FIT5146 Semester 2 2018

Case Study 4: Connectivity and Infrastructure Data to Support Connected and Automated Driving

Rapid advances in Connected and Autonomous Vehicle (CAV) development have been recognised as an opportunity to improve the safety, performance, efficiency and environmental impact of Australia's transport systems. The Australian government has commenced several initiatives to develop and deploy these new transport technologies (Austroads, 2018).

One of these initiatives is this research project, Connectivity and Infrastructure Data to Support Connected and Automated Driving. The research project is a partnership between Austroads, the Australian Department of Infrastructure, Regional Development and Cities and Monash University.

Connected vehicles are able to communicate with each other and their surrounding infrastructure. For example, navigation systems receive information about upcoming congestion through cellular signals and suggest an alternative route. The two main types of connectivity data available for current use is vehicle-to-vehicle (V2V) such as speed and position data, and vehicle-to-infrastructure (V2I) such as dynamic speed zone adjustment and event warning. Many cars are already making use of these technologies and are therefore considered connected vehicles. Connectivity is a requirement for vehicles to be autonomous, and the level of autonomy of that vehicles affects the type of data it both generates and requires.

The levels of automation for CAVs as defined by the Society of Automotive Engineers are:
- L0 (No Automation): Automated system has no vehicle control, but may issue warnings.
- L1 (Driver Assistance): Driver must be ready to take control at any time.
- L2 (Partial Automation): The driver is obliged to detect objects and events and respond if the automated system fails to respond properly.
- L3 (Conditional Automation): Within known, limited environments (such as freeways), the driver can safely turn their attention away from driving tasks.
- L4 (High Automation): The automated system can control the vehicle in all but a few environments such as severe weather.
- L5 (Full Automation): Other than setting the destination and starting the system, no human intervention is required in all driving modes.

(Engineers Australia, 2018)

The primary goal of the project is to *identify the key data and attributes that could be used to enable a CAV transport system in Australia*. The project will *identify current gaps in the data landscape and determine a range of solutions through collaboration with technical, policy and commercial experts*. The project is only concerned with data sources that are actually available or exist.

The aims of the project are as follows:
- Investigate the key road operator, infrastructure and geospatial data that will be used for future CAV systems in Australia.
- To identify gaps in-between the data provided to the developers of the CAV system currently.
- To determine what data will be required by the CAV connected infrastructure system to provide for interactive communications within the network.

- To determine the makeup of Public Private Partnerships with key network and road operators for the collection and transmission of such data assets.
- To produce a comprehensive report detailing the findings of this study with recommendations.

The project design involves working with selected participants from organisations involved in the broader research effort. Data will be collected in the form of workshops, semi-structured interviews and analysis of related documentation and data provided by the participants.

Datasets sources provided by participants will be analysed for the purposes of preliminary modelling by the NSW Data Analytics centre, as a number of the data sets are highly sensitive due to their commercial nature, this data has been licenced to the NSW Data Analytics Centre and will not be accessed or held by any other party including Monash University. This data is qualitative and is provided by Transurban. The results of the analysis and modelling will be provided to Monash University via a report that will not be published publicly. Network Blackspot data is particularly important. The 4G cellular network is critical to CAV operability. One danger this network presents are network blackspots. Telstra has partnered with the Australian Government to determine where these network black spots are and how to remedy them. This data and the other network performance data will be used for this study. Like Transurban data, it will not be held or accessed by Monash University.

Several other data sets will be retrieved and held by Monash University. These are all considered open sources and the results of their analysis will be published by Monash University. The nature of the stakeholders, the type of activities they will be involved in, and the types of data that will be collected for them are shown in the table below.

The aim for you, the **Monash University Research Data Management team,** is to **develop a Data Management Plan** based on **the first stage of this major project.** As an indication your team should consider your role in:

- Scoping the range of data sources currently available which may be of relevance to this case
- Managing contributions from key stakeholder groups to support the project development
- Working with experts to identify the key requirements for supporting collaborative data processes
- Identifying the potential data which will be produced during the early stages of the project and establishing an outline DMP

The table below contains the main stakeholder groups relevant to the case study. Note in some cases this highlights which groups may be involved in data governance, as data sources and/or data archive responsibilities. You will need to consider these in developing your Data Management Plan (DMP). Use the information here alongside the DMP Template and provide your reasons for noting particular data requirements, based on this table.

Stakeholder	Stakeholder Role	Activity	Primary Data	Prompt for your DMP
Bureau of Infrastructure, Transport and Regional Economics (BITRE)	BITRE is a government department responsible for the project oversite. They also manage a large number of datasets and act as a source of expert advice for transport policy as data analysts, modellers as well as social researchers.	• Workshops • Surveys • Provision of nominated documents	• Video recordings • Photos • Drawings and diagrams • Nominated documents	Determine two documents of relevance to developing the DMP. Provide an explanation of the rationale for referring to this type of information.
SMART Infrastructure Facility, University of Wollongong	The SMART team is a key collaborator in the project and has a facility dedicated to supporting infrastructure planning and management. They will coordinate a series of researcher workshops and public focus groups to initiate the project.	• Focus groups	• Audio recordings • Workshop and focus group materials	Outputs from these activities are a primary source of data generated from this project. However, when analysed, this data will be transformed into at least two other formats.
CSIRO – Data 61	CSIRO is Australia's leading government research agency and is responsible for large scale national research projects. They also have a strong remit to engage with business and support innovation.	• Data analysis and exploration • Workshops	• Formatted analysis and reports • Video recordings	Explore the data available via CSIRO and identify three open datasets that are appropriate for the study. Explain your rationale for choosing these

Stakeholder	Stakeholder Role	Activity	Primary Data	Prompt for your DMP
	Data61 is the primary group responsible for supporting Australia's data driven future and provides a range of resources for innovation to small and large businesses alike.	• Provision of nominated contributed data sets		and what they may contribute to the CAV system effort.
Transurban	Transurban is the lead partner in conducting tests in connected and automated vehicle trials. They hold substantial information on existing trials and have a range of expertise to support this project. They also conduct market research in interest and concerns from the public about the potential impact of CAVs	• Workshops • Provision of nominated datasets to the NSW Data Analytics Centre	• Video recordings • Drawings and diagrams • Nominated datasets	Consider the types of data that will be required from Transurban. Why are these important? How can existing data be managed in combination with collected data throughout the project?
Transport for Victoria	This group brings together Victoria's transport agencies in supporting a coordinated approach across the roads, public transport and ports development authorities. This group will have a key governance role in supporting the project and is a key stakeholder in the potential outputs.	• Focus groups • Provision of nominated data sets	• Audio recordings • Nominated datasets	Consider the types of data available from the Transport for Victoria consortia. Which might be particularly appropriate in this case?
CSIRO / Data61	CSIRO is Australia's leading government research agency and is responsible for large scale national research projects. They also have a strong remit to engage with business and support innovation. Data61 is the primary group responsible for supporting Australia's data driven future and provides a range of resources for innovation to small and large businesses alike.	• Data analysis and exploration • Workshops • Provision of nominated dataset	• Formatted analysis and reports • Video recordings Image based nominated dataset	Explore the data available via CSIRO and identify open datasets that are appropriate for the study. For example, one nominated dataset is the 'Data61 Pedestrian Dataset'. Consider and explain the ethical implications of using this open data. Why would we want to include this data in the project (hint: you may wish to read about V2P connectivity)
NSW Data Analytics Centre	The DAC is involved to support the data analytics relating to synthesising outputs	• Data analysis, exploration,	• Formatted analysis and reports	How might data be effectively shared with the Data Analytics

Stakeholder	Stakeholder Role	Activity	Primary Data	Prompt for your DMP
	from the project. The group has an interest in supporting whole of government approaches which inform policy and public spending on major projects.	modelling activities		team? What would need to be considered in enabling data sharing with another state agency?
Telstra	Telstra is the major telecoms partner in this project supporting the technical connectivity aspects, as well as broader communications infrastructure development, for the project.	• Provision of nominated datasets to the NSW Data Analytics Centre	• Nominated datasets	The data from Telstra (and Transurban) has been treated as commercial in confidence. This is a significant legal consideration in this project. Consider what this means in terms of access, storage and use.
Accident Research Centre, Monash University	The Accident Research Centre at Monash has over 30 years of research looking at a range of measures to mitigate the effect of vehicle accidents. Their data includes evaluations of road infrastructure and road safety data systems. In addition to providing data this group has an interest in understanding social behaviour with potential human driven vs automated vehicles and will run a major public survey to explore this area.	• Provision of nominated datasets to the • Workshops • Surveys	• Video recordings • Completed surveys	The surveys can be disseminated, collected and analysed either digitally or physically. Determine how this is format is considered in the DMP. How are privacy, practicality and ethics clearance managed and why are these important in this case?
ABS	ABS provides a range of secondary data required for the project.	• Workshops • Data analysis and exploration • Provision of nominated data sets	• Audio recordings • Formatted analysis and reports • Nominated datasets (see below)	You will be using geospatial data from the ABS. The ABS geographical and spatial data is of interest to this project. One issue regarding the use of this data is that the spatial and statistical communities have different operational metadata capabilities. How are different metadata environments being

Stakeholder	Stakeholder Role	Activity	Primary Data	Prompt for your DMP
				consolidated by the ABS and what influence could this have for the choice of metadata standards for this DMP?

Possible sources of data are:

- BRITE. (2018). Publications. Retrieved from https://bitre.gov.au/publications/
- CSIRO. (2018). ABOUT-DATA-GOVE-AU. Retrieved from http://kn.csiro.au/about-dataset-list/data-gov-au
- CSIRO. (2008). Data61 Pedestrian Dataset. Retrieved from https://data.csiro.au/dap/landingpage?pid=csiro:23454
- Queensland Office of the Information Commissioner. (2018). Breech of confidence. Retrieved from https://www.oic.qld.gov.au/annotated-legislation/rti/schedule-3/8-information-disclosure-of-which-would-found-action-for-breach-of-confidence/section-81/breach-of-confidence
- Australian Bureau of Statistics. (2018). Geography. Retrieved from http://www.abs.gov.au/websitedbs/D3310114.nsf/home/geography
- Australia and New Zealand Infrastructure Pipeline Portal by Infrastructure Partnerships Australia http://infrastructurepipeline.org/
- data.gov.au
- ABS http://www.abs.gov.au/
- Data from the NSW government Centre for Road Safety Cooperative Intelligent Transport Initiative (CITI). CITI Light Vehicle study http://roadsafety.transport.nsw.gov.au/research/roadsafetytechnology/cits/citi/light-vehicle-study.html
- Vicroads (2018). Vicroads Transport and Open Data API. Retrieved from http://api.vicroads.vic.gov.au/

General prompts for the DMP

Privacy and confidentiality/ Ethics

What are the potential privacy and confidentiality considerations to note in the DMP in this case?

What specific considerations in data curation does this project present?

Data storage and security

What are the data formats, both acquired and produced in this case?

Based on the potential formats, how and where should data in this case be stored? Why?

What are the likely considerations for storage and security?

Research outputs

What are the potential outputs from this project?

How will the DMP support the sharing of these outputs?

Why will making outputs more broadly available be important for this particular project?

What risks may need to be considered and noted in the DMP for sharing datasets as outputs?