UK Research and Innovation

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

Urban Ride Research

UK Research and Innovation

Contents

UK Research and Innovation

1. Use case Report Template

The DAFNI-DINI project team have created a template to collect your project information.

This will be distributed to all project use cases to gather evidence for the DINI project.

The maximum page number for this report is 12 pages.

1.1 Background and Context

The Urban Ride Research project, that DAFNI has helped support, has been setup as an enabler project. Realising the significant amount of work that must often go into data sharing, the University of Bristol-based project team devised the project in the hope of overcoming some of the many barriers, and so opening up a novel data set to as board a group of researchers as possible. The data in question is electric-micro-mobility data.

Bristol and the rest of the West of England Combine Authority (WECA) area operates one of the largest electric micromobility short term rental schemes anywhere in Europe. Roughly 10 million journeys are made each year, by roughly 100 thousand active users, across an operating area of 65km^2. To make possible the operation of this fleet, the electric scooters and bicycles record and report back their location every 5s. This vehicle telematics data, produced as a byproduct of the scheme operation, therefore details every single journey made as part of the scheme. It provides visibility of not just the origin and destination of the journey, but the complete end-to-end route selection. This data could be extremely insightful, as people's choice of route is governed not just by which is the shortest path, but a wide range of different factors such as safety, comfort and even the state of the infrastructure.

The electric micromobility rental scheme is operated by a private company called Dott, although the scheme is licensed by WECA. WECA, and their legal consultants, also write the contract of operation. When the scheme was first launched the data it generated was retained by the operator (at that time a company called Voi). It was during this deployment that we at the University of Bristol approached WECA, and demonstrated to them the potential value of this incredibly rich data set, and the large number of policy relevant questions that the data might be able to inform on. Since then, later versions of the contract have stipulated that The Combined Authority by the ultimate owner of the scheme data, and have detailed specifically what data the operator must provide (mitigating **Barrier 1.3**).

Although WECA is now both the data owner and (in some cases) the policy body whose decision making can benefit from the insights the data contains, the complexity of the data, as well as the challenge of processing such a large dataset, means that they are not equipped to handle this data on their own. For this reason, we set up the Urban Ride Research project. This project aims to put in place a data sharing agreement that will allow researchers at the University of Bristol access this data, and generate the insights that will support policy makers to make more data driven decisions. (This project therefore aims to support WECA in overcoming **Barrier 4.4**.)

The Urban Ride Research Project has a number of different strands that have been working together to try and progress different aspects of the challenge. Since the key stakeholder (WECA) is a large political body, a key focus of the Urban Ride Research has been the generation of buy-in amongst those within the organisation who might not be directly involved, but who's opinion might help or hinder the project as it progresses (essentially aiming to overcome **Barrier 4.2**). To do this we have pulled together a pan-disciplinary network of researchers who are interested in the data, and run a series of workshops, during which we presented the data. These workshops were successful in generating an extensive list of different policy relevant insights that the data can provide. These insights are not just related to transport, but cover a broad range of topics including safety, the economy and flood resilience. This list was put into a document that was then shared within WECA, to demonstrate to the organisation the many advantages of supporting this project.

A further complication to handling the WECA micromobility data is its highly sensitive nature. If viewed in its entirety, the data set provides a complete record of every single journey that every single user has ever made as part of the scheme, alongside demographic information about that user. A key challenge of this project is setting up the Trusted Research Environment (TRE) that will allow this data to be accessed and processed safely. (It is through the use of a TRE that we hope to overcome **Barriers 1.1 and 1.2**)

It is this TRE element of the project that is of greatest relevance to DAFNI. So far, the work to produce a TRE has been carried out by the University of Bristol IT Services department. Although this is acceptable while the geographical scope of the project is limited to the Bristol area, there is no reason why this same approach couldn't be adopted more widely across the UK. If this were done it would be advantageous to have data from different cities stored centrally in a single location, as this would facilitate the sharing of algorithms and insights between different regional authorities. At that point it is no longer appropriate for the University of Bristol to be the sole holder of this data. DAFNI is the clear choice for this centralised TRE facility.

Barrier 4.3, concerning the ethics of data sharing, is also of consideration with such a large and sensitive data set. Our intention is to relay on the University of Bristol Ethics Application process, which already very mature, to assess the ethics of any pilot project proposals before providing access to the data.

Although we have overcome and mitigated a number of the different barriers that can prevent data sharing, it is **Barrier 3.3** (relating to contractual issues) that has delayed our project significantly. As will be explored later in this document, a separation between WECA (who strongly support the sharing of data) and their legal consultants (who seem disincentives to finalise the data sharing agreement) seems to have delayed significantly the creation of the necessary contracts, the finalisation of which is still ongoing.

There are a large number of different stakeholders to this project, including:

- Within the University of Bristol:
    - The Urban Ride Research Project Team (**Data Provider**)
    - The Network of Researchers interested in (and adding value to) the data (**Data User** and in some cases **Data Aggregator**)
    - The IT Services Department
    - The Contracts Department
- Within The West of England Combined Authority (**Data Owner**):
    - The Officers Involved in the delivery of the electric-micro-mobility scheme
    - Policy makers in areas informed by the mobility data.
    - The Data Services Department
    - The Contracts Department
- WECA have outsourced much of their contract writing to external consultants WSP.
- The scheme operator, currently Dott (**Data Creator**)
- Other regional authorities that operate electric micro-mobility schemes and might benefit from the data they generate. We have been targeting Gloucester Council and the Greater London Assembly.
- DAFNI

This project clearly falls into the domain of **Transport**, sitting primarily within **Mobility** although also relevant to **Road**. Researchers involved in the network have demonstrated the wide range of research areas that the mobility data can inform on, including **Infrastructure**, **Public Health**, **Safety** and the local economy.

1.2 Description of Activities

The original project plan was broken down into for work packages, these were:

- WP1: Accessing Combined Authority micro-mobility data
- WP2: Researcher/Combined Authority Codesign
- WP3: Expand project to other local authorities
- WP4: Scaling up approach and exploring DAFNI role

As will be described in more detail in Section 1.4, we have encountered a significant delay when it comes to the finalisation of the data sharing agreement that will enable data to be transferred between WECA and the University of Bristol. Despite the combined authority presenting us with a "final draft" of the agreement in March, legal consultants have since then repeatedly amended the document and delayed the final signing. There have now been nine versions of the document since that initial final draft. This has resulted in a delay of over five months when compared with the original project timeline.

Those work packages that were not dependant on the mobility data have continued as planned. After the research workshops exploring the data, a list of "pathfinder" project ideas was collated. This is a list of projects that will use the data to generate policy relevant insights which WECA might then use to inform decision making. This list was shared with WECA, who reviewed it, and indicated which of these path-finder projects most closely align with their current political priorities. When data is finally made available to researchers, this list will allow the project team to prioritise the path-finder projects, maximising the initial impact of the work. Path-finder projects will then go through a co-design process, where researchers work with the relevant individuals within WECA to maximise the saliency of the insights being generated.

Work to reach out to other regional authorities has started, including a conversation with the Head of Data for London, as well as discussions with Transport for London. The further division of responsibilities within London adds extra complexities to the task of achieving buy-in, as (unlike in the WECA area) the body responsible for licensing the electric-micro-mobility scheme is not the same as the body that would ultimately benefit from many the insights in contains. The delay to the signing of the WECA contract has also resulted in delay here, as our initial plan was to be able to present to TfL and the GLA real-world results from Bristol, that would help demonstrate conclusively the feasibility of the approach. Needless to say this has not yet been possible.

The final work package, which included exploration of how the TRE might be run on the DAFNI platform, has also carried on. The project team has had a number of meetings with DAFNI Data Scientists and the DAFNI Science Lead. From these it has become clear that DAFNI is already pursuing the implementation of TRE capabilities as part of the DAFNI platform, although a decision is still to be made as to whether this will be implemented in-

house or provided by a third part supplier. Further work will be carried out to make sure that the TRE solution being developed is able to deliver the requirements required for our project.

1.3 Benefits of Data Sharing

Although, as described, no data has yet been shared, our work creating a pan-disciplinary network of researchers has demonstrated extremely effectively the massive potential benefits of data sharing. The network consists of more than thirty active participants, who's expert fields include engineering, transport, mathematics, policy studies, geographical sciences, economics, neuroscience, business, the circular economy, hydrology, behavioural science and public health.

This diversity of expertise has meant that each of the different researchers has brought a different perspective to the data, and an understanding of how this extremely rich data set would be able to map onto their specific field. This, in turn, resulted in the generation of an expansive list of possible projects that might derive benefit and insight from the data. Some of these projects have already started to attract funding, some of which can be used to support the long-term operation of the Urban Ride Research project and its data. This ability to attract funding from a wide range of different sources is another clear benefit of extensive data sharing, as it adds a level of resilience to the projects that support the data.

A key aim of this project is to demonstrate to regional authorities the inherent value of sharing data with researchers. These authorities often do not have the skills to handle large data sets, even data which they already own. Neither are they always equipped to understand the potential insights contained with the data, and they are generally not nimble enough (and lack the funding) to investigate those insights that they are aware of. Furthermore, extra value can often be added through the merging of data, which can be challenging for a regional authority to achieve. For these reasons it is advantageous for an authority to share the data with research institutions that are better equipped to carry out this work.

The potential benefits identified by the Urban ride Research project have mostly been in form of policy insights, although in some cases insights have been of a more operational nature. For example, it has been suggested that by combing the mobility data with rain fall information it will be possible to identify areas where e-scooter riders are avoiding surface water flooding. This information could be fed into some sort of early warning system, identifying to the highways authority where there are drains that needed to be unblocked. This is just one example of the more than 30 different real-world applications devised by our research network.

1.4 Barriers for Data Sharing

The biggest barrier that we have faced is the delay of the signing of the contract (**Barrier 3.3** from Table 2.3). As mentioned above, despite reassurances that this document was nearly ready to be signed back in March, the Combined Authorities legal consultants have failed to produce a final version. This has caused significant frustration not just for the University of Bristol project team and researchers, but also within WECA where there is widespread buy-in for the project. It appears that, since the lawyers causing the delay are consultants, even those working on the project within WECA have very limited ability to influence their actions.

Legal scrutiny when carrying out data sharing is obviously crucial, especially when dealing with large amounts of personally sensitive data. Much of the frustration, however, comes from the apparent lack of progress since March. At that point a "final draft" was presented by WECA's legal consultants to us, and was deemed ready to sign (baring one very minor alteration) by the university contracts team. Since that draft, however, WECA's legal consultants have continued to work on the agreement, and it has now doubled in size and gone through 9 revisions. Despite all this work we still do not have a version WECA is willing to sign.

Quite rightly, some legal barriers will be encountered by almost everyone wanting to share data, although the scrutiny under which the public sector puts itself means that it seems to take much longer for that sector to over comes these barriers. The extra complication of the lawyer in question being a consultant (and thus potentially being paid by the hour to work on this) may also have led to a lack of incentive for them to work quickly on the problem, although that is speculation. Certainly, the significant changes to the document since the first "final draft" suggest that either the earlier version was woefully lacking, or the current version is woefully bloated, or both.

It is difficult to see an easy way of mitigating the impact of a barrier like this. In theory the process could be sped up through the use of very detailed templates, or previously existing contracts upon which new contract can be based, however considering the incentive structures in place there is no guarantee that the legal consultants in question would choose to adopt those templates.

1.5 Sources of data – table

Since starting this work, WECA has formalised the data relationship between them and the system operator. The operators is now contracted to provide WECA with data from the service, which WECA owns. In total there are 18 different data tables that the operator must provide to WECA, all of which we expect to eventually have access to. The most

relevant two of these have been listed below, but for a complete list please contact the Urban Ride Research project team.

Data Source	Data Description	Purpose	Technical Details	Data restrictions and Licence	Barrier	Stakeholder
Electric Micro Mobility Trajectory Data	The 5 second 'ping' data that gives a complete history of every journey ever done using the electric micro mobility rental scheme. This is **geospatial** micromobility data, and can be **personally sensitive**.	This extremely rich data set can be used to understand a considerable amount about a city, not just transport related, but also infrastructure and societal issues.	A group of geo-tagged time series journey records. Currently stored in WECA's data lake this can be exported to many different formats, csv being most likely at this stage. The data can contains a persistent user ID, which if included makes the data highly sensitive, requiring the use of a TRE for processing.	Although the data collaboration agreement is not yet signed between WECA and the University of Bristol, it is clear from the drafts that it will not permit the onward sharing of data. The collaboration agreement being drafted is however designed to act as a template that will allow other organisations to create new collaboration agreements more easily.	The completion of the contract has been the biggest barrier to accessing this data. This has been delayed for many months. So far efforts to expedite this process have had little to no impact. This is closest to **Barrier 3.3** from table 2.3	The impact of this barrier has affected all stakeholders, although seemingly the delay sits with WECA's legal consultants.
Scheme Users' Survey and Questionnaire Responses	As well as the data generated by the micro-mobility fleet, the operator also collects data about the users through surveys and questionnaires. This includes demographic information, as well as a responses to ad hoc surveys on a range of different	Combining demographic data with usage data will provide many unique insights into the factors that govern route selection. Path-finder projects that use this data include those looking at safety and	This data set is not as large as the trajectory data, and the format may well change depending on the nature of the questions being answered.	Since this contain demographic information, and there is likely to be some personally identified information with in that, this data will be treated as extremely sensitive and stored in the TRE. As above, the	This data will be covered by the same data sharing agreement as the trajectory data, and so has also been held up by **Barrier 3.3**.	As above.

	service related topics.	social inclusion.		current drafts of the data collaboration agreement negate onward sharing.		

1.6 Results Data

The delays to the contract signing mean that as yet no direct 'results data' has been generated. There have however been many findings from the work carried out that will be useful in the future. Something that has proven to be effective is our workshop-based approach for building a pan-disciplinary network of engaged researchers. The extensive list of different possible uses of the data is far beyond what we would have originally considered possible, and this highlights the importance of breadth and diversity when examining data. It also provides a clear example of the benefits of adopting a co-designed-based approach, as long as the advantages to all parties can be effectively articulated.

Our approach for generating buy-in within WECA seems also to have been effective as they continue to strongly support the project despite the delays. This sort of approach may have implications for future data gathering endeavours, as it provides a mechanism through which bodies that are not inherently pro the sharing of data can be encouraged to do so.

On one level the raw, unprocessed e-micro-mobility data will be "results" data for the Urban Ride Research project, and much is already known about what this data will look like when the contractual barriers are overcome. Data that contains any sensitive information will be placed on a University of Bristol hosted TRE, while low sensitivity data will be stored on SharePoint. In both cases, researchers that wish to use the data will have to apply to the Urban Ride Research project team to get access to the data. Before access is given they will have to carry out University of Bristol ethics application, although the project team will supported them with this process. Clearly this raw data cannot be made public, and that will be stipulate in the WECA data collaboration agreement.

As mentioned, WECA has provided a complete list of all the data tables that the micro-mobility operator is contractually obliged to provide, and which we hope to get access to. The size of these varies between quite small (for example the anti-social behaviour report table) and extremely large (such as the historical ping data which is likely to be billions of rows long). WECA are able to extract this data from their "data lake" in many different data formats, and so we will be guided by the researchers when it comes to what format is easiest

for them to work with. Initial conversations suggest that in many cases this will simply by CSV, but that might not always be appropriate.

On another level, since the Urban Ride Research project is designed as an enabler for other projects and researchers, it is the findings and data generated by this suit of projects that should be considered as our "results" data. At this stage little is known about the specifics results data that these projects will produce, as relatively few of these projects have been funded, and none have yet had access to the raw data from which findings will be derived. Some things are known however, namely that it is unlikely that the results data of these projects will contain personal data, as extraction from the TRE this will not be permitted. Wider sharing of this set of results data is therefore likely, but this will be dependent on the specific funder of that specific project. The Urban Rise Research project is highly focused towards supporting policy decision making, and certainly the list of pathfinder projects generated so far have focused on that as an output. We will be encouraging researchers to generate policy statements summarising the findings of any research they carry out in way that can be easily digested by policy makers at WECA. We are engaging the University of Bristol policy engagement teams to ensure that this is done as effectively as possible. These policy documents will most likely be shared beyond WECA, and certainly have the potential to be made open.

1.7 Lessons Learnt and Recommendations

The most salient lesson learnt during this project has been that delays can come from anywhere, and that the non-technical aspects of data sharing can consume a huge amount of time and resource, seemingly without showing any demonstrable progress.

Incentivisation of data sharing is crucial if results are to be obtained, and in this project we have seen both sides of this. Our work to incentivise WECA through co-creation of research projects has made this work possible, and the buy-in this has instilled continues to pay dividends as they still support the project despite the delays. On the flipside, WECA's legal consultant seem completely disincentivised to produce a final version of the Data Collaboration Agreement that is ultimately required, and this has delayed the project significantly.

Understanding incentivisation is crucial for broadening the impact of data sharing, and data sharing infrastructure, as in many cases data sharing efforts are brough together by a "coalition of the willing". Through appropriate incentivisation, such as the project co-creation carried out here, it might be possible to increase the size and scope of that coalition.

Something that has worked well has been the creation of the pan-disciplinary network of researchers. Face to face workshops, where people with data science expertise can present the data, discuss the techniques that might be used, and work through the potential insights the data might contain, has fast tracked the co-creation process, and allowed us to focus researchers' efforts to policy relevant pilot projects.

We will continue to apply for finding to carry on this work. The aspiration is for the Urban Ride Research project to receive funding that can then be handed on to pathfinder projects to support and strengthen the network of researchers. With pathfinder success we will be in a stronger position when reaching out to other regional authorities, and better able to demonstrate the advantages of adopting an Urban Ride Research-like approach. Further funding would also allow the project team to continue collaborating with the DAFNI technical team, supporting the design of any TRE capabilities developed for the DAFNI platform, and ensuring that it is appropriate for the storage of electric-micro-mobility data set.