Integrative Analysis of Microscopic Spatio-Temporal Traffic State, Driving Styles, and Accidents: Towards a Comprehensive Understanding of Mutual Effect of Traffic Operation and Road Safety in Complex Motorway Corridors

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.

The dataset was created during my PhD and is continuously updated, it includes:

1.1- Traffic Data: 226 sensors across 64 stations (2019–2023), including passage time, speed, and vehicle length for each passing vehicle. Stored in monthly compressed CSVs, averaging 1.5–2.0GB per file.

1.2- Continuous Vehicle Trajectories: 1Hz data from Be-Mobile commercial traffic service provider (2019–2021), including X-Y coordinates, time, and vehicle ID. Hourly compressed CSVs range from 200KB to 12MB for each file, depending on the time of day.

1.3- CCTV Footage: Segment of R1 ring of Antwerp (2022), 8 days, 13 hours daily with 9–10 camera views daily.

1.4- Data of Variable Signs: Data from 35 gantries each has a sign per lane and 3 variable warning message signs (2019–2023).

1.5- Crash Data: All crash records across the Flanders from the traffic control center (2019–2023).

1.6- Publications and presentations. These include academic papers and presentations (for internal use in the research group or for conferences and seminars) produced by the researchers (Mohammad Ali Arman). Academic papers will be published in .pdf format, and presentations will be made in .ppt format and stored in .pdf on OneDrive.

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:

The part of the dataset that was created during my PhD was stored on a local external hard drive and delivered to my PhD supervisor, who is now my postdoctoral supervisor. Later, the dataset except for item (1.3) of the list above, which includes huge files transferred to the ManGO Active Data Management Platform of KUL ICTS, and huge files of item (1.3) are stored on large-volume storage provided by KUL ICTS.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

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.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or 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 original owner of the data is the Flemish government. It is applicable to data items (1.1) and (1.5). The data can be used only in KU Leuven with the current license agreement. Moreover, the Flemish government provides free licenses for non-commercial use upon request.

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

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).

I created a "read me" PDF file and stored it along with the data. The file explains each header and its unit.
Given that I developed the dataset in my PhD, and the dataset is related to the local traffic flow in Flanders, because different parties collected data, if any inconsistency exists, for example, regarding time and due to daylight saving, either they unified, or they mentioned it in the "read me" file.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

No, at the time, the data was only stored, but my supervisor, Professor Chris Tampère, is consulting with KUL ICTS to make data querying and metadata available via the ManGO Active Data Management Platform.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Large Volume Storage
- ManGO
- OneDrive (KU Leuven)

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

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

The data resides on KUL servers that are hosted, backed up, and secured according to KUL's professional standards. Prof. Tampère, my supervisor, takes formal responsibility, relying on the professional services of the KUL ICTS data center (responsible antenna: Rudy Rys). The ManGO Active Data Management Platform that KUL ICTS launched in March 2023 will be used in my mandate for all the above-mentioned datatypes except for 1.3. It supports secure storage, metadata management, automation of workflows, and sharing of data among my supervisor, me, and other researchers in our group based on the iRODS open-source software. In addition, for more passive (longer-term archive) storage of the data type mentioned in 1.3, which has larger data volumes requiring low I/O, KUL ICTS provides large-volume storage in its data center at reasonable rates.

The responsible antenna ensured data query or access would be possible only for authorized persons through my supervisor.

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

There are no expected costs for data storage and backup.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

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

- Large Volume Storage (longterm for large volumes)

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

There are no expected costs for data storage and backup.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?

Please explain per dataset or data type which data will be made available.

- Yes, as restricted data (upon approval, or institutional access only)

I plan to publish it to foster scientific collaboration, but the Flemish government is the original data owner, and it needs further negotiation, although they are very collaborative and provide free licenses for non-commercial use.

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

The Flemish government is the original data owner, and providing access for researchers outside KU Leuven requires further negotiation. However, they are very collaborative and always willing to provide free licenses for non-commercial use.

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 per dataset or data type where appropriate.

- Yes, other

Generally, the answer is no, there are no factors that restrict or prevent the sharing of (some of) the data, but if the original data owner, the Flemish government, grants us the required permission, the dataset does not include any sensitive, dowl use, or human-related components that make sharing problematic or impossible. Still, it is conditional on a more general agreement with the Flemish government.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)
- Other (specify below)

It depends on further agreement and consulting with the original data owner, but with my supervisor, we agreed to publish a small sample of data on my personal webpage to increase my scientific visibility.

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)
- Data Transfer Agreement (restricted data)

As previously explained, all data-sharing plans are conditional and dependent on negotiation with the original data owner.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- No

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

There are no expected costs for data sharing.

Responsibilities

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

The documentation is my (Mohammad Ali Arman) responsibility. Metadata will be managed by my supervisor, Professor Chris Tampère, via the KUL ICTS data center (responsible: Rudy Rys).

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

I will manage the data storage and backup during the research project in collaboration with my supervisor, Professor Chris Tampère, via the KUL ICTS data center (responsible: Rudy Rys).

Who will manage data preservation and sharing?

After negotiating with the original data owner, which will be conducted by my supervisor, Professor Chris Tampère, conditional on the agreement and its articles, I will manage the data preservation and sharing in collaboration with Professor Chris Tampère via the KUL ICTS data center (responsible: Rudy Rys). Moreover, with the same preconditions, I will present and share a small sample of data via my personal website.

Who will update and implement this DMP?

I, Mohammad Ali Arman, the researcher of this postdoctoral mandate, will update and implement the data management plan with the collaboration and supervision of Professor Chris Tampère.