FIT5146

RESEARCH DATA MANAGEMENT PLAN TEMPLATE

INTRODUCTION[1]

Data management planning is an important part of the responsible conduct of research. Data that is created, compiled or collected during any research project is a valuable asset that needs to be cared for over long periods of time.

This template is based on a working example used at Monash University which aims to introduce you to:

- Documenting research data management activities;
- Identifying areas of potential difficulty or conflict that need to be resolved with your supervisor;
- Finding out more about data management services and tools available at Monash University, and how to access them.

NATIONAL AND INSTITUTIONAL CONTEXT

The Australian Code for Responsible Conduct of Research was updated by the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) in June 2018. The Code requires aspects of data management such as ownership, ethics, and retention to be well-documented by researchers. While there is currently no national requirement to lodge formal data management plans, it is very likely that ARC and NHMRC will require greater evidence of data management planning in future.

Monash University's Research Data Management Policy recommends that all researchers, including HDR students, undertake data management planning. This aims to enable the University to better respond to changes in the wider policy and funding environment.

BEFORE YOU START

Your team should become familiar with the following resources:

- The Australian Code for Responsible Conduct of Research (2018)
  Monash University's Research Data Management website

ABOUT THE TEMPLATE

- In real world research the Data Management Plan may be developed over the life of the project. However key requirements need to be in place for many projects, such as where human and animal subjects are part of the research process, these must be clarified up front or ethics approval may not be granted.
- For your case study, considerations of the ethics requirements will inform how the Data Management Plan is developed in terms of specific requirements and potential constraints.
- For this exercise we recommend the team having an overview of the whole plan and then separating out sections for the each team member to cover. Note that some elements of the plan are dependent on others so you will need to be clear what you are doing as a team for the overall plan to make sense.
- Multiple-choice boxes make completing the template relatively easy to note the key expected requirements.
- However as you are developing a more comprehensive data management plan, you need to add supplementary information to explain your choices. You can also attach other relevant documents as appendices if needed.

[1] This TEMPLATE has been adapted for FIT5146 Assignment 2, Semester 2 2018, from the MONASH DATA MANAGEMENT CHECKLIST

ADDITIONAL RESOURCES

Online resources

The template contains links to a number of relevant Monash University Research Data Management Guidelines, as well as other sources of information and advice.

You should refer to these guidelines as needed in the sections where additional links are given. These guides will help inform you about the types of data and other considerations you need to make in completing the template

One-on-one support

At Monash faculty Contact Librarians provide support with Data Management Planning. Neil Dickson, Data Manager at Monash (based at Clayton) will be contributing to the **FIT5146 lecture in Week 5** and can answer questions about data requirements at Monash during the lecture.

WHAT HAPPENS WITH A COMPLETED DMP?

- Researchers are expected to retain a copy of the completed DMP with their research data as this forms part of the research record.
- The DMP also provides a discussion document when talking to others about your data management needs.
- As data management requirements may change as research progresses: policies, legislation, personnel and technologies all evolve, and refinements to research methodologies are common, the DMP needs to be updated to reflect the actual state of the research project's data management approach.

RESEARCH DATA MANAGEMENT PLANNING TEMPLATE

1. RESEARCH PROJECT TITLE:

CASE STUDY 4: CONNECTIVITY AND INFRASTRUCTURE DATA TO SUPPORT CONNECTED AND AUTOMATED DRIVING

2. YOUR TEAM STATEMENT OF CONTRIBUTION

NAME	CONTRIBUTION (BRIEFLY STATE WHICH SECTIONS EACH TEAM MEMBER COMPLETED AND WHAT WAS INVOLVED)	SIGNATURE
SIDDHANT SHARMA	**A.** DATA IN THIS DATA MANAGEMENT PLAN **B.** OWNERSHIP, COPYRIGHT, INTELLECTUAL PROPERTY (IP) **C.** ETHICS REQUIREMENTS **OTHERS :** REPORT REVIEW, CORRECTION, FORMATTING AND AGGREGATION OF ALL THE SECTIONS	**SIDDHANT**
SUYASH SATHE	**D.** DURABLE FORMATS **E.** STORAGE AND BACKUP **F.** SHARING DATA AND CONTROLLING ACCESS	**SUYASH**
HAO CHEN	**G.** DOCUMENTATION AND METADATA **H.** RETENTION AND DISPOSAL **I.** REPOSITORIES AND ARCHIVING	**HAO**

PLEASE ENSURE ALL TEAM MEMBERS SIGN THIS BEFORE SUBMISSION

PLEASE UPLOAD ONE COPY TO MOODLE PER TEAM.

3. PREAMBLE (250 WORDS MAX) Provide a brief overview of the key aspects your team will focus on in this developing this DMP based on your understanding of the case study. You should note the context and the purpose for data management of importance in your case.

A data management plan or DPM is a formal document that outlines how data are to be handled both during a research project, and after the project is completed. This data management plan has been created for a project related to connected and autonomous vehicle (CAV) development. The projects is being run in a partnership between the Australian department of infrastructure, regional development and cities and Monash University. Due the data being highly sensitive and commercials in nature, data has been licensed to NSW Data Analytics Centre.

Our team from Monash University has been assigned the task for creation of an efficient data management plan. There are quite a number of data which is being referred from third party resources and open sources under contractual agreements in association with open Data as well.

Clear articulation in relation to the version control, naming conventions and format of the data would be articulated as part this DMP. A couple of Governance committees would be required to be formed for privacy and intellectual property concerns and making sure the project runs in a smooth manner. There would be different metadata standards which would be used as part of different data files and sets. Details for the same would be included in the DMP.

Data storage plan has been also discussed and advised in the DMP with relevant explanation for the suggested decisions and recommendations.

4. **GROUP REVIEW (250 WORDS MAX)** Provide a brief summary of the key data management requirements your team has identified as a result of developing the DMP. You should note four-five key aspects identified in the DMP which you considered most relevant for your case.

As the project and case study is related to an area which is coupled with quite a number of commercial and confidentiality agreements, a structured and clear data management plan becomes as a very useful tool handy during the course of project. It assists in easy project management, clarifies the needed budget, makes sure the data is fairer to all and sets the accountability within the team.

The major elements of a data management plan which would need to be analysed are:

Assess, re-use & sharing,  servation of data, Data retention and disposal policy, Different data formats being used, Storage requirements, few institutional matters like legal requirements.

Details related to the data files which are being generated and used with the specific data formats have been discussed in relation the project. As this project is related to processing based on sensitive data, DMP discusses few topics which are highly related in detail.

Handshake of information  ngst different stakeholders has been tried to be streamlined by suggesting the governance structure.

HOW TO USE THIS DMP TEMPLATE.

Use this template alongside the information provided in your case study.

The DMP must reflect the specific requirements and situation related to your case.
Note we are asking you to consider the stakeholders, data sources as data requirements and likely data captured during the initial stage of the research project by documenting a Data Management Plan for these elements.
We are not asking you to do the research project or create a data management system for the research project.

Use the left hand column to note the data requirements considered by your team to be relevant to your case study.

In the right hand column provide a brief explanation of why you have noted particular requirements
You should refer to the prompt questions given in the case study stakeholders table.

Note there are links to additional resources throughout this template which you should refer to in completing these sections.

The key sections in the template ask for information on:

A. DATA IN THIS DATA MANAGEMENT PLAN

B. OWNERSHIP, COPYRIGHT, INTELLECTUAL PROPERTY (IP)

C. ETHICS REQUIREMENTS

D. DURABLE FORMATS

E. STORAGE AND BACKUP

F. SHARING DATA AND CONTROLLING ACCESS

G. DOCUMENTATION AND METADATA

H. RETENTION AND DISPOSAL

I. REPOSITORIES AND ARCHIVING

A. DATA IN THIS DATA MANAGEMENT PLAN					EXPLANATORY NOTES SHOULD BE ADDED IN THIS SECTION
DATA FILES PRODUCED BY THE PROJECT					
Data File	Format	Primary or Secondary source	To be archived (Y/N)?	Open access (Y/N)?	
Requirement Specification Document	WORD,PDF	Primary – Regular Updates Secondary- Archived	Y	N	Requirement Specification Documentation is the document which is used in order to capture all the requirements which are beings planned to be addressed and developed as part of this project. For this project both Word and PDF formats are used for this document. Word document is the running document in which all the regular updates are to be made. Proper version control would need to be maintained for this document. As the word document is platform and software document, this specific document would be converted into PDF format for archival purposes. PDF is a file format that is used to represent documents independent of the application software, operating system or hardware used to create them. So one top benefit of the format is its true mobility that allows it to be accessed on a huge variety of operating systems and platforms. Also to address the growing need for a standard electronic document format, the ISO agreed that PDF would be the format of choice for long-term preservation and archive of documents. Hence, for the regular updates and changes to be made, word document should be used. On the other hand for achieving purpose pdf format will be used. As this is amongst one of the most important documents from the project, hence a permanent archiving of all the version of the document would be required. One copy of the document in word format would be available for making the changes and the previous base lined document would be archived. This document would not have open access provided for the users. Only authorised users would have access to this document.
Design Specification Document	WORD,PDF	Primary	Y	N	Design Specification document is the detailed design documented for this project. It consists all the details related to the automated vehicles and the way how they would be communicating with cellular signals to commute. Similar to the above document it would be available in both word and pdf formats. Similarly it would also be archived with the framework

Vehicle Network Maps	PNG	Primary	Y	N	Vehicle network maps for the project have been designed and saved in the PNG format. PNG format has been selected for the archiving of these network maps as they have a better phototropic images and also have lossless compression while compressing them. As these are part of original project design, these would be required to be archived for a long term along with other design documents. Again due to criticality of the information access for this is not open to all. Only the authorised users can access this.
Vehicle Communication Log File	TXT	Primary	N	N	Vehicle Communication log files are normal text files which log all the activity performed by the vehicle while in use. This file need not be archived as the file is huge and this would not be required on a long run. These files would be retained into the current system for 30 days. Only the latest 30 days log file would be retained. Authorized user access only.
Vehicle Testing Report	PDF	Primary	Y	N	As part of the development for autonomous vehicles, testing would be have been performed on various levels and datasets. Same had been articulated in the document along with all the test result for future reference. This would need to be archived in the pdf format as there might be a need to references back to the test cases tested in an event of issue. Authorized user access only.
Vehicle Design Diagrams	PNG	Primary	Y	N	In this project, the design of vehicles is also very important in order to make this project a success. Hence all the designs have been archived for long term in PNG format. Authorized user access only.
Surveys related to this project	HTML/CSV	Primary	Y	N	The surveys had been conducted within the general public by the project team for understanding the expectations of public from this project. The same done by a web-form and all the required data was collected through a webpage. Later in order to preserve it better and understand the data set, it is transformed into a CSV format. CSV format is almost processed in all the applications and operating systems. Also, it is a tabulated form which assist while going the data set. This being sensitive data in relation to data which had been collected this would be required to archive with other files. Authorized user access only.
Training Videos	MPEG	Primary	Y	N	All the training videos which had been provided to the researchers into this project has been archived in the MPEG format. All the videos would be archived for training purposes of new researcher's team and employees.
Presentations	PPT,PDF	Primary	Y	N	As there were quite a number of researchers, stakeholders and partners involved, majorly presentations were discussed. As PPT is not compatible on all the platforms hence, the PPT is not archived in its raw format but again converted into pdf format. As all senior level executive were mostly part these meeting hence, the data is crucial and needs to be archived. Also authorized user access only.
Financial Documents	CSV,PDF	Primary	Y	N	As this being a commercial project, there are quite a number of financials involved, hence the financial documents are prepared and archived in both CSV and PDF formats. Financial documents have legal obligations hence would be required to be retained for at least 15 years.

THIRD PARTY DATA

Data File	Format	Primary or Secondary source	To be archived (Y/N)?	Open access (Y/N)?	
Blind Spot Details Sheet	CSV	Primary	N	N	This is a third party data collected from Telstra under licensed contract. As this is a very sensitive data and re-production of this data is prohibited. The project team is under legal obligation while using this. With improvement in the infrastructure and technology these details get changed. Also as this is a third party data, there is no requirement to archive the data. The CSV dataset might be retained until the next updated CSV file is shared by Telstra. Authorized user access only.
Traffic Details Report	JSON	Primary	N	N	Traffic detail reports are the dynamic live streaming data which is sent across to the vehicles. All the communication happens in JSON objects as these are easily consumed by the Java REST APIs. As this is a live streaming data, achieving for the same is not required. Authorized user access only.
Geospatial Data	ISO	Primary	N	N	All the geospatial data is shared in the ISO format. ISO is the disk image of an optical disk. As the geospatial data needs to be very informative the best possible format for this is the ISO format. Authorized user access only.
Surveys Conducted by Accident Research centre, Monash	TEXT/CSV	Primary	Y	N	Quite a number of surveys had been conducted by Monash Team using the flyers and capturing the required details in the same. Post survey completion the results are transformed into CSV format. Archiving for the same is required as it provides insights on some specific data points. Authorized user access only.
Transport Policy	PDF	Primary	N	N	Transport policy document is received in the pdf documents. It is used for accessing the rules and regulations for transport. No need for archival of the same. Authorized user access only.
Cellular Data	JSON	Secondary	N	N	Cellular data is a real time streaming data which is used by the vehicles to move. As these are based on the JAVA REST APIs, JSON is the most appropriate data format for this usage and purpose. This real time streaming data might not be required to be archived. Authorized user access only.

OTHER PROJECT RELATED DATA FILES

Data File	Format	Primary or Secondary source	To be archived (Y/N)?	Open access (Y/N)?	
Meeting Audio Recordings	MP3	Primary	Y	N	All the meetings being conducted had been recorded for future references. MP3 has a low compression ratio and data size. As this consists of quite a number of detailed information discussed over meeting hence the archival of these files is required. Authorized user access only.
Meeting Notes	Electronic/TXT	Primary	Y	N	These are informal notes which are prepared by the meeting conveyer during the meeting discussions. It can be electronic or textual format. Archival of this is required for future references. Authorized user access only.
Video Recordings for meeting and presentations	MPEG	Primary	Y	N	Video recordings for the presentation presented in the meeting with the senior executives have been recorded and would be required to be archived for future references. All these meeting had critical discussions in relation to project. Authorized user access only.
Emails	PST	Primary	Y	N	Emails had been the backbone of communication channel during the project execution a planning. All these emails have been working on a different dedicated exchange server which had been allocated separately for this project in order to maintain privacy. The dedicated exchange server have been built with the self-archival policy. Authorized user access only.
Research Papers from journals and Books	PDF	Secondary	N	Y	As part of the data management planning and the execution of actual project while a number of research papers had been reviews and studies. All these journals and research papers where in open data space. No Archival required for the same. Can be access by anyone is the open data space.
Zipped Address Details	ISO	Secondary	N	N	Zipped address details have been used as part of the project on autonomous commuting of vehicles. These details have been shared in the ISO format. As this is a secondary data set, archival for the same is not required. Authorized user access only.

PUBLICATIONS

Data File	Format	Primary or Secondary source	To be archived (Y/N)?	Open access (Y/N)?	
Privacy Policy Document	PDF	Primary	Y	N	Privacy Policy Document would be published by the privacy committee during the course of this project. It lets the researchers and other stakeholders know the privacy related laws. This would be required to be archived. Authorized user access only.
Intellectual Policy Document	PDF	Primary	Y	N	Intellectual Policy Document would be published the Intellectual policy committee of the project. It would clearly articulate the rules and regulation around the intellectual property on the project and assist in having clear polices for the stakeholders and project partners. The document would be required to be archived. Authorized user access only.
Data Retention and Disposal Document	PDF	Primary	Y	N	Data retention and disposal document would clearly articulate the data retention policy and disposal policy for the data which is being used and generated by automated vehicles. As this is a process document, retention of the same is required. Authorized user access only.
Metadata Inventory Document	PDF	Primary	Y	N	Metadata inventory document is an important document in relation to having all the details related to metadata and description of data types. As this is a one stop document to understand all the data sets which are being used as part of the project hence archiving of the same is required.
Research Papers on the Patents	PDF	Primary	N	Y	During the project, quite a number of new inventions were made by the team of researchers. Patents and research papers for all these inventions were published into various journals and research books. As post publishing it with the journals, the research papers are available in open data space. Hence archival for the same is not required. Access is open to all.
Final Research Report	PDF	Primary	Y	N	This is most important document of this project. The complete synopsis and detailed report is present in the report along with all the required artefacts and relevant examples. This report needs to be confidential and even not available for the partners of the project openly. It should be accessed through a licensed or contractual agreement.

B. OWNERSHIP, COPYRIGHT, INTELLECTUAL PROPERTY (IP)	KEY DOCUMENTS ON THIS TOPIC	EXPLANATORY NOTES SHOULD BE ADDED IN THIS SECTION
COPYRIGHT PROTECTION 1. ■ The data is protected by copyright. This will apply to most research data. 2. ■ The data will be collected, created or compiled	Research data management guidelines: ownership, copyright and IP Intellectual Property Framework Statute 11.2 IP and Copyright and IP Regulations Explanatory Memorandum for IP Statute and Regulations Copyright at Monash website	**1.** As copyright is a form of intellectual property which specifies the legal rights attached to an original work, in this project all the datasets, documents and relevant data/information would be protected under the copyright laws. Copyright of the data sets being used in the project could be sold, transferred or licensed for use by other organizations on needs and usage basis.

■ in Australia - Australian copyright applies

As part of this project quite a number of stakeholders are involved hence, the ownership and rights around the data being used as part of research would be required to be determined at the early stages of project planning.

Based on the consensus of all the partners and stakeholders of the project, an Intellectual Property Committee should be setup which would further appointment an Intellectual Property Officer, who would have a defined set of roles and responsibilities agreed by the committee. Officer would have the accountability to make sure that an effective governance plan for this project is place and being adhered to by all the stakeholders.

A process would be required to be setup for addressing any copyright infringements reported for this project.

All the copyright data should have a detailed, clear and informative metadata which articulates the details about the custodian, licensing arrangements and details for the permitted users. Clear mentioned of the datasets which would follow Creative Common license should be made which would provide permission in advance for the range of uses.

2.

As this project is a futurist project which would require quite advanced technologies, the data to be collected, created or compilation of the same is possible both within Australia and outside Australia.

2.1

All the data which is created within Australia is covered and needs to be abided under the Copyright Law of Australia which is defined in the Australian Copyright Act of 1968. This gives ownership to the stakeholders and project partners on the data which is being used in the projects. Each stakeholder would have their data copyrighted separately and could have different license, contract or usage features as part of that. Understanding the sensitivity of data related to this project, key project members or at least project managers should be provided sufficient training on Australian copyright laws in order to prevent any infringement of the same.

2.2

There might be some data which might be created for this project outside the Australian boundaries which could be during some international research conferences, presentations or meetings. Same should be governed under the International Federation of

		Reproduction Rights Organisations (IFRRO) and abide to Universal Copyright Convention (UCC) which was concluded by United Nations Education, Science and Cultural Organisation (UNESCO).
■ outside of Australia	*Consult the Copyright Advisers*	
OWNERSHIP OF COPYRIGHT AND IP		
3. The copyright and other IP in the data is owned by: ■ Monash University ■ The researcher	In general, Monash University owns intellectual property rights, including copyright, in research data originated by all staff including academic staff, except where the data created by academic staff falls within the definition of a 'scholarly work', Monash will also own copyright in data generated by students as part of a 'collaborative research activity.' These terms and defined in Part 5 of the Vice-Chancellors regulations..	**3.** **Monash University :** The basic principle of copyright policy in Monash is that the ownership of the data lies with the individual who creates it but in-case if the work is created for an organization and is created by the employees or specially appointed team, in that case the organization will own the copyright for the work.

- Monash University (joint ownership)
  Research conducted by Monash University in collaboration: copyright and IP ownership are documented in an agreement between the organisations.
  Reference:

- Someone else owns the data

A common example would be a collaboration with researchers from another university.

Provide a reference number or copy of the agreement

In relation to the this project, the ownership of the work would lie with the partners of this project and based on the confidentiality agreements along with the licenses with NSW Data Analytic Centre.

As per the project requirements, Monash would not be able to retain any project related records or classified information. But, as the Data Management Planning team is from Monash, all the work related to creation of Data Management Plan would owned and a copyright artefact of Monash University. All the meetings conducted, artefacts and documents generated during discussions for creation of this Data Management Plan would be an Intellectual Property of Monash University.

The Researcher :

There would 2 types of researchers working in this project.

The first would be the team of Monash which is working on the creation of Data Management plan for the stakeholder. As the team is working on behalf of Monash, the ownership of this data would lie with Monash.

The second would be the researchers actually working in the project as employees for the stakeholders. Any work performed or data generated by these members would be an Intellectual property of partners in this project.

Monash University (Joint Ownership) :

The partners would still need to be informed about the particulars of the artefacts which would be owned by Monash as part of a Confidentiality Agreement. Also, the partners could access that information from Monash after the access agreement contract has been accepted and signed by both Monash and the stakeholders of the project.

Someone else owns the data :

As this project would relate to very sensitive and confidential data which is governed by copyrights and licensees, any other party or individual would not have the legal right to access, use or re-publish this data and information.

☐ Monash University Graduate Research student is creating/collecting the data, but another party owns the copyright and IP.
Reference:

For more information see Ownership of research (Copyright site)
A common example is research funded by a company that wants to retain copyright/IP.
This might also apply to communities (e.g. indigenous groups) that participate in the research and negotiate ownership or co-ownership of the data.
Provide a reference number or copy of the agreement.

☐ Monash University Researcher is using data originating from another party and that party owns the data.
Reference:

Common third parties are government or commercial agencies with existing datasets.
Provide a reference number or copy of the agreement.

THIRD PARTY DATA

Copyright at Monash – Using Third Party Content

4. If someone else owns the data, how did you obtain it and what terms and conditions apply to your use of it?

- Purchased or licensed the data commercially
  Reference:

Provide a reference number or copy of the agreement.

- Obtained data under an open access license
  Reference:

Provide a reference number or copy of the agreement.

4.

Third Part Data :

Third party data is usually subjected to licensing and agreed contracts between the different parties and stakeholders. Below are the considerations which would need to be taken care as part of this project.

■ Obtained data through other means Details:	*Formalise an arrangement with the data owners as soon as possible.* Tips on Seeking Permission from Copyright Owners	**4.1** Purchased or licensed the data commercially Reference Data related to blind spots which would be provided by Telstra would need to be purchased under a strict contract of not reproducing or sharing this information any further in any form. **4.2** Obtained data under an open access license Reference : Data from Accidental Research Centre Monash University, ABS, Trans urban and CSIRO could be gained access under the open access license agreements with these organizations. But us the copyright is normally retained by the owner, there might be ethical requirements which would need to be considered by the project team. Researchers using third party data would be required to get the ethical approval from the original owner of the data, and make sure the issues related to privacy and other requirements imposed are taken care off. **4.3** Obtained data through other means Details : There would be data generated on daily basis which could be part of daily stand-up meetings, technical discussions, design reviews and commercial engagements. As this data is being generated while working for the project, the ownership and IP of the data lies with the organizations and partners of this project.
C. ETHICAL REQUIREMENTS	KEY DOCUMENTS ON THIS TOPIC	EXPLANATORY NOTES SHOULD BE ADDED IN THIS SECTION
5. ■ The research involves human subjects. This includes re-use of data from or about people, e.g. from health agencies.	Research data management guidelines: ethics and consent NHMRC *National Statement on Ethical Conduct in Human Research* [especially Section 3.2: Databanks] *AIATSIS Guidelines for Ethical Research in Indigenous Studies[pdf 97 kb]*	**5.** The data as part of this projects is not related to human subjects or any health agencies. But there are still some data sets like the property locations, blind spots and details related to accidents which should be made processed in such a way that none of the [illegible]cal are identifiable regardless of the technology bein[illegible]ed.
6. ■ A Human Ethics Application has covered / will cover requirements in the following areas:		

■ Privacy
■ Confidentiality
■ Cultural sensitivity
✓ Other
Ethics application ref. no.:

Ethical requirements will impact on how you share and control access to the data.

Provide a reference number or copy of your application.

Information and forms from the Monash Research Office

6.

Ethical requirements apply to this project in consideration to the laws and regulations to protect privacy, confidentiality and other sensitive data for which data de-identifying strategies should be planned and actioned.

6.1

Privacy :

During the project as the research data would be having some datasets which would consists details related to addresses and movement of vehicles, the data privacy restrictions would apply in this project. As the vehicles are considered as entities in this project, all the information and data related to vehicles would be considered as critical information for privacy.

Privacy legislations such as Information Privacy Act and the Australian Commonwealth Privacy Act would be applicable.

A Privacy officer would be required to be appointed by a committee working on privacy related concerns for making sure to maintain an unbiased equilibrium between stakeholders.

6.2

Confidentiality :

The most important requirements for this project is to maintain the confidentiality of the reports and project work which is being performed. We could classify the any information or data which could not be released to public as confidential information.

There would be a requirement to introduce some legal mechanism for controlling the dissemination of data into public. Generally the information would need to kept safe in an encrypted and password protected manner.

As part this project, there would be a need create contract laws for maintaining the confidentiality of sensitive data. Data related to blind spots from Telstra, vehicle movement and traffic management would require non-disclosure agreements to be signed and agreed upon.

6.3

Cultural Sensitivity :

This project does not contain any data or information related to cultural sensitivity. Generally research involving cultural information requires special data requirements for decision

making and processing but in this case it is no applicable as none of the cultural related data set is used in the project.

6.4

Others :

Project team and researchers would need to be provided a training in order to understand and manage potential issues between ethical requirements and other requirements such as retention policies and de-identifying data strategies.

It need to be ensured that privacy is maintained by anonymising data and controlling access to the sensitive-critical data.

7	✓	Other special requirements for managing data have been negotiated in addition to the Human Ethics Application. Details:	Requirements additional to the standard ethics process may apply, e.g. participants in community-based research may specify where data is stored and who can access it. *Provide any available documentation that helps to clarify ethical use of the data.*
7a.	✓	The research involves animal subjects	

7.

This point is not applicable in respect to human ethics. But adding to this the standard ethics process would still apply in which information gathered from different resources should be based on the mutual consent or contract of data sharing.

7.A

The research does not involve any study or information assessment related to animal subjects.

D. DURABLE FORMATS	KEY DOCUMENTS ON THIS TOPIC	EXPLANATORY NOTES SHOULD BE ADDED IN THIS SECTION
8. Data formats used: Details:	**Research data management guidelines: durable data formats** *Consider the durability of all data formats, including digital, print and physical samples.*	**8.** **Data Formats Used :** Containers: These data formats are used for sharing the data sets and information is a packaged form. TAR : Tape Archive GZIP : GNU Zip ZIP : Compressed File ISO : Disk image of optical disk

9.	Data formats meet the following criteria:	Structured Data: Data files present in a structured format from sources like Trans urban, CSIRO are generally in structured format. XML : Extensible Markup language CSV : Comma Separated Value JSON : JavaScript Object Notation Audio: All the meeting recordings and Audio files for information related to noise while vehicle movement is stored, shared in the below format. MP3 : Moving Pictures Expert Group (3) Geospatial: Data related to geospatial locations had been shared in the below formats. GeoTIFF : Georeferenced Tagged Image File Format NetCDF : Network Common Data Form Videos: Video recordings have been done during the project while presentation and demos. Same have saved and preserved in below formats. MPEG : Moving Pictures Expert Group AVI : Audio Video Interleave Still images: Quite a huge number of images have been generated as part of project while assessment of infrastructure based on datasets from SMART infrastructure facility, University of Wollongong. PNG : Portable Network Graphics JPEG : Joint Photographic Experts Group Tabular data: Tabular data has been the backbone of this projects. All the locational data, statistics and exploration data was majorly used and created in tabular format using below formats. CSV : Comma Separated Value Excel : Microsoft Excel

		Text: Textual data was one of the most frequently used and highest volume contributor data format. All the reports and notes were generated in either of these formats. PDF : Portable Document Format WORD : Microsoft Word TXT : Text File PPT : Presentations Email: As the project team size was big, in order to maintain efficient communication at all times, Microsoft Outlook Email were used. The emails are generated in the below data format. PST: Personal Storage Table
✓ Endorsed by an international or national standards agency ✓ Widely used ✓ Accepted as best practice in this discipline Details:	*Choosing standards-based formats assists with long-term access and preservation.*	**9.** The above data formats are inter-operable among diverse platforms and applications along with being fully published and available royalty-free which could be independently accessed and worked upon by multiple software providers on multiple platforms without any intellectual property restrictions for necessary technology. Also, the file formats are developed and maintained by open standard organizations which have a well-defined inclusive process for evolution of the standard improvements. Being non-proprietary, uncompressed and unencrypted make these the ideal characteristics for data archival process also. These have documentations which are publicly available, endorsed by Australia standards and widely used-accepted by the research community.

10. Special hardware or software requirements (for digital data) Details:	*Special hardware and software requirements may have an impact on long-term access and preservation.*	**10.** As all the above data formats are in open source, hence would be available for long term availability for creating, manipulating and accessing research data-information. The level of technical support and time to be spend by researchers while using above data-formats is very less. Currently there is no licenced software which is being used in the project to generate or maintain data files in above data formats, but having said that any specific software and hardware requirements should be clearly documented and be retained for future references.

E. STORAGE AND BACKUP

DIGITAL DATA

KEY DOCUMENTS ON THIS TOPIC

Research data management guidelines: storage and backup

EXPLANATORY NOTES SHOULD BE ADDED IN THIS SECTION

11. Digital data is **stored** in:

- ✓ On Monash University allocated storage (e.g. Network: S drive, Google Drive)
  Location:
- ✓ Monash University managed data storage
  Location:

11.

On Monash University allocated storage :

As the requirement of the project is keep all the data very secure and stored only in NSW data Centre, hence no data would be stored by either the project team or the Data Management Planning team in the Monash University allocated storage drives.

✓ Other
Details:

12. Digital data is **backed up**:

☐ On Monash University managed data storage – automatic nightly backup

☐ On Monash University allocated storage (e.g. Network: S drive, Google Drive)
Frequency of backup:

Monash University managed data storage:

There would still be some amount to data which would be generated by the Monash research team which is working along with project team for creating the data management plan. As Monash has a joint ownership of this data and considering the sensitivity of this data, a customized storage managed by Monash would be required to store the same. This data might contain the information related to research conducted for creation of Data Management Plan.

Others :

One of the important requirements for this project was to store all the data-sets and information related to this project in NSW Data Analytics centre. As part of the DMP research team we would advise 2 data storage architectures to be used and created and by NSW Data Analytics team.

For the sensitive data which is to be not shared openly with anyone, Monash's MeRC storage structure would be desirable.

Since this research involves sensitive data and requires long term retention, MeRC storage offers the best solution. This type of storage will also be used to store the results of the analysis performed during the project phases.

Another set of data which could be used and accessed with wider range of stakeholders under agreed contracts and confidentiality agreements, a structure similar to Monash Figshare could be used. This provides cloud-based solution to access the files locally and remotely for sharing and collaboration.

12.

On Monash University allocated storage :

The data would be backed up as per the Monash automatic nightly backup policy.

Monash University managed data storage:

☐ Other backup arrangements
Details:

No data is stored here.

Others :

As the data is sensitive, it needs to be backed up daily. Ideally, it is backed up at the location where it is stored along with another location which serves as the disaster recovery location.

NSW Data Analytics centre would be required to maintain the data archival in more than 2 archival repositories.

NON-DIGITAL DATA

13. Data in non-digital formats is/will be stored in:

☐ Secure facilities located in the school, institute, or centre
Location:

✓ Other
Location:

13.

Non-digital data may comprise of documents like charts, diagrams, drawings, workshop documents and minutes of meeting etc. As it is efficient and cheaper to store the digitized documents as compared to non-digital document, a process for conversion of non-digital documents into digitized format would be setup.

As part of that process all the non-digital documents would be converted into digital documents and stored along with other digital document in respective date centres. Based on the source of the documents, they could be stored in either Monash managed customized storage of NSW Data Analytics data centres.

DATA VOLUMES (ALL FORMATS)

14.	Estimated data storage size: Details:	*Estimate data volumes and discuss these with your storage providers well in advance.*	**14.** As this project involves vehicle movement for vehicles which is classified into 6 different categories, they are generate huge amount of streaming data. Both access and production of huge amounts of data in real time is required. For this the storage capacity in NSW Data Centre should be sufficient. We could plan to setup 2 different storage capacities in NSW Data centre: **Active Storage : 500 TBs (Tera Bytes)** This would be required to store all the live streaming which is being generated by the vehicles. Maximum latest past 7 days data can be stored in the storage. The data once generated for 7 + 1 day would be archived to passive storage. **Passive Storage : 100 PBs (Peta Bytes)** This storage would be required to store all the archived data and reference data which is required for successfully functioned the automated vehicle system. 

F. SHARING DATA AND CONTROLLING ACCESS

This section relates to data sharing during the course of the project.

Longer-term data sharing is covered in the section below on deposit in a repository or archive.

SHARING DATA DURING THE PROJECT

15. During the project, research data will be **shared**

- ✓ Not at all – I am the only person that will have access
- ✓ Internally – Monash University staff (e.g. colleague or supervisor) and students
- ✓ Externally – (e.g. staff at another institution, research participants, funding agency)
  Details:

16. Data will be shared with external parties in the following ways:

- ☐ Data stored at Monash University will be accessed by others, e.g. using an online shared workspace
  Details:
- ✓ Data will be **transferred** to the other location/s

KEY DOCUMENTS ON THIS TOPIC

Research data management guidelines: sharing and disseminating data

Only share data if you are sure about the implications of doing so, in terms of copyright and IP ownership, and ethical requirements.

This includes access to Monash University systems for storing and managing digital data that are provided by the Monash e-Research Centre.

EXPLANATORY NOTES SHOULD BE ADDED IN THIS SECTION

15.
Not at all :

This data would be shared amongst both internal and external stakeholders.

Internally :
All the data and information which needs to be shared internally within the project team should be strictly done under the consideration and governance of privacy officer along with Intellectual property officer.

Externally :
Data being shared with external stakeholders need to governed and acknowledge by both the parties under the legal agreements and contracts.

As per the Code for Responsible Conduct of Research in Australia, researchers should be sharing all the research data if possible. The same is also encouraged by Australian Research Council (ARC) Discovery Projects.
As per this project, very limited access to datasets and final report is shared.

16.
Project involves highly sensitive commercial data that is licensed to the NSW Data Analytics centre. This data cannot be held or accessed by any other party including Monash University until

☐ using data transfer tools
☐ using secure data transfer tools
☐ on physical media, e.g. hard disk drives, tapes, etc.
✓ by other means
Details:

CONTROLLING ACCESS

17. Access to the data during the project will be

☐ Unrestricted

✓ Restricted

accepted, acknowledged and approved by partners along with NSW data Centre.

However, several other data sets will be retrieved and held by Monash University. Those can be accessed by few selective forum based on the ethical consent and agreement.

Data would be transferred and accessed only electronically using the encrypted primary and secondary keys.

Audiences for this work may be available outside of the research project also, but the same would not be shared with them.

17.

The access to this information is restricted and could be accessed only by authorised team members by VPN access using multi-factor authentication system.

Strict guidelines would be issued and need to be adhered to for generating the passwords.

In relation to physical storage, the data centres would be having dedicated security team for protecting all the data centres round the clock.

✓ password-protected (digital data only)
✓ encrypted (digital data only)
✓ physical security, e.g. locked filing cabinets, check-out procedures
Details:

Both Monash Data Centre and NSW Data Centre would be required to access specialised infrastructure and services that have been independently assessed and accredited to the ISO 27000 standards.

G. DOCUMENTATION AND METADATA

18. The following documentation and/or metadata (information about the data) will ensure data can be retrieved and used:

✓ Inventory of data assets
Details:

KEY DOCUMENTS ON THIS TOPIC

Research data management guidelines: organising data

EXPLANATORY NOTES SHOULD BE ADDED IN THIS SECTION

18.

Inventory of data assets :
The Australian Code for Responsible Conduct of Research requires all researchers to maintain a list of their research data assets.

- ✓ Metadata standards
  Details:
- ✓ Protocols for identifiers / reference numbers
  Details:
- ✓ Protocols for folder & file naming
  Details:
- ✓ Protocols for document structures – e.g. column headings, document properties etc.
  Details:

Some metadata can be stored internal to the data object that is being described, while some documentation and metadata would usually be stored externally. As part of this project we would be creating a well-structured and detailed documentation which contains details in relation of all the data-sets and its versioning. This would be stored in the NSW Data Centre along with other reference data in the passive storage.
In order to attain the maximum information about the datasets, UK Data Audit framework methodology would be used. This would require to have the information clearly articulated as name,

- ✓ Data dictionaries, data definition files and schema Details:

description, reference, criticality of the data set along with any other general comments.

Metadata standards :
Understanding the different types of the data sets on which the project team would be working the below Meta Data Standards have been advised to be used.

Dublin Core - All the generic data sets and information like the videos, surveys, minutes of meetings. This would be collected as part of the project progression.
Content Standard for Digital Geospatial Metadata (CSDGM) - For the geospatial locations. This would be collected from CSIRO, TransUrban and Telstra.
Core Scientific Meta-Data Model (CSMD) - All the scientific calculated and collected data sets. Data-sets from Transport for Victoria and Accident Research centre.

Protocols for identifiers / reference numbers :
The identifier is unique or at least unique within the particular system and will not change over time, therefor once we have the identifier number, we can retrieve the data.
Both primary keys and handles would be used controlled and organized vocabulary.

Protocols for folder & file naming :
File naming conventions should be finalized early in a research project, and agreed upon by the research team members before data is created.
Generic recommendations which generating folder structure on MeRC and Figshare storage clusters.

Protocols for document structures :
Project team members would be required to ensure that all the digital files are well-structured internally. General information like document name, authors, contact details, creation date, version information along with the heading of column for tables and spreadsheets greatly increase the ability of the data to be managed and interpreted over time.

Data dictionaries, data definition files and schema :
Rules relating to how data is entered, and how the information should shored internally and externally. As the documents would be

- ☐ Other documentation or metadata requirements (including software that may be used to managed documentation and metadata)
  Details:

retained for long-term, and it will provide valuable context to the data over time.

H. RETENTION AND DISPOSAL

MINIMUM PERIODS

As a researcher, you should identify the likely retention period for your data as early as possible in the research and ensure that requirements for retention and disposal are met.

19. The University's Retention and Disposal Authority (RDA) provides four main retention provisions for research data:

- ✓ Data and datasets created as part of research activities within the institution. This does NOT include data created for specific research activities for which additional regulatory requirements apply, including: clinical trials, gene therapy and research involving children.
  - ☐ Disposal allowed 5 years after completion of research activity
- ☐ Data and datasets created as part of research activities within the institution which involve minors.
  - ☐ Disposal allowed 15 years after child reaches the age of 18
- ✓ Data and datasets created from clinical trials as part of research activities within the institution.
  - ☐ Disposal allowed 15 years after completion of research activity
- ✓ Data and datasets created as part of research activities within the institution, which are of regulatory or community significance.
  - ☐ Permanent Archive

KEY DOCUMENTS ON THIS TOPIC

Research data management guidelines: retention

EXPLANATORY NOTES SHOULD BE ADDED IN THIS SECTION

19.
A record management team would need to be created in order to maintain and ensure the retention and disposal data generated and used as part of the project. The Records Management team would be required to perform systematic and controlled destruction of data.

Disposal after 5 years:
All the data created and captured during the planning and development data management plan would be retained for 5 years. Even accident and blind spot related would also have the same policy.

Disposal after 15 years:
All the data related to special datasets collected from TransUrban and CSIRO along with the transport related data from transportation of Victoria would be retained for 15 years.

Permanent Archive :
Decisions related to retention of selected datasets permanently as part of archival needs to be made at an early stage of the research project.

All the original design documents, along with requirements and specifications for the project to be retained indefinitely.
Technology-based decisions relating to storage media, software, and digital file formats might impact upon the length of time that data can be easily retrieved and used.

LONG-TERM AND PERMANENT RETENTION

Can be completed by the researcher in draft form at the start of the project and should be reviewed regularly as research progresses.

20. Data should be considered for permanent retention / archiving because the research

☐ is controversial
✓ is of wide public interest
✓ uses an innovative technique for the first time
☐ shifts the paradigm in this field of inquiry
✓ would be costly or impossible to reproduce
✓ will be of enduring value to researchers in this discipline
☐ will be of enduring value to researchers in other disciplines
✓ supports a patent application or other formal IP process
☐ Other
Details:

20.
Data should be considered for permanent archiving :
The research is of wide public interest. The project is an opportunity to improve the safety, performance, efficiency and environmental impact of Australia's transport system which can be consider as wide public interest.

It is uses an innovative technique for the first time in Australian. Because the Australian government also has commenced several initiatives to develop and deploy these new transport technologies. Which means those technologies are new for them.

The project data were collected in the form of workshop, semi-structured interview and analysis of related documentation and data provided by the participants. As number of the data sets are highly sensitive due to their commercial nature. Which means the research would be costly to reproduce , will be of enduring value to researchers in this discipline and also will be of enduring value to researchers in other disciplines which is similar to this.
Therefor data should be considered for permanent archiving.

21.	The ability to permanently retain / archive the data and make it accessible will be affected by: ☐ legal issues around ownership of copyright and IP ☐ ethical requirements for data to be destroyed ☐ legal or ethical requirements for access to data to be restricted ☐ technical issues, e.g. obsolete data formats or software ✓ other Details:	*Consider these issues during data management planning so that a strategy can be developed for resolving any conflicts.*	**21.** **Others :** As part of this project, a dedicated data management Team had been setup and engaged to plan for a long term data management plan. Hence considering the current state and plan as part of DMP, there should not be any issues in the ability to retain some amount to data permanently.

I. DEPOSIT IN A REPOSITORY OR ARCHIVE	KEY DOCUMENTS ON THIS TOPIC	EXPLANATORY NOTES SHOULD BE ADDED IN THIS SECTION
22. ✓ The data will be deposited in a repository or archive. ✓ Monash University Data Repository ✓ another repository or data archive Details:	**Research data management guidelines: sharing and disseminating data** *Check the deposit requirements of your scholarship and funding agreements. Some agencies e.g. ARC, NHMRC, may require deposit in a repository or archive as a condition of receiving funding. In some cases secure storage on University systems may satisfy the requirements.* *Always read and understand the terms and conditions of deposit.*	**22.** There would be both possibilities depending on the type of data and information being stored or planned to be archived. As suggested in the previous sections, all the research data collected and generated by Monash research team for generating the data management plan would be stored in the Monash University Data Repository. Few other data sets which have the changing real time information would be retained on NSW data centre for 15 years. All the technical and design documents along with metadata documents would be archived for permanent retention.
23. ✓ Data cannot be deposited in Monash University Data Repository or other online repository or archive. ☐ Data and documentation stored on Monash allocated data storage will be associated with the academic unit. ☐ Data and documentation stored on Monash Network Drives will be associated with the academic unit. ☐ A copy of local data (e.g. from a personal laptop or home computer or on removable media) will be provided to the academic unit. Details: ✓ Data will be securely destroyed. Details:	*Make other arrangements for long-term storage of the data* Monash managed data storage: merc@monash.edu **Research data management guidelines: security**	**23.** All the data which would be disposed after 5 or 15 years duration would be destroyed. Destruction process must be irreversible, meaning that there is no reasonable risk that any information may be recovered later. Extra care would need to be taken care of when dealing with these records as they may contain sensitive information. General Retention & Disposal Authority for Records of Common Administrative Functions would be taken into consideration while destroying the records.
24. ✓ CHECK HERE TO NOTE THAT YOUR TEAM HAS CROSSCHECKED THE DMP AND SIGNED OFF YOUR CONTRIBUTIONS BEFORE SUBMITTING	*Provide any available documentation that helps to clarify arrangements that have been made for long-term management of the data.*	

ADDITIONAL INFORMATION CAN BE ADDED HERE

PLEASE LINK THIS TO THE RELEVANT SECTION (A – I) OF THE DMP TEMPLATE.