relevant for the previously described application case of a tourist and/or interested person in Stolpersteine, i.e. represent the fitness for use. For this purpose we follow (Homburg and Boochs, 2019) to define a requirement profile to define a good/acceptable Stolperstein representation as defined in Table 1.

In our opinion a Stolperstein description is only of use if the distance between the Stolperstein geocoordinates is sufficiently small. We assume that a deviance more than 10m from a Stolperstein geocoordinate to the gold stan-

Criteria	Threshold	Priority	Crucial?
Attribute Completeness	$\geq 80\%$	1	No
Distance from official data	$\leq 10\text{m}$	2	No
Number of mismatches	≤ 2	2	No
Missing Name	1	2	Yes
Missing Coordinate	1	2	Yes
Missing Address	1	2	No

Table 1. Requirement profile depicting the requirements of a well-represented Stolperstein set with given threshold values and priority assignments (1-4) whereas 1 is the highest priority

dard (official data) depicts a situation in which the Stolperstein will not be found by the average tourist. In addition, there are some crucial elements making a Stolperstein description useful. Firstly, the name of the victim should be present in the dataset. Without, it may prove difficult to match the data given by the app to the stone in reality, especially if many Stolpersteine exist next to each other. Secondly, a street address can be of use which in combination with a geocoordinate usually allows the unique identification of at least a group of Stolpersteine if the name is not present.

5. Results

This section presents the results of our comparison in terms of completeness, positional accuracy, valuable attributes and common conflicts.

5.1 Completeness

At first we measured the completeness of Stolpersteine in the respective areas. As can be observed in Table 2, with

City	OSM	Wikidata	Official
Aschaffenburg	93 (87%)	0 (0%)	107
Mainz	48 (21%)	21 (9%)	228
Wiesbaden	129 (20%)	22 (3%)	652

Table 2. Completeness Of Stolpersteine

the exception of Aschaffenburg, the completeness of Stolpersteine in OSM is just a fraction of the Stolperstein data gathered by the local communities displayed in official data. We theorized that in Wiesbaden and Mainz this might be due to a preference of OSM users to map features in city centers, as they provide a natural area of interest. (see Figure 4) We therefore repeated the completeness analysis for the areas of city centers in Table 3 giving a better coverage result of more than 50% for those areas.

City	OSM	Wikidata	Official
Aschaffenburg	47 (100%)	0 (0%)	47
Mainz	48 (75%)	21 (32%)	64
Wiesbaden	112 (57%)	15 (8%)	195

Table 3. City Center Completeness of Stolpersteine

5.2 Valuable Attributes

In a first step we define valuable attributes as all attributes present in the official data set. Under this assumption Table 4 shows to which extent the attributes present in official data can be substituted using OSM and Wikidata attributes. In Aschaffenburg, if a Stolperstein is annotated in OSM, 80% of the information found in the gold standard

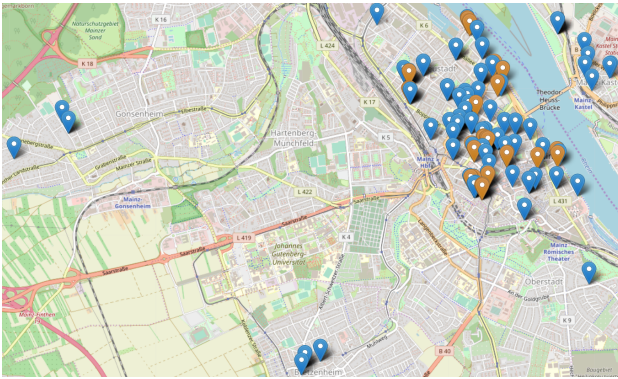


Figure 4. Coverage of Stolpersteine in Mainz: Orange icons indicate that a linking to Wikidata or OSM is present. Blue icons indicate official data. Linkings exclusively exist in the city centre.

City	OSM	Wikidata	Combined
Aschaffenburg	80%	0%	80%
Mainz	50%	70%	65%
Wiesbaden	70%	70%	80%

Table 4. Completeness of attributes compared to the gold standard. If items are annotated in Wikidata and/or OSM, this table shows the coverage percentage of valuable attributes vs. the gold standard

is represented. In fact, only one attribute is consistently missing, the time of deployment. Images of the Stolperstein are annotated 80% of the time, all other attributes are present in 100% of the cases in which an equivalent Stolperstein could be found in OSM. In Mainz, we can observe a coverage of 50% of official data attributes in OSM. The most prominent attributes in OSM are the name and the inscription of the Stolperstein. Information in Wikidata, despite annotated on far fewer Stolpersteine typically matches 70% of attributes in the gold standard, only missing the donator and the inception date. This pattern repeats in Wiesbaden, with a higher OSM completeness of street addresses.

Additional attributes in OSM and Wikidata

In addition, about 20% of OSM annotated Stolpersteine provide an attribute [wheelchair](#) indicating that the Stolperstein can be approached by people with certain disabilities. While this attribute is not present in the official data it certainly provides an extra value to the end user. In addition, for some Stolpersteine the date of birth and date of death of the commemorated person was added as an attribute. In Wikidata we found a lot of attributes we would not consider as valuable for our assessment. [material used](#) (P186) describes the material used to build the Stolperstein which is always brass. In addition we find information about the creator [Gunter Demnig](#) (Q195796), which also is the same for every Stolperstein. Sometimes the [country](#) (P17) is annotated, which might be considered valuable, but could also be inferred from the GPS coordinates.

5.3 Positional Accuracy

Figure 5 shows the distances between OSM coordinates and official data coordinates for Stolpersteine with equivalences in OSM/Wikidata. In Aschaffenburg, we can observe a high correlation of the coordinates which leads us

to believe that either the local community also contributes to OSM or that OSM contributors chose data of the local community to enrich the map. Mainz and Wiesbaden ex-

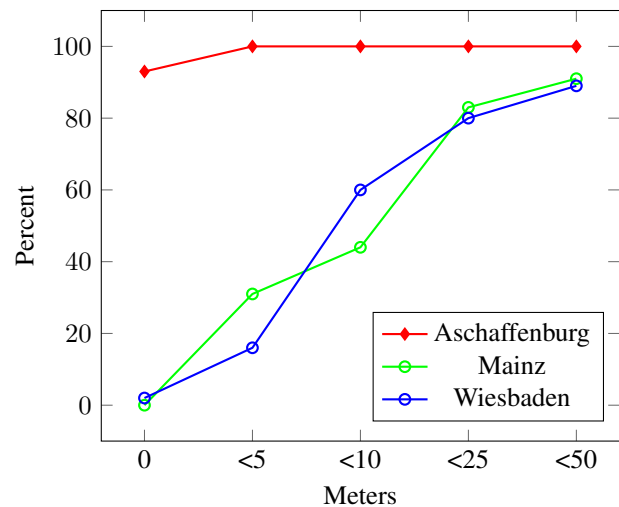


Figure 5. Stolperstein distances in Aschaffenburg, Mainz and Wiesbaden. In Aschaffenburg, distances to official data are negligible possibly due to the adoption of official coordinates in OSM. In Wiesbaden and Mainz more than 80% of all recorded Stolpersteine can be found in the vicinity of 25m

hibit a very similar behavior concerning the positional accuracy. Between 40% and 50% of Stolpersteine have been mapped within a distance of 10m, but more than 80% have been mapped within a distance of 25m to the official data set. Concerning why geocoordinates deviate we found a variety of possible reasons.

No precise GPS positional knowledge

Figure 6 shows that a mapper might not necessarily know the exact position of the Stolperstein but annotates according to the street address. Figure 6 shows, that a Stolperstein may be located at the same street address but the distance to the official data coordinate could exceed 30m in certain situations.

Semantic mistakes

A few mistakes yielding distances of greater than 1km, originated in the fact that coordinates had been copied from another Stolperstein in the vicinity, therefore again pointing to annotation mistakes in the (linked) open data set. Also, the name of the person which is commemorated might have been annotated wrongly, yielding distances of greater than 1km using our data mapping algorithm.

Mapping of groups

We found that in official data, groups of Stolpersteine have occasionally been assigned the same geocoordinates per Stolperstein. Such groups occur fairly regularly (Figure 7) when a whole family has been deported at a specific address, yielding a Stolperstein for each of the family member. Usually, the deviations of Stolperstein coordinates are within a range of 10m, so that we do not consider such deviations to be disturbing to the average tourist.

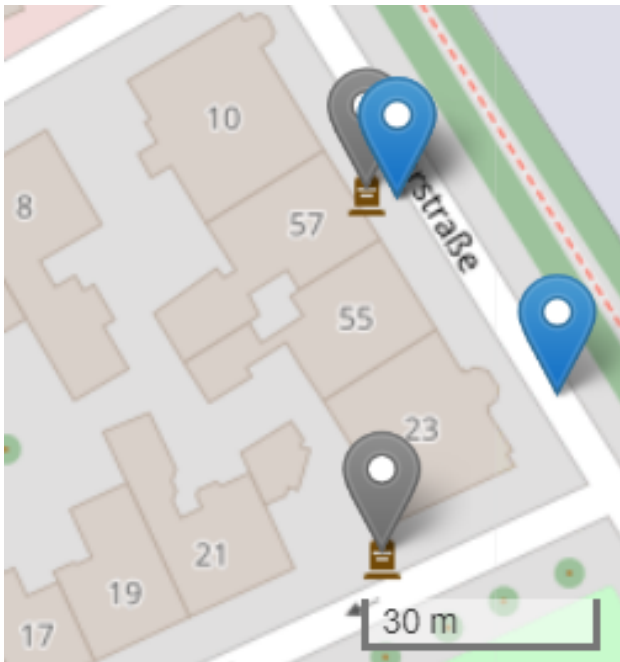


Figure 6. Distance of official geocoordinate and OSM geocoordinate in front of the same house. Here, the exact Stolperstein position might not be known to the OSM annotator and a guess to where the Stolperstein is located was done based on the street address yielding in a distance of 30m between the coordinates.

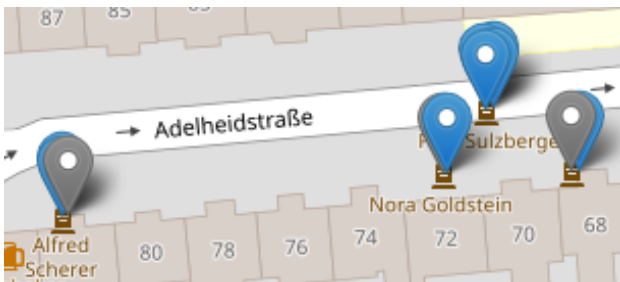


Figure 7. Groups of Stolpersteine in Wiesbaden

Corner cases

Finally, in all three cities we found about 3% of Stolpersteine in the official dataset not depicting any coordinate or missing a name and one to three Stolpersteine being present in OSM but not in the official dataset. This observation shows minor inconsistencies in both the official data set and possibly in OSM.

5.4 Attribute matches

We define an attribute match in two different ways. The first way is a syntactic match which occurs when an equivalent tag and the same value is found (denoted in green in Figure 9). The second attribute match is a semantic attribute match for which an equivalent tag and an equivalent value is found. If the equivalent value conveys the same semantic (or an extended) meaning compared to the official data we count it as a match. The semantic match was performed by a human annotator. Table 5 shows the differences between the matching types. Using semantic matching, additional information does still yield a match if non-additional information matches. This fact is impor-

Type	Value1	Value2	Match
Syntactic	Neugasse 8	Neugasse 10	False
Syntactic	Neugasse 8	Neugasse 8, Mainz	False
Syntactic	Neugasse 8	Neugasse 8	True
Semantic	Neugasse 8	Neugasse 10	False
Semantic	Neugasse 8	Neugasse 8, Mainz	True
Semantic	Neugasse 8	Neugasse 8	True

Table 5. Differences between semantic and syntactic matchings as applied in our evaluation

tant, as the schemas of three data sources OSM, Wikidata and Official data are not aligned.

5.5 Common conflicts

Figures 8 and 9 show typical conflicts which have been observed in the testing areas. Semantic conflicts occur when a misinterpretation of the human annotator happens. Syntactic conflicts either occur because of typing mistakes or because of a lack of convention on how to fill the respective attribute value. A distinction of how mistakes are treated in the evaluation is shown in Table 5. In Figure 8 it is un-

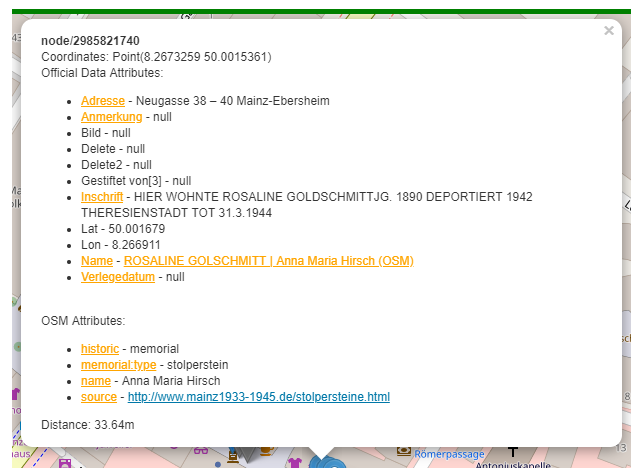


Figure 8. Semantic conflict: Despite a fitting address, the name of the person depicted on the Stolperstein does not match with the official dataset

clear if the name of the person depicted in the Stolperstein should be annotated with or without the maiden name of the person. In Wikidata for example, maiden names are classified separately (Q1376230), whereas OSM and the official data sources we investigated do not make such distinctions.

Frequency Analysis

In order to investigate how often the aforementioned conflicts occur, we conducted an edit distance analysis using the Levenshtein distance (Yujian and Bo, 2007) metric between official data, OSM and Wikidata respectively. In addition we checked if a given String is the substring of its comparison counterpart. Figure 10 shows examples of the calculated edit distances using the essential name attribute. We found that high edit distances without a substring correlation were pointing to inconsistencies supposedly in the linked open data data set which would need to be corrected (semantic mistakes). Non-substring correlated results with a low Levenshtein distance usually turned out as typing mistakes or syntactic name inconsistencies. Table 6 shows the frequency of these mistakes.



Figure 9. Writing convention: The name of the person depicted by the Stolperstein may be written using the name including the maiden name or without it. Both representations are correct and a matter of convention.

- **Name - HUGO LÖWENSBERG | Hugo Löwensburg [1] (OSM)**
- **Name - ALBERT LEOPOLD | Albert Leopold Vogel [6] (OSM)**
- **Name - ELLA RAPHAELSON | Berta Erlanger [15] (OSM)**

Figure 10. Edit distance examples: Typing mistake (low edit distance, no substring), incomplete name (substring and low edit distance), different name (no substring and high edit distance)

5.6 Overall quality of LOD Stolperstein data

Revisiting the requirements set in Section 4.1 defining well-represented Stolpersteine for the case of an interested tourist, we answer the question for which of the examined areas there is a sufficient amount of such Stolpersteine are representative. Table 7 shows the results of this analysis. In column 1 and 2 we can see that only a fraction of the Stolpersteine which have been annotated in linked open data have crucial flaws like a missing name, address or coordinate. The percentage of well-represented Stolpersteine in general has been between 72% and 95% proving that if annotations were made, they usually have sufficient quality. Column 3 shows that except for the city of Aschaffenburg, the coverage of well-annotated Stolpersteine including the ones with no equivalence is still very low.

6. Interpretation

Our experiment showed that the coverage of Stolpersteine is quite heterogeneous throughout the cities we have investigated. In Mainz and Wiesbaden, a significant amount of Stolpersteine is present in the city center, many of which are interlinked in OSM and Wikidata. In Aschaffenburg, almost no Wikidata elements are present and interlinked but OSM shows in terms of completeness a good coverage of

Area	Syntactic	Semantic	Sem. Correct
Aschaffenburg	100%	35%	65%
Mainz	31%	44%	66%
Wiesbaden	97%	55%	45%

Table 6. Distribution of syntactic and semantic mistakes in the testing areas with the amount of semantically correct attributes in the third column

Area	Well-represented	Crucial	Total
Aschaffenburg	87 (95%)	5 (5%)	81%
Mainz	36 (72%)	4 (8%)	21%
Wiesbaden	109 (84%)	3 (2%)	16%

Table 7. Amount of well-represented Stolpersteine for the use case of a tourist in relation to the total amount of Stolpersteine present in the linked open data resources (Wikidata and OSM attributes combined)

Stolperstein elements in the city. The positional accuracy analysis showed that with the exception of Aschaffenburg where OSM seems to be managed by the local community as well, most Stolpersteine can be found within a radius of 25m around the official point. About half of them can be found within a distance of 10m. Here, we must note that a variety of factors can influence the accuracy of geospatial measurements¹⁰. We assume that not all measurements have been taken using professional equipment and can therefore suggest that distances smaller than 10m could be due to non-precise equipment being used. (Zandbergen and Barbeau, 2011) In Wiesbaden and Mainz this could explain the discrepancy of about half of the Stolpersteine investigated. Another 10% of Stolpersteine are erroneously annotated or captured in Wiesbaden and Mainz. The measurements in between points roots in our opinion from missing information about the Stolperstein's precise location, a possibly imprecise mapping approach using satellite images or a combination thereof. Attributes which we considered as valuable have been matched in the respective areas. We found that in Aschaffenburg, the coverage of these valuable attributes is 80% mainly only omitting the deployment date. In Mainz and Wiesbaden a slightly lower completeness rate of the gold standard attributes has been observed. For the use case of a tourist visiting a city for the interest of Stolpersteine, several factors point to linked open data already being useful. The completeness in the city centers is sufficient, between 70% and 80% of all gold standard attributes are maintained and the positional accuracy of most Stolpersteine were correct.

7. Conclusions

In this publication we have investigated if a mobile application which builds up on LOD - in our case a PWA app, can deliver an equivalent quality to currently available locally build applications showcasing Stolperstein stones in a certain area. We showed that it is feasible to build an app powered by linked data to view semantically-enriched Stolperstein information anywhere in the world. In the three areas of Germany, the coverage of Stolpersteine in the respective city centers would be sufficient to guide a tourist to many of the Stolpersteine in the immediate vicinity and to get to know the most basic information about the name of the victim and the inscription. We have also shown that the attributes which are covered in open data, especially when combining the two linked open data sources Wikidata and OSM are to a great extent semantically correct with respect to the gold standard. However, the linked data coverage of Stolpersteine is currently far from complete in this knowledge domain in any of the three cities we have investigated. This may change with initiatives like the

¹⁰https://wiki.openstreetmap.org/wiki/Accuracy_of_GPS_data

Stolpersteine working group in Wikidata and would provide the opportunity to repeat this study in a few years to investigate its progress. In conclusion, this study suggests, that linked data applications for Stolpersteine seem to serve the needs of a casually interested tourist especially in the city centers we have investigated, but currently fails to provide a basis for more detailed applications with a focus for the local communities. In comparison to locally created apps, users can for the time being still expect a more detailed description and completeness of information using a locally prepared app version. However, for the purposes of planning a city tour or to gain rudimentary information about Stolpersteine, linked data currently can provide a sufficient quality. We see much potential in applying our approach to further point of interest data in OpenStreetMap or Wikidata. Examples could be historical monuments like castles or UNESCO world cultural heritage sites to evaluate data quality.

7.1 Future Work

Possible future work might be to investigate more cities within Germany and/or in Europe in order to determine the frequency of the mistakes we found in this publication in other areas. In addition, it could be of interest to conduct a sample of test measurements using professional GPS equipment in order to verify the geocoordinates of both the official data set and the (linked) open data set. The current analysis is based on the assumption that the geocoordinates of the official dataset by the local communities is superior. However, this assumption may be invalid, as we cannot be sure that the local communities use professional equipment for measuring Stolperstein coordinates, nor that they possess the necessary knowledge to capture them correctly. Another interesting area of research could be if Stolperstein information could be completed by using reasoning approaches, generating more information about the Stolperstein data sets through inferencing. Finally, the results of the string similarity analysis can be used to create a tool which not only points out inconsistencies in Stolperstein map data using the given data sets, but giving means to a local community to correct the mistakes found.

References

- Auer, S., Lehmann, J. and Hellmann, S., 2009. Linkedgeo-data: Adding a spatial dimension to the web of data. In: *International Semantic Web Conference*, Springer, pp. 731–746.
- Battle, R. and Kolas, D., 2011. Geosparql: enabling a geospatial semantic web. *Semantic Web Journal*.
- Berners-Lee, T., Hendler, J., Lassila, O. et al., 2001. The semantic web. *Scientific american* 284, pp. 28–37.
- Butler, H., Daly, M., Doyle, A., Gillies, S., Hagen, S. and Schaub, T., 2016. The geojson format.
- Egenhofer, M. J. and Franzosa, R. D., 1991. Point-set topological spatial relations. *International Journal of Geographical Information System* 5(2), pp. 161–174.
- Franke, D. and Demnig, G., 2009. *Stolperstein*.
- Guptill, S. C. and Morrison, J. L., 2013. *Elements of spatial data quality*. Elsevier.
- Hecht, R., Kunze, C. and Hahmann, S., 2013. Measuring completeness of building footprints in openstreetmap over space and time. *ISPRS International Journal of Geo-Information* 2(4), pp. 1066–1091.
- Homburg, T. and Boochs, F., 2019. Situation-dependent data quality analysis for geospatial data using semantic technologies. In: W. Abramowicz and A. Paschke (eds), *Business Information Systems Workshops*, Springer International Publishing, Cham, Switzerland, pp. 566–578.
- Pôças, I., Gonçalves, J., Marcos, B., Alonso, J., Castro, P. and Honrado, J. P., 2014. Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring. *International Journal of Geographical Information Science*.
- Prud, E., Seaborne, A. et al., 2006. Sparql query language for rdf.
- Raifer, M. et al., 2018. Overpass turbo. *Overpass API*.
- Ramm, F. and Topf, J., 2010. *OpenStreetMap: Die freie Weltkarte nutzen und mitgestalten*. Lehmanns.
- Vrandečić, D. and Krötzsch, M., 2014. Wikidata: a free collaborative knowledge base.
- Wick, M., 2006. *GeoNames*. GeoNames.
- Yujian, L. and Bo, L., 2007. A normalized levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence* 29(6), pp. 1091–1095.
- Zandbergen, P. A. and Barbeau, S. J., 2011. Positional accuracy of assisted gps data from high-sensitivity gps-enabled mobile phones. *The Journal of Navigation* 64(3), pp. 381–399.