# Building an Information Ecosystem for Public Transport in Rural Areas

Peter Edwards, David Corsar, John Nelson, Nagendra Velaga,
Mark Beecroft, Jeff Pan, Somayajulu Sripada
dot.rural Digital Economy Hub
University of Aberdeen
Aberdeen, UK
{p.edwards,dcorsar,j.d.nelson,n.r.velaga,m.beecroft,jeff.z.pan,yaji.sripada}@abdn.ac.uk

## ABSTRACT

Passenger information systems (PISs) providing real-time information are valuable tools for public transport users; however, in many rural areas, for various reasons, such systems do not exist. As part of the Informed Rural Passenger (IRP) project we are exploring how linked data from various sources, including the passengers themselves, can be used to construct an ecosystem upon which a range of applications and services can be built. In this paper, we discuss the need for such a solution, and describe its use within IRP to provide a rural PIS; we also discuss field trials of the *GetThere* iPhone app on three rural bus routes in Scotland.

## Categories and Subject Descriptors

H.4.0 [**Information Systems Applications**]: General

## General Terms

Design

## Keywords

Linked Data, Trust, Intelligent Transport Systems, Public Transport, Rural, Crowdsourcing

## 1. INTRODUCTION

The main aim of any Passenger Information System (PIS) is to provide users with accurate and efficient real-time travel and transport information. There are various benefits associated with the incorporation of real-time information in such systems, for example reduced perceived waiting time for passengers [6]. However, few real-time PIS are available for rural areas; reasons include: lack of infrastructure (e.g. vehicle tracking systems); fewer passengers, and so reduced operator incentive to provide real-time information; problems with mobile communication systems; and use of request stops by many passengers.

As part of the Informed Rural Passenger (IRP) project[1] we are developing a PIS which aims to address some of the existing issues in rural areas. Crowdsourcing techniques [3] play an important role, by allowing passengers to act as both consumers and providers of the information within the PIS. Through the use of smartphones, users are able to provide various types of real-time transport information, for example, location, vehicle occupancy levels, and facilities, which can then be used as part of our PIS.

The user supplied information is integrated using linked data principles [1] with open data from various sources, to construct an information ecosystem upon which a range of applications and services can be built. Use of open data and crowdsourcing potentially introduces imperfect (e.g. incomplete, erroneous, or fraudulent) data, making it necessary to perform assessments into the quality and trustworthiness of such data [4]. In addition, given the nature of the location data and other details supplied by passengers, the ecosystem must ensure their privacy, along with the security of the data they provide. We are investigating these issues, which we argue apply to any system which incorporates crowdsourcing, within the context of the IRP project.

## 2. RELATED WORK

Several projects have used the crowd to provide transport and travel information. The most relevant here is the Tiramisu transit system [7], which uses a mobile app to enable urban bus users to provide real-time bus location, occupancy levels, and free-text problem reports. Depending on availability, other users are presented with the real-time, historical, or timetable information. Tiramisu does not address the issues associated with the introduction of imperfect data into such systems, which is the main focus of our work.

Research in the area of provenance (a record of the processes and entities involved in producing a resource) indicates that it can play an important role in addressing these issues. Simmhan et al. [5] discuss how provenance can be used to support discovery of online information and trust evaluation, while Hartig and Zhao [2] describe its role in quality assessment

## 3. INTEGRATING AND ASSESSING DATA

Figure 1 outlines the information ecosystem that we are developing to address the challenges associated with construction of trusted linked data applications. The ecosystem consists of both open data sets and local data repositories only accessible within the ecosystem. Figure 1 shows a selection of the transport data sets currently being used by IRP. In addition, the ecosystem must support a range of annotations describing provenance, quality, and trustworthiness. These annotations are generated and used by an associated set

[1]http://www.dotrural.ac.uk/irp

*Digital Engagement '11* ,
November 15 – 17, 2011, Newcastle, UK

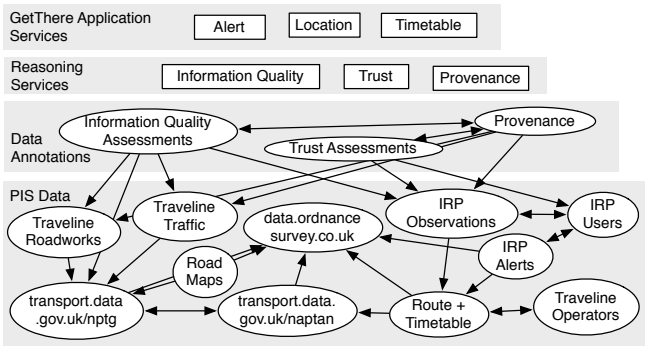


**Figure 1: Outline of the information ecosystem instantiated for IRP**

of reasoning services, which have access to both the annotations and the data. Finally, application services provide application functionalities using any of the other services, annotations, and data as appropriate.

# 4. THE GETTHERE SYSTEM

The GetThere system is a set of services and client applications, built using the ecosystem, that provide registered users with the ability to contribute and view travel information (for example estimated bus arrival times) via an iPhone app and website. Both the iPhone app and website allow new users to register with the system, login, view bus locations, and manage alerts. The website also provides more sophisticated tools for managing alerts, viewing previous journeys, and anonymising/deleting their data; whilst the app also allows users to upload information about the location of the bus on which they are traveling[2].

# 5. FIELD TRIALS

As part of our initial evaluation of the ecosystem, we have trialled the GetThere app using a number of iPhones on three rural bus routes in the North-East of Scotland. The purpose of these field trials was to investigate the performance of the app in a realistic, rural environment; and the suitability of the locations produced by the iPhone for use within the PIS. Table 1 presents a summary of the location data gathered during these journeys.

Analysis of the data collected found that, despite the expected unreliable mobile network coverage, and fewer number of location points generated by the iPhone 3GS, a sufficient number of location points were collected to enable real-time estimation of bus locations. The accuracy of the locations obtained from the device's GPS (those with accuracy under 100m), also aided this process. These findings were promising, as they show crowdsourcing rural bus locations using passenger's phones acts as a viable proxy for the real-time vehicle location data required by the IRP PIS.

# 6. CONCLUSIONS AND FUTURE WORK

We have demonstrated that the use of a range of open data sources combined with user observations can create a rich information ecosystem. However, given the sources of much of this information, imperfect content is likely, which may negatively impact on system performance and the user experience. We are therefore exploring how provenance and quality annotations and associated services can be used to address this problem, within the context of a real-time PIS for rural areas.

We are shortly to begin a one week pilot study with around 15-20 users on a bus route in the Aberdeen area. We are also planning a more extensive study with a community in the Scottish borders.

[2]This includes the longitude, latitude, estimated accuracy, speed, heading (movement direction), and timestamp as provided by the mobile's GPS device.

| Data set | Device | Mobile operator | # location points | # outliers (accuracy >100m) |
|---|---|---|---|---|
| 1 | iPhone 4, iOS 4 | o2 | 3694 | 27 |
| 2 | iPhone 4, iOS 4 | Orange | 4615 | 0 |
| 3 | iPhone 3GS, iOS 4 | o2 | 407 | 9 |

| Data set | Accuracy (in metres) | | | |
|---|---|---|---|---|
| | Average | Min | Max | Standard Deviation |
| 1 | 10.3 | 5 | 100 | 13 |
| 2 | 9.7 | 5 | 100 | 8.2 |
| 3 | 36 | 5 | 100 | 33 |

**Table 1: Summary of location data collected during field trials.**

# 7. 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

# 8. REFERENCES

[1] T. Berners-Lee. Linked Data. *IJSWIS*, 4(2):1, 2006.

[2] O. Hartig and J. Zhao. Using Web Data Provenance for Quality Assessment. In *The Role of the Semantic Web in Provenance Management Workshop at ISWC 2009*. SWPM, 2009.

[3] J. Howe. Wired 14.06: The Rise of Crowdsourcing, 2006.

[4] S. Ramchurn, T. Huynh, and N. R. Jennings. Trust in multiagent systems. *The Knowledge Engineering Review*, 19(1):1–25, 2004.

[5] Y. L. Simmhan, B. Plale, and D. Gannon. A survey of data provenance in e-science. *ACM SIGMOD Record*, 34(3):31–36, 2005.

[6] K. E. Watkins, B. Ferris, A. Borning, G. S. Rutherford, and D. Layton. Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A: Policy and Practice*, 2011.

[7] J. Zimmerman, A. Tomasic, C. Garrod, D. Yoo, C. Hiruncharoenvate, R. Aziz, N. R. Thiruvengadam, Y. Huang, and A. Steinfeld. Field Trial of Tiramisu : Crowd-Sourcing Bus Arrival Times to Spur Co-Design. In *CHI 2011*, pages 1677–1686, 2011.