Open Accessibility Data Interlinking

Chaohai Ding, Mike Wald, and Gary Wills

Web and Internet Science Group, ECS, University of Southampton, Southampton, UK
{cd8e10,mw,gbw}@ecs.soton.ac.uk

Abstract. This paper presents the research of using Linked Open Data to enhance accessibility data for accessible travelling. Open accessibility data is the data related to the accessibility issues associated with geographical data, which could benefit people with disabilities and their special needs. With the aim of addressing the gap between users' special needs and data, this paper presents the results of a survey of open accessibility data retrieved from four different sources in the UK. An ontology based data integration approach is proposed to interlink these datasets together to generate a linked open accessibility repository, which also links to other resources on the Linked Data Cloud. As a result, this research would not only enrich the open accessibility data, but also contribute to a novel framework to address accessibility information barriers by establishing a linked data repository for publishing, linking and consuming the open accessibility data.

Keywords: Linked Data, Open Accessibility Data, Information Retrieval, Data Interlinking.

1 Introduction

Accessible tourism is listed as a significant research topic proposed by the European Commission because of the difficulties faced by people with disabilities [6]. Due to the complex conditions of travelling, to address the discordance between the expectation of people with disabilities and transport patterns would be a significant challenge [4]. There have been several research projects proposed for improving accessibility for people with special needs or disabilities [1], [5], [7]. According to the literature review [3], there have been some problems identified in these projects, such as accesibility data isolated from different systems, missing accessibility related information or difficulties for fetching useful content. Another issue is that most current standards for accessibility metadata research are focussed on the user interface or Web content rather than the guidelines for publishing the accessibility metadata to describe facilities in the real world, such as buildings, train stations or restaurants. The motivation of this research is to address the information barriers between users' needs and accessible facilities or places by integrating heterogeneous accessibility related resources and then applying the Linked Data principles to establish a linked open accessibility data repository. This paper firstly presents the survey of four different accessibility related datasets and then demonstrates the approaches and challenges for

K. Miesenberger et al. (Eds.): ICCHP 2014, Part II, LNCS 8548, pp. 73–80, 2014.

mapping and interlinking these datasets. Finally, this paper indicates the opportunities and potential benefits of publishing and consuming open accessibility data.

2 Open Accessibility Data

Open accessibility data is the data related to the accessibility issues and associated with geographic data, such as step-free access, accessible entrances, accessible toilets, accessible parking, large print or systems to aid hearing. Accessibility data also refers to the data that benefits people with special needs, such as baby change facilities, staff help, carrying large luggage or travelling with a baby pushchair. This section presents a brief summary of open accessibility data within the UK based on the datasets retrieved from four different sources, namely Wheelmap[1], Factual[2], Step-free Access Guide Feed (London Tube)[3], and National Rail[4].

Wheelmap is a crowdsourcing-based online map service to provide information about wheelchair accessible places around world. It is based on Open Street Map (OSM)[5] project and the nodes are the most primitive geographic entities in OSM using WGS84 reference system to represent geographic information. Based on the dataset extracted from Wheelmap on 29/11/2013, there are 421666 nodes within UK and 1.11% places are annotated as wheelchair accessible, while 0.24% nodes are not wheelchair accessible. 0.21% places have limited wheelchair accessibility and there are 98.44% entities without the wheelchair accessible information.

Factual is a location platform that provides over 65 million local businesses and points of interest in 50 countries. The only UK based accessibility related dataset is the Restaurants-UK dataset. There are 210613 restaurant entities in the dataset provided by the Factual Team on 10/01/2014, 8904 restaurants are annotated as wheelchair accessible while 1786 entities are indicated as not accessible. However, there are 94.92% restaurants that are unknown for this attribute.

Transport for London (TfL) provides a series of Open Data for developers, which includes the step free tube guide data, station facilities and some real time data. The step-free tube guide data contains the step-free access information of London tube, DLR and Overground stations. According to the data we downloaded on 10/01/2014, there are 362 tube station entities and approximately 50% stations are annotated as either wheelchair accessible or not, while the rest of the entities are unknown. There is an accessible interchanges attribute, which presents the accessible metadata for interchanging to other public patterns.

UK national railway stations accessibility data was crawled from the national rail website on 14/01/2014 and we converted the extracted data into the JSON format. There are 2601 railway stations data associated with accessibility information, such as

---

[1] `http://wheelmap.org/en/`
[2] `http://www.factual.com/`
[3] `http://www.tfl.gov.uk/syndication/feeds/step-free-tube-guide.xml`
[4] `http://www.nationalrail.co.uk/`
[5] `http://www.openstreetmap.org`

ramps for train access, step free access, accessible toilets, accessible ticket machines and accessible car parking. There are also some attributes for people with special needs, such as toilets, car park and baby changing facilities. Most attributes are annotated with metadata and the percentage of entities without annotation is just 1%. But for some attributes, such as baby changing, accessible toilets, accessible payphones and accessible ticket machines, there are more than 50% entities annotated as unknown.

According to the brief survey of open accessibility data presented in the previous section, there are some accessibility related datasets available online, which refer to the real world locations. These accessibility datasets are from multiple heterogeneous sources, and currently there is no standard guideline for specifying the attributes to annotate the accessibility metadata for public places or facilities. Crowdsourcing is not enough for obtaining open accessibility data due to the different data schemas or data quality cross the different systems or applications. Therefore, in order to fulfil the missing information and improve the quality for accessibility related data, we applied the Semantic Web technologies and Linked Data principles to map and interlink these datasets as a public Linked Data repository for accessibility metadata research, such as the research of accessibility metadata schema for publishing accessible place information or the research of guidelines to add accessibility metadata for existing buildings or facilities resources on the Web.

3 Accessibility Data Interlinking

As a core data layer of the Semantic Web, Linked Data exposes the advanced characteristics for knowledge sharing, such as human and machine readable data, well-structured data, standard data format, ontology-based reasoning and domain specified [2]. Ontology is the formal, explicit specification of a shared conceptualization. There are three different ways to integrate data based on ontology, namely the single ontology approach, multiple ontologies approach and hybrid approach. However, a single ontology approach would face the problems of the mappings between low dimensional entities (only one wheelchair accessible attribute) and high dimensional entities (multiple attributes stands for wheelchair accessible). The mappings between the multiple ontologies are still a challenging issue, which is a research area involved a large number of research efforts. According to the fact of multiple heterogeneous schemas among the datasets, we proposed two different ontology driven approaches to integrate the datasets.

The first approach is a hybrid approach, which is to design an individual ontology for each dataset to represent the entities based on its own schema and some existing ontologies, namely the accessible tube stations ontology, the wheelmap ontology, accessible restaurants ontology and accessible train stations ontology. Then we design a top-level ontology to define the basic accessibility terms in the accessible buildings domain, which could reduce the difficulties for multiple ontologies mapping. However, building specific ontologies for each dataset is difficult and not scalable for future integration. For example, Figure 1 presents the ontology graph for the data schema extracted from Wheelmap. If the data provider changes the data schema into another different structure, then the ontology needs to be redesigned. There is also another issue due to the fact that the ontology should be the formal, explicit specification of a shared conceptualization rather than the specific data schema in different applications.

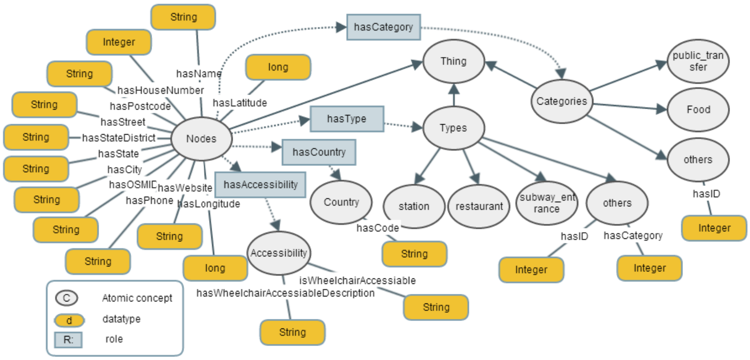

Fig. 1. Ontology Graph in Wheelmap

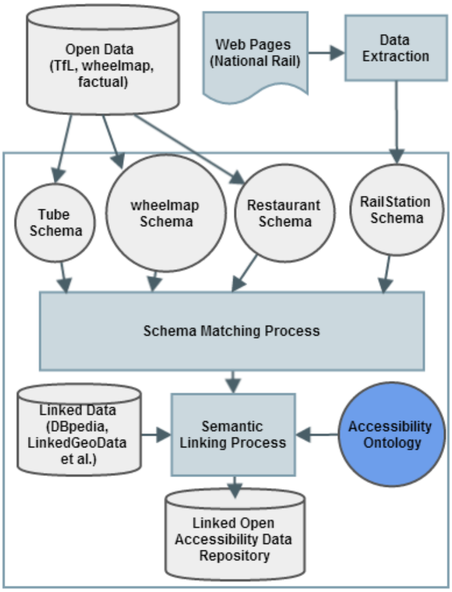

Fig. 2. Architecture of Mapping and Interlinking Open Accessibility Data

The second approach is the single ontology approach, which is the approach applied in this paper. Figure 2 presents the architecture of mapping and interlinking open accessibility data. Due to the problem of mappings between low dimensional properties and high dimensional properties in ontology reasoning, we firstly developed a basic schema of matching rules to map the same entities in different datasets by observing the data schemas in different datasets. Figure 3 demonstrates the example of two equivalent entities in both wheelmap and National Rail datasets. There are some basic properties to describe the place information, such as name, geographic data, phone, address. For accessibility related properties, there is only one property called Accessibility with wheelchair and the description for wheelchair accessible. By contrast, National Rail provides more than 50 properties to describe the accessibility issues. Another problem is the rules for entity mapping. As indicated in Figure 3, the name of train station in National Rail is London Euston Rail Station compared with Euston in wheel map. The latitude and longitude of the entity in two datasets are also slight different.

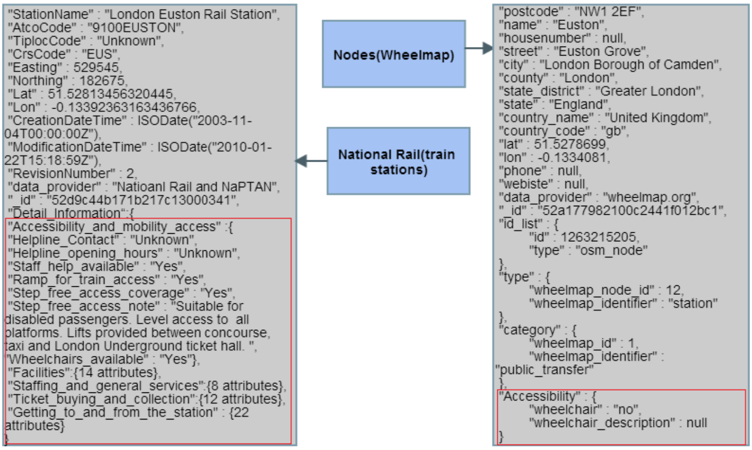

Fig. 3. Same Entities in Wheelmap and National Rail Datasets

Due to the fact that there is geographical data available for each entity in these four datasets, therefore, we develop a simple mapping rule including geographic distance matching, entity name matching and other information matching, such as address, postcode, phone number and other specified identifiers. The Mapping Rules for each entity is described as follows:

- **Step one**, the Entity Type Matching: there are 12 primary categories and 130 different types of places in wheelmap dataset, such as train station, tube station, platform and restaurants. Therefore, it is easier to map other datasets to wheelmap datasets based on the entity categories and types.
- **Step two**, the Name Similarity Matching: this step is mainly for both entities with name. If not, then go to step three. In this step, it applied the Regular Expression Matching and Jaro-Winkler Distance algorithm to get the name similarity score between two entities.
- **Step three**, Geographic Distance: calculating the geographic distance based on latitude and longitude for each entity.
- **Step four**, Additional Information Matching: matching any additional information available for both entities, such as telephone, postcode, website or identifiers.

- **Final Step**: based on the scores calculated in previous matching steps, we defined the candidate matched entities should be satisfied with following criteria: Geo Distance<=200 and Name Similarity Score>=0.85 (if there is name similarity score). If there is more than one pair of entities in the candidate list, the entities with smallest geo distance would be matched.

Table 1. Entities Mapping between Wheelmap and Other Datasets

	Train Wheelmap	Train Stations	Tube Wheelmap	Tube Stations	Food Wheelmap	Factual restaurant
Total	3384	2601	222	362	56970	10629
With Name	3323	2601	119	362	52901	10629
No Name	61	0	103	0	4071	0
Mapped	2438		90		1907	

Table 1 indicates the data mapping results based on this mapping rule. However, there is some mismatching of entities or redundancies, which would be evaluated in the following work. According to the schema observing in Figure 3, there is a data conflict to present the same entity in different schemas. The properties to describe wheelchair accessibility in the National Rail dataset indicate that the station London Euston Rail Station is wheelchair accessible, while the same entity in wheelmap displays it as not wheelchair accessible. There are also other data conflict issues for other properties, which are not related to accessibility, such as postcode, telephone or even address. Therefore, we examined the data conflict issues in our mapped entities. Table 2 compared the wheelchair accessibility properties in mapped entities. For train station datasets mapping, 50 of 2436 entities are conflicted. 4 of 90 entities are conflicted in tube station datasets mapping and 52 of 1907 entities are conflicted in restaurant datasets mapping. However, there is large amount of accessibility data enrichment archived during this datasets mapping. For instance, there are 2225 train stations, 26 tube stations and 1794 restaurants with wheelchair accessibility data mapped to the same entities in wheelmap dataset.

Table 2. Open Accessibility Data Mapping

ItemA	ItemB	No.	ItemA	ItemB	No.	ItemA	ItemB	No.
Train wheelmap	Train Station	2436	Tube wheelmap	Tube Station	90	Food Wheelmap	Factual restaurant	1907
Null	No	1068	Null	No	12	Null	Yes	1794
Null	Yes	1157	Null	Yes	14	Yes	Yes	61
Null	Null	27	Null	Null	39	No	Yes	23
Yes	No	25	Yes	No	1	Limited	Yes	29
Yes	Yes	85	Yes	Yes	3			
Yes	Null	3	Yes	Null	0			
No	No	46	No	No	3			
No	Yes	4	No	Yes	1			
No	Null	0	No	Null	14			
Limited	No	10	Limited	No	1			
Limited	Yes	11	Limited	Yes	1			
Limited	Null	0	Limited	Null	0			

Then we designed the ontology related to the place accessibility called Place Access Ontology[6] to describe the basic concepts and semantic relationships for different places. The place class in Place Access Ontology mainly refers to the place class in Schema Ontology[7]. There are 50 different categories and 120 types in schema ontology and there are some existing properties to describe the place class, such as name, geo (latitude and longitude), address, telephone or the url to present the place on the Web. There is no existing ontology to describe the accessibility facilities for the buildings or places. Therefore, we need to construct the accessibility related classes for facilities, such as Braille, LargePrint, SignLanguage, AutomaticDoor, HearingSystem, HelpPoints, Lift, AccessibleParking, AccessibleToliets, StepFree and AccessibleChangingRoom as well as some object properties and data properties to define the relationships, constraints and reasoning rules. As described in Figure 4, we stored all extracted data in MongoDB, a NoSQL document-based database and dumped the database into JSON-LD[8], a W3C standard format for Linked Data by annotating the semantic metadata to each entity in the database with Place Access Ontology. Based on the entities mapped in previous sections, we used equivalent assertions (owl:sameAs) to interlink the same entity from different data sources. By applying named entities recognition, the entities in linked open accessibility repository could also be linked to other resources in the Linked Data Cloud, such as DBpedia, Freebase and LinkedGeoData.

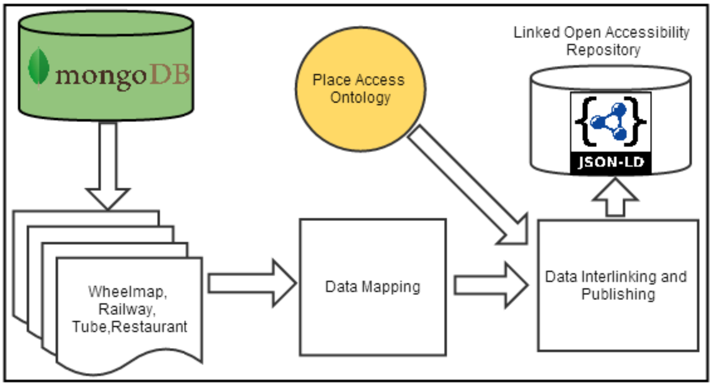

Fig. 4. Mapping and Interlinking Process

However, there are still several problems and challenges exposed in this experiment. The first problem is to improve the mapping rules for the same entity from different datasets. The second problem is to address the data conflict for the accessibility properties. Another problem is mappings between low dimensional properties and high dimensional properties based on ontology reasoning and inference rules. The

[6] Available on: `http://waisvm-cd8e10.ecs.soton.ac.uk/2014/2/ontology/placeaccess.owl`

[7] `http://schema.org/docs/schemaorg.owl`

[8] `http://www.w3.org/TR/json-ld/`

question is how to determine the place object is accessible based on its accessibility level of several facilities available on the Web. All these problems will be our future research work.

4 Conclusion

Although crowdsourcing is a powerful approach to fetch millions of data entities, it is still difficult to obtain the accessibility data with good quality, compared with the accessibility data published by government. Therefore, we proposed the approach to map and interlink different accessibility related datasets into a public Linked Open Data repository. According to the demonstration of the mapping and linking results in previous sections, this approach could enrich the metadata and improve the data quality related to accessibility issues. There are also some problems and challenges proposed for this linking approach. The next step of this research is the evaluation of mapping methods and results as well as the test based on the linked open accessibility data repository. Based on this research, researchers could evaluate the accessibility level of different areas while developers could use the service to produce more accessible applications or contribute to the repository. Users could get more useful accessibility related information based on their special needs. As a consequence, this research would not only propose a method to improve and enrich the open accessibility data, but also contributes a novel way to research the addressing of the accessibility issues with the Linked Data principles.

References

1. Bekiaris, E., Panou, M., Mousadakou, A.: Elderly and disabled travelers needs in infomobility services. In: Stephanidis, C. (ed.) UAHCI, Part I, HCII 2007. LNCS, vol. 4554, pp. 853–860. Springer, Heidelberg (2007)
2. Bizer, C., Heath, T., Berners-Lee, T.: Linked data-the story so far. International Journal on Semantic Web and Information Systems (IJSWIS) 5(3), 1–22 (2009)
3. Ding, C., Wald, M., Wills, G.: Travelling for all. In: Proceedings of the International Conference e-Society 2013, IADIS, pp. 519–522 (2013), http://eprints.soton.ac.uk/350984/
4. Packer, T.L., McKercher, B., Yau, M.K.: Understanding the complex interplay between tourism, disability and environmental contexts. Disability and Rehabilitation 29(4), 281–292 (2007), doi:10.1080/09638280600756331
5. Vanderheiden, G.C.: Using Distributed Processing to Create More Powerful, Flexible and User Matched Accessibility Services. In: Stephanidis, C. (ed.) UAHCI 2009, Part II. LNCS, vol. 5615, pp. 438–444. Springer, Heidelberg (2009)
6. Westcott, J.: Improving information on accessible tourism for disabled people. Office for Official Publications of the European Communities (2004)
7. Wiethoff, M., Sommer, S.M., Valjakka, S., Van Isacker, K., Kehagias, D., Bekiaris, E.: Specification of information needs for the development of a mobile communication platform to support mobility of people with functional limitations. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 595–604. Springer, Heidelberg (2007)