UK Research and Innovation

Use case report for the Data Infrastructure for National Infrastructure project (DINI)

The Use of Quasi-Real-Time Data in Transport Infrastructure

Toby E. Willis

Dr G. Punzo

Prof. Daniel Coca

UK Research and Innovation

Contents

UK Research and Innovation

1. Use case Report

1.1 Background and Context

The specific focus of this project is in understanding the barriers and opportunities that exist around the use of Quasi-Real-Time data (QRT) in transport infrastructure. QRT is defined to be data that is used in an approximate 3-60 minute window of the events that it describes. We differentiate it from real time data which is used more immediately, with examples being the reporting of the exact location of buses, trains and aircraft, or the immediate guidance of autonomous vehicles. We also differentiate it from Batch Data, used on a long term scale (minimum 1 day but often months or years) which provides statistical information. Examples of batch data usage are the identification of optimal locations for physical interventions in infrastructure, such as the building of a new road, or the expansion of an existing one, as well as addressing required infrastructure maintenance in a timely manner.

This project contributes to the DAFNI DINI project by collating existing work identifying barriers and opportunities in transport infrastructure, as well as opening these barriers and opportunities to direct scrutiny and comment by key stakeholders in the area of national infrastructure.

1.2 Description of Activities

This project has three main phases. The first phase involved conducting a literature review into QRT, barriers that exist generally in transport and infrastructure and specific barriers that affect QRT in transport infrastructure. This would both contribute to the final report, but also assist in the identification of key stakeholders for the second phase as well as preparation of questions and background information.

The second phase involved conducting a series of interviews with representatives of key stakeholders. The findings of the literature review were put to them, as well as asking a series of other questions about both best practice and barriers in the area of transport infrastructure. Finally, they were given the opportunity to speak generally about their thoughts on the future of QRT in transport infrastructure. The full list of topics appears below, with the interviewees asked variations on the topics in order to best suit them to their area of expertise.

1. Describing the nature of their business and the way it relates to QRT.

2. In practice applications of data standards and the extent to which the alignment of data standards poses a barrier to QRT.
3. Their knowledge of the F.A.I.R principles and whether or not they are being followed.
4. The extent to which there is provision of open access data both from and to other stakeholders and barriers that prevent that provision where it does not occur.
5. A comparison of best practice between the key stakeholder and the others that they interact with directly compared to international best practice.
6. The specific limitations that prevent greater data sharing, in particular having the users address each of the 5 key areas identified in the literature review. That is:
    a. Legal
    b. Commercial
    c. Cultural
    d. Security
    e. Technological
7. A comparison of the use of pre-processed data and post-processed data as a means of communicating with other users. This has relevance in the level of technical skill required by the data users.
8. A discussion of the future of QRT.
9. What will remove barriers they currently face and allow more data sharing within transport infrastructure.

The third phase involves collating these opinions into a report, combined with presenting the collated opinions to each of the key stakeholders, ensuring that their words and opinions are represented accurately, as well as providing the opportunity for secondary comments.

1.3 Benefits of Data Sharing

Access to datasets about the current state of infrastructure can give useful information about the upcoming status of pieces of transport infrastructure.

One example of this is in horizontal data-sharing between digital twins. A digital twin uses a mixture of real time and quasi real time data to simulate the status of a networked transport infrastructure, allowing targeted interventions (such as the use of variable speed limits, diversions and road closures) to manage congestion and improve outcomes for the infrastructure users. These systems are affected by the impact of adjacent regions however, and a major crash closing a piece of infrastructure beyond the scope of the digital twin can still have major knock-on impacts.

Another example is in vertical data-sharing between infrastructure owners, data aggregators and commercial enterprises providing a service to the public. This information improves the

experience of the end-users of infrastructure in a number of different ways. This can be an improvement of waiting experiences (such as while waiting for a bus) by providing QRT updates on the arrival time, as well as information about the amount of free space available, both generally, but also specifically for those requiring specific accommodations such as the availability of accessible seating.

1.4 Barriers for Data Sharing

Through the literature review, a number of key barriers to data sharing were identified. These can be characterised as:

- Legal barriers (in particular, the requirements to maintain anonymity through the data)
- Commercial barriers (the costs associated with making high quality data available are sometimes significant)
- Cultural barriers
- Security barriers
- Technological barriers
- Additional further barriers have been identified through interviews with stakeholders
- Availability of skilled personnel capable of using data
- The balance of capital vs revenue based investment in public funding.

1.6 Results Data

The result from this project will be a report, shared with DAFNI for publication.

The data produced by this activity are interview transcripts with stakeholders, that, will be analysed and combined in a report together with the literature review conducted.

A data management plan was set up at the beginning of the project. Based on this,

- The interviews conducted will have the transcripts made available through the reports appendix.
- MP4 format recordings of the interviews from which the transcripts were made will be stored through the University of Sheffield's Network drive repository (X-drive).
- Recording and transcripts are to be made available upon approved request through the principal investigator.

A consent form for the collection and use of the interview data was signed by all the people interviewed.

The report will aim to characterise the:

- Owners and data types
- Specific gates from either individual data owner or individual data type (these could be delays)
- The way that QRT data are included in sharing agreements, that is whether they are shared as time batch, real-time API, the right to future access/removal, the ownership of the data and its transfer from the producer to the user

It will also include

- background information about the state QRT data accessibility worldwide,
- a literature review focussing on the barriers to QRT accessibility informed by our stakeholder interactions
- Case studies of QRT data successfully made available.
- A strategic development agenda toward QRT data accessibility.

1.7 Lessons Learnt and Recommendations

Recurring key opinions from stakeholders included repeated examples of international best practice, identifying the Dutch “Talking Traffic” as providing world-leading service, as well as identifying Japanese systems as providing a similar type of service. However, other stakeholders shared opinions that while the Dutch system is very strong, it presupposes public run infrastructure management. This stakeholder identified that a potential alternative system involving open access to public data supported by commercial interests utilising this data as efficiently as possible has the potential to provide a stronger service with competition driven innovation.

Overall consensus is that data standards are not a major barrier to additional data sharing, with most stakeholders identifying that they were able to access and use data shared without difficulties. A stakeholder representing a data aggregator and merchant identified that too stringent data standards had the potential to suppress innovation through over-prescribing technical standards preventing new entrants to the field from being able to compete.

Identifying data owners with access was said to be more difficult however, with stakeholders depending on personal knowledge and connections to bypass barriers.

Future funding in the area of Quasi-Real time data may be used to support the creation of a business model whereas the availability of data becomes

- more immediate and open to all private enterprises,
- conditional to the provision of a public good, different from the commercialisation targets. This may include, for instance, clean and accessible data products for the use of the public and researchers downstream.

There is more over the need to fill the gap in expertise on the problem-owner sides, so that skills about how to use the information in the data and how to get to it can be owned by the same set of professionals.