developing a contemporary public bus network

A Data Management Plan created using DMPonline.be

Creator: Maarten Wens

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Maarten Wens

Grant number / URL: 1SH2L24N

ID: 202142

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Over the years, several service types for public transport by road have been proposed, each promising an improvement over the current situation. These services range from traditional, fixed bus lines to a fully flexible door-to-door Uber service. However, multiple case studies report that the implementation of these services is not straightforward. The identified research gap between theory and practice is that these services are studied separately and not as part of a bigger transportation system. When combined with other transport options, some services might become too expensive for their added value. Therefore, the main goal of this research is to find a methodology for integrating several systems optimally without compromising the level of detail of each service.

Last modified: 11-03-2024

developing a contemporary public bus network
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

Two types of data will be used:
1) artificial data to test the developed algorithms. These datasets consists of either benchmark instances (.txt or sometimes even custom formats) used by other researchers or created by ourselves.
2) Data form public bus operators: .csv, .txt and .shp (or other file types for geographical data) files with travel times, current lay-out of the transit network, travel demand (anonymous), geographical data, and general data about population density.
Except from this data, software code will be produced (.py files) to solve the problem of this research (<1GB).
No additional data storage will be necessary.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Designation of responsible person: Pieter Vansteenwegen
2. Storage capacity/repository
    - during the research: Data is kept on the research computer of the researcher, a back up is kept at the local drives of the research group, as well as the KU Leuven onedrive.
    - after the research: The relevant data will be put in the data management tool of KU Leuven, RDR. A back up will be kept at the local drives of the research group, as well as the site of the department (e.g. https://www.mech.kuleuven.be/en/cib/lp).

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

/

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

All data regarding travel patterns of people is aggregated and therefore completely anonymous. This should minimise any ethical issue.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

/

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
benchmark instances	to compare with other researchers	reuse data	digital	simulation data	.txt, NA	<1GB	
operator data	Third-party data on the daily operations of bus lines, such as number of lines, current routes	reuse data	digital	aggregated data, simulation data, observational data,...	.csv, .txt, .gtfs,...	<100GB	
demand data	data on the current and future estimated amount of passengers of busses and other modes of transportation	reuse data	digital	simulation data, observational data	.csv	<100GB	
Research codes	Research codes developed within the research unit (stored in ITScrealab on Gitlab) or found online as open-source code	Generate new data + reuse existing data	Digital	Software	code in .py	<1GB	
OSM	OpenStreetMap data; open data source for maps and geographical features of the land	reuse existing data	Digital	Compiled/aggregated data	online database in .osm, can be extracted and stored in tabular format in .csv	< 100 GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Example benchmark instances: https://users.cs.cf.ac.uk/C.L.Mumford/Research%20Topics/UTRP/Outline.html

Operator's/demand data: contact Pieter Vansteenwegen
OSM: https://www.openstreetmap.org/about

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Datasets will be aggregated in such a way that no individual personal data is available to the researcher or can be constructed. aggregated datasets: travel demand and travel patterns.

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

The research codes and the model (the simulation of developed scenarios) might have the potential for commercial valorization. We work with open source licenses with the restriction that everyone who uses our data or tools/code must also make their tools/code or data available open source as well. For commercial use or other use types that are excluded from the open-source license, the license needs to be negotiated with KUL/LRD.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

The data provided by the railway companies is confidential and cannot be shared with other researchers without permission from the data owners. Still, specific (benchmark) instances might be made available for other researchers to verify our work and to build further on or improve our work.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Re-used data will either be referred to the original source or a relevant copy will be uploaded to the KU Leuven data management tool, which has required meta-data documentation fields. Relevant code (including README.md files) will be stored in git-lab and will be documented

according to documentation standards and the available time of the researcher to improve re-usability.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

RDR, the data management tool of KU Leuven has such a standard. For code, standard practice concerning readability and documentation will be used.

3. Data storage & back-up during the research project

Where will the data be stored?

The data will be stored on the research computer of the researcher, with a back up at the internal storage drives of the research group. Cloud services, such as KU Leuven onedrive will be used as well. Gitlab is used for code.

How will the data be backed up?

The internal drive of the research group provided by the department is used as a backup, this drive has an auto backup. For code, gitlab works with version control. So if necesarry, a rollback to a previous version of the code can be made as well.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

KUL OneDrive and Gitlab quota are sufficient for the entirety of the project.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Security checks are in line with KU Leuven regulations (e.g. dual factor authentication for accessing all platforms where data is shared, such as gitlab, teams, onedrive,...).
Some platforms, such as gitlab, furthermore use leveled access. This makes it impossible for non-authorised to access these files.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

No additional cost are expected

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

KU Leuven research data management policy stipulates that all relevant data generated are retained for a period of minimally 10 years after the

end of the project. This will be followed.
The GitLab repository of the group including all the developed codes stays intact irrespective of people finishing their research and moving on to other projects. All data related to publications will necessarily be retained along with other important unpublished data.
The possibility of publication of open-source tools as formal publications with permanent public identifier is currently being examined for other packages in the group and will become our standard for open-science publishing.

Where will these data be archived (stored and curated for the long-term)?

After the research is concluded, the data will be kept at RDR as well as the internal storage drive of the research group and the gitlab.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The stored data is assumed to be small enough to not incur additional costs.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

If allowed by the third parties providing this data, the data will be transformed to become a new benchmark instance for other researchers to try and find improved solutions for the same problem.
The end result that will be public will include: code which uses geographical data,(adapted) travel demand data, and input about the allowed number off resources to provide a contemporary public transit network. Except for the code and input data, example instances and solutions will be provided as well.

If access is restricted, please specify who will be able to access the data and under what conditions.

Access to internal information is restricted to the research group only.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Privacy aspects

3th party must agree that (part of) their data is used and shared in an academic context.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

A github/gitlab repository will be used, but is not made yet.
The department of mechanical engeneering furthermore has sites with publishes and stores research data:

- https://www.mech.kuleuven.be/en/cib/lp
- https://www.mech.kuleuven.be/en/cib/op

When will the data be made available?

Along with each publication, relevant code and data will be made public. Under the code of open science, all relevant data that can be shared will be shared, to increase verifiability and reproducibility of the methods and their conclusions.

Which data usage licenses are you going to provide? If none, please explain why.

Datasets will be published under GNU General Public License version 3 (https://www.gnu.org/licenses/gpl-3.0.en.html) which is an open source software license. It has the restriction that everyone who uses our data or tools must also make their tools or data available open source as well. For commercial use or other use types that are excluded from the open-source license, license needs to be negotiated with KUL/LRD.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

What are the expected costs for data sharing? How will these costs be covered?

No additonal costs are expected.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Maarten Wens and Pieter Vansteenwegen

Who will manage data storage and backup during the research project?

Maarten Wens and Pieter Vansteenwegen

Who will manage data preservation and sharing?

Pieter Vansteenwegen

Who will update and implement this DMP?

Maarten Wens

developing a contemporary public bus network
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

developing a contemporary public bus network
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable