A Survey of Open Accessibility Data

Chaohai Ding
Web and Internet Science Group
ECS, University of Southampton
Southampton, UK
cd8e10@ecs.soton.ac.uk

Mike Wald
Web and Internet Science Group
ECS, University of Southampton
Southampton, UK
mw@ecs.soton.ac.uk

Gary Wills
Web and Internet Science Group
ECS, University of Southampton
Southampton, UK
gbw@ecs.soton.ac.uk

ABSTRACT

This paper presents the research of using Linked Data for enhancing accessibility data, especially for accessible travelling. With the aim of addressing the gap between users' special needs and accessibility data, this research initially explores the current situation of open accessibility data. Open accessibility data is the data related to the accessibility issues and associated with geographical data, which could benefit people with disabilities or special needs. This paper proposed a survey of open accessibility data in UK based on the datasets retrieved from five different resources. After examining the features of each dataset, a mapping approach using Semantic Web technologies is proposed to interlink these datasets together to generate a linked open accessibility repository and link this repository to other resources on the Linked Open Data Cloud (LODC). As a result, this research would not only benefit people with disabilities, but also contribute to a novel method to address accessibility information barriers by establishing a linked open accessibility data repository for publishing, integrating and consuming the accessibility data.

Categories and Subject Descriptors

K.4.2 [**Computer and Society**]: Social Issues ─ *Assistive technologies for persons with disabilities*; H.3.5 [**On-line Information Services**]: Data sharing and Web-based services

General Terms

Design, Human Factors

Keywords

Linked data, open accessibility data, information retrieval, data interlinking

1. INTRODUCTION

There are approximately 11.5 million people with at least one type of impairment or disability in the UK (Family Resource Survey of 2010/11). The impairment types of mobility, lifting and carrying, and manual dexterity are listed as the top three prevalent disabilities. Most of these people are facing barriers and problems in their daily activities. Accessible tourism is listed as a significant research topic, which is also proposed by the European Commission because of the difficulties faced by people with disabilities [7]. 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

[5]. There have been several projects proposed for improving accessibility for people with special needs or disabilities [1], [6], [8]. According to the literature review [3], there have been some problems identified in these projects, such as accessibility information isolated in different systems, information overloading and difficulty for fetching useful content. Therefore, the motivation of this research is to address the information barriers between users' need and accessibility data by integrating heterogeneous accessibility related resources and establishing a public linked open accessibility data repository for future research. The second part of this paper mainly presents a survey of open accessibility data, while the third section proposes the potential challenges and solutions for ontology driven accessibility data integration. The last section is the conclusion and the future work.

2. OPEN ACCESSIBILITY DATA SURVEY

Open accessibility data is the data related to the accessibility issues associated with geographical data for people with disabilities, such as step-free access for public transport stations, accessible entrance, accessible toilets, accessible parking or hearing loops. Accessibility data also refers to the data that brings benefit for people with special needs, such as baby change facilities, staff help available points, customer help points, travelling with a baby pushchair or carrying large luggage. This section demonstrates a survey of open accessibility data in the UK based on the datasets retrieved from five different resources, namely Wheelmap, Factual, Step-free Access Guide Feed (London), National Rail and AccessTogether.

2.1 Wheelmap

Wheelmap[1] is an online map service for searching, finding and marking wheelchair accessible places by crowdsourcing. It is mainly based on Open Street Map (OSM)[2] project, an open source, collaborative, free and editable online map service. There are three fundamental geographic entities in OSM, namely Nodes, Ways and Relations. Nodes are the most primitive entities using WGS84 reference system to represent geographic information with latitude and longitude. Wheelmap provides the RESTful API for data querying. Table 1 indicates the UK based accessibility data, which was extracted from this API on 29/11/2013. There are 421666 nodes within the UK with their own unique and valid OSM Node ID in Wheelmap. 4687 of these nodes are annotated with wheelchair accessible, while 1001 nodes are not wheelchair accessible. 898 nodes are limited wheelchair accessible. 98.44% of these nodes are annotated as unknown for the wheelchair accessible information. All these nodes are classified into 12 categories and 130 types. Each category is associated with several types. There are 12 core attributes for each entity including OSM ID, Name, House Number, Street, City, County, Country, Postcode, Phone, Website and Wheelchair Accessible.

---

[1] http://wheelmap.org/en/

[2] http://www.openstreetmap.org

Table 1. Wheelchair Accessible Data in Wheelmap

Categories	Yes	No	Limited	Unknown
All nodes	4687 (1.11%)	1001 (0.24%)	898 (0.21%)	415080 (98.44%)
Transfer	1445 (0.59%)	272 (0.11%)	158 (0.06%)	244094 (99.24%)
Food	1038 (1.82%)	367 (0.64%)	430 (0.75%)	55135 (96.78%)
Leisure	100	13	27	5744
Bank	351	43	33	15204
Education	132	15	32	11597
Shopping	767	118	123	33552
Sport	19	0	13	1851
Tourism	88	45	15	11459
Accommodation	66	35	13	6940
Misc	484	78	35	20648
Government	26	3	2	1910
Health	171	12	17	6946

2.2 Factual

Factual[3] is a location platform that provides over 65 million local businesses and points of interest in 50 countries based on the Mapbox map[4], which is also powered by OSM. There are three extended attributes datasets, namely restaurants, doctors and hotels. Both restaurants and hotel datasets provide an accessibility related attribute, called wheelchair accessible. However, only the Restaurants-UK dataset provides the accessibility information which is located within UK. Table 2 indicates the 210613 restaurant entities extracted from Factual API on 10/01/2014, 8904 entities are annotated as wheelchair accessible while 1786 entities are indicated not accessible. However, there are 199923 (94.92%) restaurants that are unknown for wheelchair accessibility. Besides the Wheelchair Accessible attribute, each restaurant contains 13 core business attributes including Name, Address, Phone, Category, Lat/Long, and URL as well as 43 restaurant specific attributes, such as Cuisine Types, Meal Types, Price Range, Rating, Hours of Operation, Payment Options, and Wi-Fi, which would also be of benefit for people with special needs.

Table 2. Restaurants UK Dataset in Factual

Wheelchair Accessible	Yes	No	Blank
Restaurants-UK	8904 (4.23%)	1786 (0.85%)	199923 (94.92%)

2.3 Step-free Access Guide, London

Transport for London (TfL) provides a series of Open Data for developers, which includes the step free tube guide data[5], station facilities and some other real time datasets. The step-free tube guide data contains the stations' step-free access information of London tube, DLR and Overground as well as the information of the step height and gap between the platform and the train for people with limited mobility. According to the XML based dataset downloaded on 10/01/2014, the step free access tube data is published by TfL and updated to 16/08/2013. There are 362 tube station entities with 5 core attributes, such as Station Name, Naptans, Links, Accessibility Interchanges and Accessibility. Table 3 states the 6 sub-attributes in the attribute named as Accessibility including accessible car park spaces, taxi outside station, specific entrance, accessible toilet, lift access and access type (shown in Table 4). Approximately 50% of all these 6 sub-attributes are annotated as either accessible or not, while the rest of the entities are displayed as unknown.

Table 3. Step Free Guide Dataset in TfL

Items	Yes	No	Unknown
Blue Badge Car Park Spaces	147 (40.61%)	35 (9.67%)	180 (49.72%)
Taxi Ranks Outside Station	13 (3.59%)	169 (46.69%)	180 (49.72%)
Specific Entrance	16 (4.42%)	141 (38.95%)	205 (56.63%)
Accessible Toilet	54 (14.92%)	128 (35.36%)	180 (49.72%)
Access Via Lift	98 (27.07%)	84 (23.20%)	180 (49.72%)
Limited Capacity Lift	8 (2.21%)	174 (48.07%)	180 (49.72%)

There is an Accessible Interchanges attribute, which presents the accessible metadata for interchanging to another public transport patterns. The problem for using this attribute is to associate with other interchange datasets to determine the initial interchanges in the stations. The information about the step and the gap between platform and train is indicated in Table 4. The attribute Step Max is the maximum height of the steps, which is useful for people with mobility disabilities or carrying the large luggage. The Gap Max attribute is the maximum height of the gap between platform and train. If the maximum gap is over 181 mm, wheelchair users would need assistance to access the train.

Table 4. Accessibility Dataset in TfL

Items	Full	I/C	Partial	None
Access Type	131	25	26	180
Items (mm)	**0 – 50**	**51 – 120**	**121+**	**Unknown**
Step Max	90	15	53	220
Items (mm)	**0 – 85**	**86 – 180**	**181+**	**Unknown**
Gap Max	33	112	23	220

2.4 UK National Rail Stations

As there are not open APIs or open datasets related to UK national railway stations' accessibility data, we extracted the data from the national rail website[6] and converted the unstructured data into a well-structured data. Based on the data we crawled on

---

14/01/2014, there are 2601 railway stations' data associated with accessibility information, such as ramp for train access, step free access, wheelchair available, accessible toilets, accessible ticket machines and accessible car parking for people with disabilities. There are some other attributes for the people with special needs, such as toilets, public Wi-Fi and baby changing facilities.

Table 5. Accessibility related data in National Rail data

Items	Yes	No	Unknown
Ramp for train access	1726 (66.36%)	842 (32.37%)	33 (1.27%)
Staff help available	1199 (46.10%)	1370 (52.67%)	32 (1.23%)
Step free access coverage	1347 (51.79%)	1220 (46.91%)	34 (1.31%)
Wheelchairs available	374 (14.38%)	2195 (84.39%)	32 (1.23%)
Toilets	792 (30.45%)	1778 (68.36%)	31 (1.19%)
Baby changing facilities	365 (14.03%)	426 (16.38%)	1810 (69.59%)
National key toilets	168 (6.46%)	623 (23.95%)	1810 (69.59%)
Accessible public telephones	152 (5.84%)	320 (12.30%)	2129 (81.85%)
Customer help points	1726 (66.36%)	843 (32.41%)	32 (1.23%)
Accessible ticket machines	995 (38.25%)	285 (10.96%)	1321 (50.79%)
Car Park	1548 (59.52%)	910 (34.99%)	143 (5.50%)
Accessible car park equipment	392 (15.07%)	1210* (46.52%)	999 (38.41%)
Public Wi-Fi	200 (7.69%)	0	2401 (92.31%)

Table 5 states the data annotation for each accessibility related attribute. The percentage of unannotated entities for almost have the attributes is approximately 1%. Some attributes, such as baby changing, national key toilets (accessible toilets), accessible phones and accessible ticket machine have more than 50% unknown annotation. Moreover, there is no geographic data in extracted data. Therefore, we combined the extracted data with rail references data download from National Public Transport Access Nodes (NaPTAN)[7] to generate a Nation Rail dataset with 9 core attributes, namely Station Name, AtcoCode, TiplocCode, Easting, Northing, and 13 accessibility related attributes as well as another more than 40 attributes.

2.5 AccessTogether

AccessTogether[8] is another crowdsourcing platform to provide additional accessibility data for geographical data. According to its schema, there are approximately more than 58 attributes related to accessibility, which could be divided into mobility disability,

---

* Wheelchair users may require assistance in this car park.
[7] http://data.gov.uk/dataset/naptan
[8] http://www.accesstogether.org

hearing disability and visual disability. Some attributes are very useful for people with disabilities, such as audio-descriptions, background-noise, doorways, fm-audio-system, sign-language-tours. However, these attributes are too complex and specific for both data reusing and crowdsourcing. Moreover, there is a Factual ID and Foursquare ID for each entity to reference to the entities in these two applications. Based on the dataset extracted on 10/01/2014, there are 4827 entities across 43 countries and 3788 entities annotated with 'none' for Factual ID. There are only 46 entities are in the UK, while 1147 entities in the US.

3. LESSONS LEARNED

According to the survey of open accessibility data demonstrated in previous section, there are several lessons learned:

- There are some accessibility related datasets available online, which refer to the real world locations.
- Accessibility data is from multiple heterogeneous resources. Some applications, such as Wheelmap, Accesstogether, Step-free guide feed and Factual, present the accessibility data in standard formats (XML, JSON and CSV et al.). Some data sources like national rail station data are mainly extracted from web pages and need to be converted to standard formats.
- There are various accessibility related attributes in the open accessibility datasets, but there is no standard guideline to specify attributes for accessibility due to the differences between various applications. For example, there is only one wheelchair accessibility attribute related to rail station in Wheelmap rather than a list of accessibility attributes in the national rail station dataset.
- The classification of accessibility data could be divided into three types to satisfy the needs of people with disabilities, namely people with mobility disabilities (wheelchair users), people with hearing disability and people with visual disabilities. However, it would be very complex if it refers to the people with special needs. Therefore, some datasets are restricted to one specific need, such as Wheelmap and Factual.
- Crowdsourcing is not enough due to too much incomplete information. For example, 98.44% of the entities in Wheelmap are annotated as unknown for wheelchair accessible and 20% of the entities are annotated with blank name. 94.92% restaurant entities in Factual are named with blank for wheelchair access. However, there are more than 50% of all entities annotated with accessibility attributes in both national rail station and TfL step-free access guide datasets, which are primarily published by the government.

Therefore, in order to fulfill the incomplete information and improve the accessibility data quality, we propose to apply the semantic web technologies and the Linked Data principles to interlink these datasets into a core accessibility dataset.

4. DATA INTERLINKING

The Semantic Web is known for its principles of knowledge sharing rather than document sharing in the traditional Web. As the core data layer in the Semantic Web, the Linked Data principle exposes the advanced characteristics for knowledge sharing, such as human and machine readable data, well-structured data, standard data format, domain specified, semantic linking and openness [2]. Ontology matching is the key to interoperability of the Semantic Web and a solution to address the problem of semantic heterogeneity. There are three different ways to integrate data based on ontology, namely the single ontology approaches, multiple ontologies and hybrid approaches.
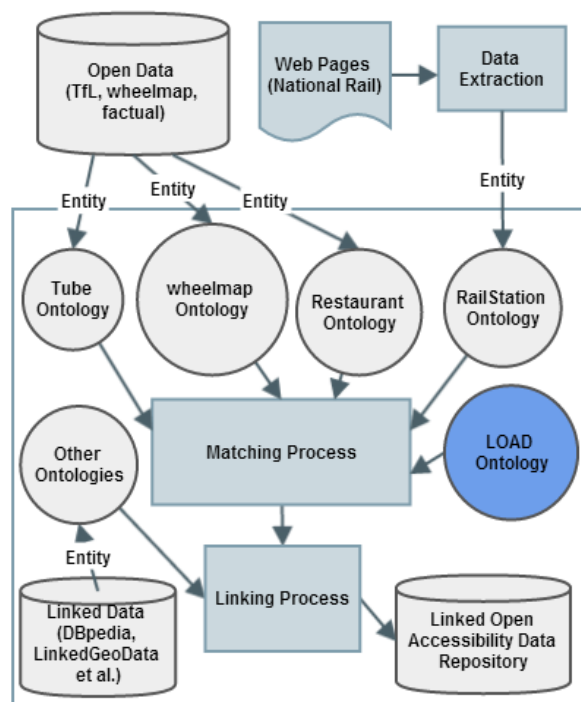
Figure 1. Linked Open Accessibility Data Architecture

Figure 1 presents the architecture of Linked Open Accessibility Data, which interlinks the datasets introduced in previous section. Currently, we would not map the AccessTogether dataset due to its complex schema and over specified attributes, which is very difficult to interlink to other datasets. All the three integration approaches could be applied in interlinking these data. However, the single ontology approach would be faced with the problems of the mappings between low-dimension entities (only wheelchair accessible) and high-dimension entities (multiple attributes for wheelchair accessible). Moreover, the mappings between the multiple ontologies are also a challenging issue, which is currently a research area involved a large number of research efforts. Therefore, we applied the hybrid approach by designing a top-level ontology called Linked Open Accessibility Data (LOAD) Ontology to define the basic accessibility terms in this domain, which could reduce the difficulties for multiple ontologies mapping. Based on the multiple heterogeneous schemas among the datasets, we design the individual ontology to represent the entities from each retrieved dataset based on its own schema, namely the tube ontology, the wheelmap ontology, the factual restaurant ontology and rail station ontology. There is geographical data available for each entity. However, the geographic reference system in National Rail dataset is using OSGB36 rather than WGS84. Therefore, we convert the geographic reference system to one standard system and use equivalent assertions (owl:sameAs) to match the same entities with their geographical data. We could also use some unique IDs across existing different datasets to map the entities, such as Factual ID, OSM Node ID, NaPTAN ID and Foursquare ID. The data integration progress also involves the mapping between the entities in open accessibility repository and other geographical resources in the LODC, such as DBpedia, Freebase and LinkedGeoData. There are some challenges for ontology reasoning and inference when querying this linked open accessibility data with disability types, such as the complexity rules for ontology reasoning and the matching between the needs

and open accessibility data [4]. Therefore, both the evaluation of this data integration process and rules for ontology reasoning are the main research for our next step future work.

5. CONCLUSION

We have described a survey of open accessibility data within the UK. Based on the open accessibility data extracted from several systems, we have learned some lessons from the current situation of these accessibility datasets. Although the crowdsourcing is a powerful approach to fetch millions of data, it is still difficult to obtain the accessibility data with good quality, compared with the datasets published by government departments. Therefore, we proposed the approach to interlink these accessibility datasets into a public Linked Data repository and also link this repository to other geographical resources in Linked Open Data Cloud. The next step of this research would develop a decision support application for accessible travelling to test and evaluate this linked open accessibility data repository. In the future work, we would also develop a crowdsourcing platform associated with other social networking services to gather much more data and improve the accessibility data quality. As a result, researchers could evaluate the accessibility level with these accessibility related geographical datasets and developers could use the RESTful API or SPARQL endpoint to produce more accessible applications. End users could get more useful accessibility information for their special needs as well as contribute to this dataset. As a consequence, this research would not only, benefit people with disabilities, but also contributes a novel way to the research of addressing the accessibility issues with the Linked Data principles.

6. REFERENCES

[1] Bekiaris, E. et al. 2007. Elderly and disabled travelers needs in infomobility services. *Universal Acess in Human Computer Interaction. Coping with Diversity (2007)*. (2007), 853–860.

[2] Bizer, C. et al. 2009. Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*. 5, 3 (2009), 1–22.

[3] Ding, C. et al. 2013. Travelling for all. *Proceedings of International Conference e-Society 2013. Lisbon, PT* (2013), 519–522.

[4] Li, Y. et al. 2012. RailGB : Using Open Accessibility Data to Help People with Disabilities. *Proceedings of the Semantic Web Challenge co-located with ISWC2012*. (2012), 1–8.

[5] Packer, T.L. et al. 2007. Understanding the complex interplay between tourism, disability and environmental contexts. *Disability and rehabilitation*. 29, 4 (Feb. 2007), 281–92.

[6] Vanderheiden, G.C. and Trace, R. 2009. Using Distributed Processing to Create More Powerful , Flexible and User Matched Accessibility Services. *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*. (2009), 438–444.

[7] Westcott, J. 2004. *Improving information on accessible tourism for disabled people*.

[8] Wiethoff, M. and Sommer, S. 2007. Specification of information needs for the development of a mobile communication platform to support mobility of people with functional limitations. *Proceedings of Universal Access in HumanComputer Interaction Ambient Interaction*. (2007), 595–604.