



Deliverable D5.1 Data Management Plan

Project acronym:	MOVINGRAIL
Starting date:	01/12/2018
Duration (in months):	25
Call (part) identifier:	H2020-S2R-OC-IP2-2018
Grant agreement no:	826347
Due date of deliverable:	Month 4 (2nd issue Month 12)
Actual submission date:	15/01/2020 (1 st issue 02/04/2019)
Responsible/Author:	Rob Goverde (TUD)
Dissemination level:	PU
Status:	Issued

Reviewed: yes

Document history		
<i>Revision</i>	<i>Date</i>	<i>Description</i>
0.1	06/03/2019	First draft
0.2	29/03/2019	Second draft for review by partners
1.0	02/04/2019	First issue
2.0	15/01/2020	Second issue, extended data summary
2.1	09/03/2020	Update WP2

Report contributors		
<i>Name</i>	<i>Beneficiary Short Name</i>	<i>Details of contribution</i>
Rob Goverde	TUD	First draft, inserting comments partners, second issue, updates
Leonhard Pelster	TUBS	Input data WP1
Martin Scheidt	TUBS	Input data WP1
Gemma Nicholson	UoB	Input data WP2
Bill Redfern	PARK	Input data WP3
Egidio Quaglietta	TUD	Input data WP4

Funding

This project has received funding from the Shift2Rail Joint Undertaking (JU) under grant agreement No 826347. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Shift2Rail JU members other than the Union.

Disclaimer

The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The content of this document reflects only the author's view – the Joint Undertaking is not responsible for any use that may be made of the information it contains. The users use the information at their sole risk and liability.

Table of Contents

1.	Introduction	4
1.1	Project summary	4
1.2	Purpose of this document.....	4
1.3	Context.....	4
2.	Data Summary	5
3.	FAIR data.....	8
3.1	Making data findable, including provisions for metadata.....	8
3.2	Making data openly accessible	8
3.3	Making data interoperable	8
3.4	Increase data re-use (through clarifying licences).....	8
4.	Allocation of resources	9
5.	Data security.....	9
6.	Ethical aspects	9
7.	Other.....	10

1. Introduction

1.1 Project summary

MOVINGRAIL ('MOving block and VIRTUAL coupling New Generations of RAIL signalling') is a Shift2Rail project addressing the topic 'Analysis for Moving Block and implementation of Virtual Coupling concept'. The aims of MOVINGRAIL are

- To identify and assess the most suitable methodology in order to test and bring into service Moving or Fixed Virtual Block contributing to the definition of the Operational Procedures and highlighting the differences with the traditional signalling systems.
- To analyse the potential business and market response thanks to the application of the Virtual Coupling concept identifying pros/cons in terms of performance and cost, and to assess the needs and work done for the Train-to-Train (T2T) both in IP1 and IP2 and propose convergence of technical communication solution(s).

1.2 Purpose of this document

This document has been prepared to provide the Data Management Plan (DMP) which addresses the way research data is managed in the MOVINGRAIL project within the Open Research Data Pilot (ORD Pilot). The ORD pilot aims to improve and maximise access and re-use of research data generated by Horizon 2020 projects, considering the need to balance openness and protection of sensitive information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, as well as data management and preservation of questions.

DMPs are a key element for good data management, as they describe the management of the data to be collected, processed and published during a research project, creating awareness about research data management topics such data storage, backup, data access, data sharing, archiving and licensing. MOVINGRAIL hereby states the adherence to the FAIR data principles, whereby research data is made Findable, Accessible, Interoperable and Re-usable for the community, responsibly considering possible data restrictions on public sharing.

It is acknowledged that a DMP is a living document and, therefore, as the implementation of the project progresses and significant changes occur, this plan is updated accordingly on a finer level of granularity at the end of each project period (M12 and M24).

1.3 Context

The present document constitutes the Deliverable D5.1 "Data Management Plan" in the framework of the TD2.3 of IP2 (Moving Block) task 2.3.1 (Moving Block Operational and Engineering Rules) and task 2.3.6 (Test Specifications), as well as the TD2.8 of IP2 (Virtual Coupling) task 2.8.3 (Feasibility Analysis) and task 2.8.6 (Impact Analysis).

2. Data Summary

MOVINGRAIL collects various kinds of data:

1. Semantic data
2. Stated preference data from surveys and workshops
3. Simulation data.

The responsibility to define and describe all non-generic data sets specific to an individual work package is with the WP leaders. The WP leaders formally review and update the data sets related to their WP. All modifications/additions to the data sets are provided to the MOVINGRAIL Coordinator (TUD) for inclusion in the DMP.

The table below shows the various data collected with the purpose of the data collection and its relation to the objective of the project.

Work Package	Data
WP 1 (TUBS)	Semantic data of railway signalling
Purpose	The data supports the operations analysis of train centric signalling
Types and format	Excel and PDF
Reuse of existing data	Semantic data from X2RAIL-1
Origin of data	X2RAIL-1 and own work
Expected size of data	Less than 100 MB
Data utility	Useful for anyone working on railway signalling engineering and operations
WP 1 (TUBS)	Glossary
Purpose	The data supports the operations analysis and terminology for describing various scenarios.
Types and format	mysql, php, flatfile, pdf, epub, html
Reuse of existing data	Various literature as specified in references
Origin of data	Literature and own work
Expected size of data	Less than 100 MB
Data utility	Useful for anyone working on railway signalling engineering and operations, accessible via https://glossary.ivev.bau.tu-bs.de/tiki-index.php and www.movingrail.eu under a Creative Commons Attribution 4.0 International License (CC BY 4.0).
WP 1 (TUBS)	Symbol library
Purpose	This TikZ library is a toolbox of symbols geared primarily towards creating track schematic for either research or educational purposes. It provides a TikZ frontend to some of the symbols which may be needed to describe situations and layouts in railway operation.
Types and format	TeX, TikZ, pdf, png
Reuse of existing data	-
Origin of data	own work
Expected size of data	Less than 50 MB
Data utility	Useful for anyone working on railway signalling engineering and operations, accessible via CTAN (Comprehensive TEX Archive Network) under an ISC license at https://ctan.org/pkg/tikz-trackschematic

WP 2 (UoB)	Stakeholders requirements data
Purpose	The data supports the identification of gaps in ETCS Level 3 testing, current issues and requirements needed for an effective system testing, validation and certification.
Types and format	PDF questionnaires and PDF survey results
Reuse of existing data	-
Origin of data	The data derives from questionnaires made originally on paper in a workshop and then aggregated and anonymized.
Expected size of data	Less than 100 MB
Data utility	The data can be used for developing operational concepts and testing strategies for the verification and validation of moving block signalling systems that draws on best practice and meets all stakeholder requirements. It is available at https://beardatashare.bham.ac.uk/getlink/fiNYac39GLAxPfs7s5WWRvi9/
WP 3 (PARK)	Stakeholders requirements data
Purpose	The data will establish and refine the communications requirements for Virtual Coupling and the Performance of communications architectures and equipment's including developments relating to autonomously driven cars.
Types and format	It is expected that the primary new data used in WP3 will be in the form of textual requirements from stakeholders via questionnaires and workshops, anonymized in accordance with GDPR.
Reuse of existing data	It is expected that data will be received and shared from the complementary projects (CONNECTA-2, X2RAIL-3) which will be in accordance with the Collaboration Agreements. In addition, we will re-use data from the public domain, and other data made available, from ASTRail, CONNECTA-1, ETALON, IN2RAIL, MISTRAL, Roll2Rail, Safe4Rail-1, Safe4Rail-2, X2RAIL-1-WP3, and X2RAIL-2-WP3/4/5 projects are also expected to be of use to MOVINGRAIL WP3.
Origin of data	Original research, industry, preceding and collaborating Shift2Rail projects.
Expected size of data	Less than 1 GB
Data utility	The data will be useful to the signalling industry to identify virtual coupling technical communication requirements and solutions; review previous studies and projects into virtual coupling; analyse solutions against requirements for virtual coupling; investigate the application, solutions and dynamics of automated car driving; and evaluate the applicability of autonomous vehicles to the railway field.
WP 3 (PARK)	Requirements data
Purpose	The data will establish and refine the communications requirements for Virtual Coupling and the Performance of communications architectures and equipment's including developments relating to autonomously driven cars.
Types and format	Statistical performance data on communications systems.
Reuse of existing data	It is expected that performance data will be subject to commercial

	confidentiality.
Origin of data	Original research, industry, preceding and collaborating Shift2Rail projects.
Expected size of data	Less than 100 MB
Data utility	The data will be useful to the signalling industry to identify virtual coupling technical communication requirements and solutions.
WP 4 (TUD)	Stated preference data from surveys and workshops
Purpose	The data supports the assessment of market potentials and impact assessment of Virtual Coupling for different railway segments
Types and format	Surveys from railway experts to gather feedback and opinions about actual technological and operational feasibility of Virtual Coupling
Reuse of existing data	Part of the information about operational scenarios from X2RAIL-3 WP6 & 7 will be reused to make surveys to railway experts.
Origin of data	The data will derive from surveys made originally on paper and then electronically transferred to an Access database
Expected size of data	Less than 1 GB
Data utility	The data produced in WP4 will be useful to railway industry stakeholders, academic researchers to assess feasibility and multi-dimensional impacts of Virtual Coupling as well as to make predictions/plans about development and implementation plans for such a technology. Furthermore it is useful to other experts of the broader transport industry and statisticians to estimate environmental repercussions that Virtual Coupling could have by potentially attracting more passengers towards the railways.
WP 4 (TUD)	Simulation data
Purpose	Investigate applicability and impacts on safety, costs, and performance of Virtual Coupling
Types and format	Simulation data will have different formats, specifically .xslm (Excel), .csv files, RailML, InfraAtlas and plain text files
Reuse of existing data	Input data from railway traffic simulation models already built during other national and international projects (e.g. ON-TIME) are expected to be re-used.
Origin of data	Input and output of simulation models and multi-criteria analyses.
Expected size of data	Less than 10 GB
Data utility	The data produced in WP4 will be useful to railway industry stakeholders, academic researchers to assess feasibility and multi-dimensional impacts of Virtual Coupling as well as to make predictions/plans about development and implementation plans for such a technology. Furthermore it is useful to other experts of the broader transport industry and statisticians to estimate environmental repercussions that Virtual Coupling could have by potentially attracting more passengers towards the railways.

3. FAIR data

3.1 Making data findable, including provisions for metadata

The data will be securely stored at 4TU.Centre for Research, which is a Trusted Digital Repository for technical-scientific research data in the Netherlands that complies fully with H2020 requirements of making data findable, accessible, interoperable and reusable (FAIR). See

<https://researchdata.4tu.nl/en/home/>

Data collections, processed data and data representations will be stored for 15 years after the end of the project. Research data that is not privacy sensitive will be open access available through the data repository mentioned above as far as this is compatible with and does not infringe IP requirements of the partners. These data, including the metadata that ensures that others can find and use the data, will be stored and made available in the TU Delft data archive 4TU.ResearchData.

The Dublin Core (www.dublincore.org) metadata standard will be adopted and further information on the data will be delivered in file headers or in separate documentation files. This also applies to documentation files (e.g. reports) and other types of data including experimental data and input/output tabular data for code training, testing and validation.

3.2 Making data openly accessible

Once scientific journal publications are published (in Open Access), publishable data (according to the Consortium Agreement) will be publicly archived for the long term via the 4TU.Centre for Research Data archive (documentation, experimental data and tabular data), following their metadata standards (Dublin Core). TUD researchers can upload up to 1 TB of data per year free of charge. This should suffice for the data that will be archived for the long term. The 4TU.Centre for Research Data Archive ensures data will be well-preserved and findable in the long term (each uploaded dataset is given a unique persistent digital identifier).

Open and standard formats will be preferred for archived data files (e.g., .csv, .txt). Proper documentation files will be delivered together with the datasets in order to facilitate reuse of data.

3.3 Making data interoperable

All publishable data will be delivered in open and standard data formats. Discipline specific metadata is currently under discussion. If applicable, metadata will be delivered in XML format together with the data (depending on the chosen format). Proper documentation (README) files will be delivered accordingly. Tabular data will be archived with informative and explanatory headers to facilitate data re-use and interoperability.

3.4 Increase data re-use (through clarifying licences)

All data that cannot be disclosed will be kept at the respective institutional server for the long term (at least 4 years after the end of the project); accessed only by team members within the institution, for auditing and validation purposes. It is also acknowledged that, for some of the outcomes, conditions for exploitation as stated in the Consortium Agreement may apply.

Since the results from this project will make a strong impact in the railway sector, we find it is extremely important to share the data responsibly. Hence datasets that will be open to the public will be released along the journal scientific publications after proper discussion with partners. The datasets will be published via repositories such as the 4TU.Centre for Research Data Archive. In the same way, and in order to motivate re-use of data, the journal articles associated to these datasets will be published in open access and/or self-archived on the MOVINGRAIL website and subject repositories, following the publisher's self-archiving policies.

4. Allocation of resources

TUD researchers can upload up to 1 TB of data to the 4TU.Centre for Research Data Archive (per year) free of charge. Also, the storage capacity and privately accessed drives managed by each partner are already available.

For internal document sharing between partners we make use of SURFdrive, a password protected cloud storage service. Each TUD staff member may use SURFdrive (100 GB storage, access via institutional account). The SURFdrive is shared between all MOVINGRAIL partners.

The WP leaders are in charge of the management of the data from their work package.

Work package	Responsible partner
WP 1	TUBS
WP 2	UoB
WP 3	PARK
WP 4	TUD

5. Data security

Some data will be processed in work laptops of research team members only when allowed. Master copies will be kept at the drives of each respective institution. The IT departments of each institution will maintain the data regarding backups (redundancy) and secure storage (protected access to only team members). Only team members within each institution will have access to the data during the research project. Such data access will be set up by the respective IT departments of each institution. The data that will remain close to the public will be archived at each partner's servers for at least 4 years after the end of the project.

SURFdrive will be used for temporal data storage and for data sharing among different partners coordinated by TUD (coordinator).

6. Ethical aspects

There are no ethical issues that have an impact on data sharing.

It is important to mention, in case there are ethics-related questions or issues arising throughout the project, these will be reported to the scientific coordinator and will be discussed accordingly among team members. Extra advice can be discussed with the Human Research Ethics Committee of TUD (at HREC@tudelft.nl).



7. Other

MOVINGRAIL will make use of the TUD Research Data Framework Policy which can be found via <https://www.tudelft.nl/en/2018/library/researchdatamanagement/tu-delft-research-data-framework-policy-published/>