

GetThere: A Rural Passenger Information System Utilising Linked Data & Citizen Sensing

David Corsar, Peter Edwards, Chris Baillie, Milan Markovic,
Konstantinos Papangelis, and John Nelson

dot.rural Digital Economy Hub,
University of Aberdeen, Aberdeen, UK
{dcorsar,p.edwards,c.baillie,m.markovic,k.papangelis,j.d.nelson}@abdn.
ac.uk
<http://www.dotrural.ac.uk>

Abstract. This demo paper describes a real-time passenger information system based on citizen sensing and linked data.

Keywords: provenance, data quality, citizen sensing, linked data, semantic infrastructure, transport

1 Introduction

Real-time passenger information (RTPI) systems provide details about public transport, allowing passengers to plan and make decisions regarding their journeys. Typical requirements for RTPI systems include: 1) listing available public transport services; 2) providing timetable (schedule) information for those services; 3) providing (real-time) vehicle locations; and 4) providing details of disruptions. However, few RTPI systems exist in rural areas for a variety of reasons, including a lack of infrastructure for obtaining and providing real-time information [7]. As part of the Informed Rural Passenger project¹, we are developing *GetThere*, an RTPI system for rural areas. The *GetThere* system consists of a smartphone app, supported by a semantic infrastructure that integrates data from multiple sources (including users). This system has been deployed in the Scottish Borders, UK in partnership with First Group.

This demonstration² will show a typical use of the *GetThere* app to view timetabled and real-time vehicle locations for a selected route, contribute vehicle locations while making a journey, report a disruption event, and assess the quality of real-time locations with and without the presence of disruption. The demo will utilise the datasets and services shown in Fig. 1.

2 Information Ecosystem

GetThere is supported by a semantic information ecosystem (Fig. 1) itself underpinned by a series of ontologies. Semantic web and linked data technologies are

¹ <http://www.dotrural.ac.uk/irp>

² A video of the demo is available at <http://www.gettherebus.com/iswcdemo/>

used for data representation and storage within the ecosystem as they provide an effective approach for large scale data integration [4]. Further, accessing and storing data via SPARQL endpoints allows storage to be handled by technologies appropriate for the characteristics of individual datasets; for example, using RDF streams or a database with a R2RML wrapper for high throughput data.

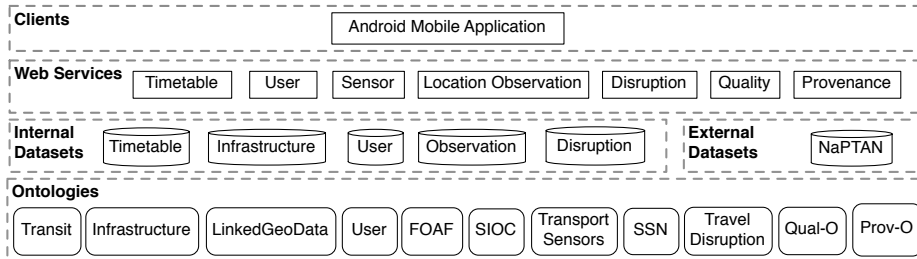


Fig. 1. Real-time passenger information ecosystem.

Details of public transport services and timetables are stored in the Timetable dataset³ and represented by the Transit ontology⁴. This dataset is used by the Timetable Service to provide details of available transport services, timetable, and vehicle location information to the *GetThere* app. The Infrastructure dataset provides details of the road networks used by public transport vehicles. This data is extracted from openstreetmap.org, and is represented using the Infrastructure ontology⁵ (which defines bus route maps) and the LinkedGeoData⁶ ontology. NaPTAN⁷ provides details of bus stops, including their IDs and locations.

The User ontology⁸ and dataset describe user profiles using SIOC⁹ and FOAF¹⁰, a description of each user's mobile device(s), along with details of public transport journeys made while using the *GetThere* app. The Observation dataset uses the Transport Sensors ontology¹¹ which extends the W3C Semantic Sensor Network (SSN) ontology¹² to describe observations (e.g. of vehicle occupancy level, vehicle location) obtained from users of the *GetThere* app. The Sensor service provides an API for storing and retrieving sensor and observation

³ Timetable information is received in the ATCO-CIF format (<http://www.travelinedata.org.uk/CIF/atco-cif-spec.pdf>); the RDF conversion program is available at <https://github.com/dcorsar/ecosystem.timetable>.

⁴ <http://vocab.org/transit/terms/>

⁵ <http://www.dotrural.ac.uk/irp/uploads/ontologies/infrastructure.owl>

⁶ <http://linkedgeodata.org>

⁷ <http://data.gov.uk/dataset/naptan>

⁸ <http://www.dotrural.ac.uk/irp/uploads/ontologies/user.owl>

⁹ <http://rdfs.org/sioc/spec/>

¹⁰ <http://xmlns.com/foaf/spec/>

¹¹ <http://www.dotrural.ac.uk/irp/uploads/ontologies/sensors.owl>

¹² <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn>

descriptions expressed using the SSN ontology; the Location Observation service handles real-time locations provided by app users. The Travel Disruption ontology describes different types of disruption, based on an analysis of existing travel disruption information sources [5]. Disruption reports from app users are managed and stored by the Disruption service and dataset.

Given the open nature of this data, issues such as data quality and trust naturally arise [6]. Examples range from malicious users and inaccurate devices to out-of-date information (e.g. timetables). As part of addressing these issues, the ecosystem features a service that can evaluate data quality. The quality ontology (Qual-O¹³), and its associated quality assessment service are discussed in detail elsewhere [1]. Briefly, the service is configured with a set of quality metrics encoded as SPARQL rules expressed against the relevant ontologies. These guide a SPIN reasoner [3] to perform quality assessment, producing quality scores which can be utilised by other services to filter low quality data.

Our current quality metrics are focused on real-time locations, and have been developed following several deployments of the system. They include: *Timeliness* - timely observations are under 1 minute old; *Accuracy* - accurate observations have a GPS error margin of less than 25 metres; *Relevance* - relevant observations are no further than 500 metres from the expected route of travel; *Availability* - observations with a high availability score have no more than a 1 minute delay between being created on the device and published by the ecosystem.

The provenance service uses the W3C Prov-O ontology¹⁴ to maintain a record of the entities, agents, and activities involved in producing data within the ecosystem. Uses of provenance include: associating users with location observations generated by their mobile device, which can support detection of potentially malicious users; and recording dataset provenance to ensure the latest timetable information is provided to users [2].

2.1 The *GetThere* Smartphone App

The ecosystem has been designed to support a range of applications through the creation of relevant application services. At present we have used the ecosystem to support the *GetThere* RTPI system, which is provided via an Android smartphone app (see Fig. 2). The app invokes the web services, which execute SPARQL queries against relevant datasets, process the results, and send a response to the app. Users are presented with a list of available bus routes; after selecting a route (and direction, either inbound or outbound), vehicle locations are displayed. These locations include both estimates based on the timetable and real-time locations obtained from other users on that route (Fig. 2, left screenshot). Bus stops along the route are also shown. The user can access timetable information for the previous and next arrivals at a particular stop. When the user boards the bus, they have the option of pressing a button to have their

¹³ <http://sensornet.abdn.ac.uk/onts/Qual-0.ttl>

¹⁴ <http://www.w3.org/TR/prov-o/>

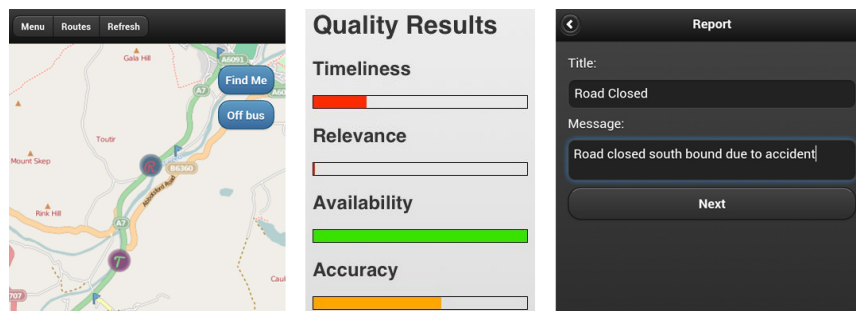


Fig. 2. Screenshots of the *GetThere* app showing (left to right): vehicle locations; the results of invoking the quality assessment service; and creating a disruption report.

location uploaded to the ecosystem every minute. The uploaded location is then used as the vehicle’s real-time location provided to other users.

Users can view quality assessment results for a real-time vehicle location by tapping its icon. We are working with users to determine an appropriate visualisation of quality results. Currently each assessed dimension is shown with a colour-coded bar representing its quality score (Fig. 2, centre screenshot).

Acknowledgements The research described here is supported by the award made by the RCUK Digital Economy programme to the dot.rural Digital Economy Hub; award reference: EP/G066051/1

References

1. C. Baillie, E. Edwards, P. Pignotti, and D. Corsar. Short paper: Assessing the quality of semantic sensor data. In *Proc. of The 6th International Workshop on Semantic Sensor Networks*, page to appear, October 2013.
2. D. Corsar, P. Edwards, N. Velaga, J. Nelson, and J. Pan. Exploring provenance in a linked data ecosystem. In P. Groth and J. Frew, editors, *Provenance and Annotation of Data and Processes*, volume 7525 of *LNCS*, pages 226–228. Springer Berlin Heidelberg, 2012.
3. C. Furber and M. Hepp. Swiqa - a semantic web information quality assessment framework. In *19th European Conference on Information Systems*, pages 922–933, 2011.
4. V. Lopez, S. Kotoulas, M. Sbodio, M. Stephenson, A. Gkoulalas-Divanis, and P. Aonghusa. Queriocity: A linked data platform for urban information management. In *The Semantic Web - ISWC 2012*, volume 7650 of *LNCS*, pages 148–163. Springer Berlin Heidelberg, 2012.
5. M. Markovic, P. Edwards, D. Corsar, and J. Pan. Demo: Managing the provenance of crowdsourced disruption reports. In *Provenance and Annotation of Data and Processes*, volume 7525 of *LNCS*, pages 209–213. Springer Berlin Heidelberg, 2012.
6. S. D. Ramchurn, T. D. Huynh, and N. R. Jennings. Trust in multiagent systems. *The Knowledge Engineering Review*, 19(1):1–25, 2004.
7. N. R. Velaga, M. Beecroft, J. D. Nelson, D. Corsar, and P. Edwards. Transport poverty meets the digital divide: accessibility and connectivity in rural communities. *Journal of Transport Geography*, 21(0):102 – 112, 2012.