D 1.2 Data Management Plan

Project acronym:	IN2ZONE
Starting date:	01/12/2020
Duration (in months):	30
Call (part) identifier:	H2020-S2R-OC-IP3-01-2020
Grant agreement no:	101014571 – IP/ITD/CCA – IP3
Due date of deliverable:	31/03/2021
Actual submission date:	30/03/2021
Responsible/Author:	UNIVLEEDS
Dissemination level:	CO
Status:	Version 4

Reviewed: Yes

Document History

Revision	Date	Description
1	4th Mar. 2021	First Draft
2	18th Mar. 2021	Second Draft
3	26th Mar. 2021	Third Draft
4	29th March	Final Draft

Document Contributors:

Name:	Organisation:	Contribution Details:
Mr Nick Fletcher	UNIVLEEDS	Initial Draft & Final Draft
Dr Zsolt Szepessy	EVOPRO	Input Data WP5
Marta Garcia	UNIFE	Input Data WP7 / GDPR
Vitek Lojda	LANKHORST	Input Data WP3
Dr David Connolly	UNIVLEEDS	Input Data WP6
Dr Valeri Markine	TUDELFT	Input Data WP4

Table of Contents:

1. Executive Summary

The aim of the Data Management Plan (DMP) is to manage data used and generated within the IN2ZONE project. It describes how data – information in digital form that can be transmitted/processed – will be collected, processed, stored and managed from the perspective of external accessibility and long-term archiving. It takes account of the particular characteristics of the IN2ZONE project, which has features such as diverse data sources and formats and greater initial uncertainties typical of coordination and support actions. The DMP is therefore designed for flexibility to meet emerging needs.

The DMP addresses the following points for each of the project's WPs:

1. Type of data to be utilised and generated within IN2ZONE. This section identifies and describes the (existing) input data that will be utilised and the output data to be generated by the project.

2. Standards to be used, metadata and quality issues. GDPR and compliance issues are covered as appropriate.

3. How data are exploited and shared/accessed for their verification and reutilisation. The exploitation of data will follow the strategies of each partner concerning their business potential, in accordance with the exploitation plan produced in WP7, and in accordance with the access to data by the partners specified in the CA. Specific restrictions and confidentiality aspects are clarified.

4. Data storage and conservation. Where the data will be held and the arrangements and responsibilities for managing, updating and maintaining the data.

This document can be found under the Cooperation Tool, which guarantees the document integrity and alignment with other plans and principles.

It targets the following participants within the project and in cooperating projects:

- IN2ZONE WP2-7 leads;
- Shift2Rail Joint Undertaking; and
- Other CFM and/or S2R JU projects as relevant.

The following criteria shall be used to update this document: major changes made to the project organization, changes made to the initial project organization structure, changes made to project engineering methods and tools, or amendments that have significant repercussions on the Project or Engineering system process changes.

Each evolution of this document will be translated to a new version which must be verified and approved in accordance with the Configuration Management principles of the project.

2. Abbreviations and Acronyms

Abbreviation / Acronyms	Description
CA	Consortium Agreement
DMP	Data Management Plan
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
LIDA	Leeds Institute for Data Analytics
S2R JU	Shift2Rail Joint Undertaking
TMT	Technical Management Team
WP	Work Package
WPL	Work Package Leader

D 1.2: Data Management Plan v4.0
Monday, 29th March 2021

3. Background

The main aim of IN2ZONE is a S2R project to develop the next generation of transition zone, offering a step-change reduction in maintenance requirements, compared to existing solutions. The associated high-level objectives are closely aligned with the call, and are:

1. Reduction in service affecting delays due to fewer track geometry defects and associated failures (for example, due to track settlement or a localised loss of rail support);

2. Increased network capacity in terms of more frequent trains and higher speeds, due to improved vertical geometry and reduced degradation rate;

3. Reduced lifecycle costs due to a reduction in maintenance and extended operational life of the track and associated assets;

4. Reduction in noise and vibration at the transition locations, due to the provision of a sustained smooth transfer between areas of differing support stiffness; and,

5. To provide a solution for optimum and sustained track support conditions, that is compatible with the next generation track solutions developed within completed IN2TRACK2 and the complimentary, on-going IN2TRACK3 Project.

4. Objective

The purpose of this document is to describe the data management plan for IN2ZONE, The Next Generation of Railway Transition Zones Project, Contract Number 777564. IN2ZONE is a Horizon 2020 S2R Research and Innovation action under Innovation Programme 3 (IP3): Cost-Efficient and Reliable High-Capacity Infrastructure.

This document has been prepared to provide the DMP which addresses the way research data is managed in the IN2ZONE 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. IN2ZONE 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 reporting period (M18 and M30).

This document constitutes the Deliverable D1.2 "Data Management Plan" for the IN2ZONE project, a 30-month co-ordination and support action for transversal exploratory research activities to benefit the railways, within the overall framework of Shift2Rail, which is developing the fundamental building blocks that will allow the creation of the future railway interoperable system.

The partners recognise the challenges of this type of project for data management, so have committed themselves to devising and honouring a formal process to ensure the data are managed and maintained in a way which will mitigate the complexity and diversity of the data sources and ensure an efficient and sustainable process to deliver the project objectives and the ongoing usefulness of its products for the success of the rail sector and its contribution to delivering S2R's wider social, economic and environmental objectives.

5. Data Summary

IN2ZONE collects various kinds of data:

1. Semantic data
2. Stated preference data from surveys and workshops
3. Simulation data
4. Existing/previous research

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 IN2ZONE Project Co-ordinator (UNIVLEEDS) for inclusion in the DMP.

5.1 Data Collected

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
WP2 (Deltares)	**State of the art technical specification**
Purpose	No data generation per se, but a literature review of existing concepts
Types and format	Excel and MSWord/PDF
Reuse of existing data	Not applicable
Origin of data	Literature review of existing/previous research, industry, preceding and collaborating S2R projects
Expected size of data	Less than 1GB
Data utility	The data will be used to support subsequent WPs through the selection of design(s) and the completion of technical specification.
WP3 (LEP)	**Detailed design**
Purpose	The data – in conjunction with WP4 – will look at design aspects of the next generation transition zone – new and contemporary materials, transportation structures, environmental impact, safety methods and hazard identification
Types and format	Word (.docx), Excel (.xlsx), Powerpoint (.pptxp), PDF, CAD (.dwg)
Reuse of existing data	IN2RAIL, existing TUD simulation models (e.g. FEM) and own work

D 1.2: Data Management Plan v4.0
Monday, 29th March 2021

Origin of data	Original research; industry; preceding and collaborating S2R projects and own work
Expected size of data	300 GB
Data utility	The date gathered will go towards creating design guidelines for the next generation transition zone
WP4 (TUD)	**Numerical modelling to support validation of designs**
Purpose	Overlap with WP3 to support/improve validated designs
Types and format	Txt, bin
Reuse of existing data	DELFTCLUSTER; IN2TRACK2; ROSE
Origin of data	Original research; industry; preceding and collaborating S2R projects and own work
Expected size of data	More than 1 TB …
Data utility	Validation of numerical modelling, simulations and transition zone design
WP5 (EVOPRO)	**Resilience based monitoring**
Purpose	Develop a usable and practical for monitoring the condition of transition zones
Types and format	Excel, MS Word, PDF, JSON, XML, various types of CAD systems, Python, C, C++ sources
Reuse of existing data	CENSIS; Assets4Rail; IN2DREAMS
Origin of data	Original research; industry; preceding and collaborating S2R projects and own work (Track/Vehicle and air/space-borne sensing)
Expected size of data	More than 1 TB
Data utility	The use of data fusion to integrate these large datasets, in-tandem with existing datasets provide by railway infrastructure managers, providing a step-change in transition zone monitoring and correction.
WP6 (UNIVLEEDS)	**Large-scale laboratory testing**
Purpose	Understanding transition zone solution behaviour
Types and format	Excel
Reuse of existing data	Only IN2ZONE produced partner data
Origin of data	Original research; industry; preceding and collaborating S2R projects and own work
Expected size of data	200GB
Data utility	Validation modelling and transition zone design
WP7 (UNIFE)	**Impact and dissemination**
Purpose	Dissemination and Communication purposes
Types and format	eps, ai, png, or jpg, xls, doc or jpg

Reuse of existing data	Not applicable
Origin of data	Original research; industry; preceding and collaborating S2R projects and own work
Expected size of data	Variable
Data utility	• Images and logos from partners participating in the project • Database of liaison with Shift2Rail activities: This database contains data such as name, e-mail, company, telephone and field of the partners participating in the Complementary projects or any Rail Stakeholder that will participate in the liaison with Shift2Rail activities.
Specific data of WP7	There are different data and tool that need to be included in the dissemination and communication activities. Included in section 5.

5.2 Specific data contemplated in dissemination activities

Dissemination type	Usage	Policy
Website	Main reference point for project dissemination and data description	Creative Commons whereapplicable. External rightsclearly marked.
Deliverables	Deliverables to the EU and the public. Disseminated through the project website www.in2zone.eu and the EU Cordis system.	Dissemination level set per deliverable. All the public deliverables, appointed as such in the proposal, will be uploaded in the webpage.
Social Media	Support of communicationactivities	Public under the approval of all members of the consortium through the different social media sites of the partners..
Brochure and Newsletters	Regular updates and links towebsite and other channels	Public under the approval of all members of the consortium.
Publications	Scientific and other publications arising from theproject	Open Access to publications.
Research data as laid out in exchange of data outside the project	Underlying project research data	Open Access with limitations due to privacy, as detailed below, in accordance with the FAIR guidelines on Data Management in H2020.

Any other data	To be decided when the data is released or published	Wherever possible, open through open licenses. 'As open as possible, as closed as necessary'; and 'open by default'.

6. FAIR Data

6.1 Making data findable, including provisions for metadata

The data will be securely stored at the Leeds Institute for Data Analytics, University of Leeds, which is underpinned by the Integrated Research Campus, a highly secure and scalable computational infrastructure, which is a Trusted Digital Repository for technical-scientific research data in the UK. It achieved accredited certification in 2017 to the international standard for information security management, ISO/IEC 27001:2013, as well as fully complying with H2020 requirements of making data findable, accessible, interoperable and reusable (FAIR). See the following link for more information:

https://lida.leeds.ac.uk

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 access the data through a Virtual Research Environment (VRE), will be stored and made available in the University of Leeds, LIDA data archive.

6.2 Making data openly accessible

IN2ZONE interfaces with several complimentary Shift2Rail projects, meaning project data will not solely be kept within the project, but shared with other projects associated with the S2R JU (e.g. the complimentary CFM call). This sharing will be beneficial to IN2ZONE by sense-checking the research directions being pursued and allowing for the validation of preliminary results.
IN2ZONE has a strategy for the internal and external sharing of data:
a) Internal data. The exchange of information, documents and deliverables among IN2ZONE project partners will be performed in-line with the Data Management Plan; and
b) External data sharing. Deliverables will be treated differently depending upon whether they are:
  - Public Deliverables. All Public Deliverables will be freely available to all interested stakeholders;
  - Confidential Deliverables. IN2ZONE project partners will provide access to non-public deliverables to the Shift2Rail Joint Undertaking, for the continuation of research activities in future years. Further, IN2ZONE will also provide access to its outcomes to the correspondent project in the CFM Call, and to IN2TRACK2 and S-CODE where required. The details of access procedures will be detailed in the Grant Agreement; and
  - Scientific Publications. All partners will disseminate the project results by disclosing them to the public, including scientific publications in any medium. The partners will make their research data findable, accessible, interoperable and reusable (FAIR). A focus on Open Access publication will be adopted as far as possible – the preferred option is Gold Open Access, but self-archiving via Green open access is also encouraged. All publications from IN2ZONE will be available for download through the project's website. The impact of this activity will be
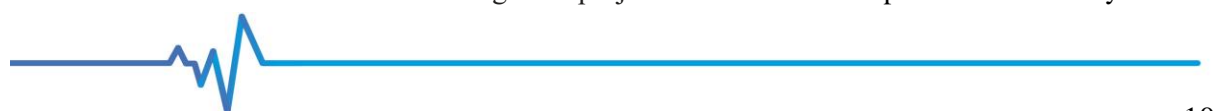
measured with the number of publications, journal impact factor, number of citations, conferences publications etc.

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.

6.3 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.

6.4 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 have 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 LIDA 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 IN2ZONE website and subject repositories, following the publisher's self-archiving policies.

7. Data privacy and personal data

Detailed requirements and descriptions of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants will be described by the tasks that implement them. Where necessary, data will be anonymised or pseudonymised. Data minimisation principles will be followed in line with applicable legislation. The relevance of data collected for tasks is considered in this document.

8. GDPR Compliance

Respective partners will also follow their internal data protection and European GDPR regulations. In line with GDPR, individual beneficiaries are responsible for their own data processing, and they will ensure the implementation and compliance of the procedures and protocols in line with internal processes and national regulations. This also includes options to withdraw consent and procedures that must be in place to deal with privacy violations in a timely manner. Processing of personal data will respect the Data Protection Principles as set out: Lawfulness, fairness and transparency; Purpose limitation; Data minimization; Accuracy; Storage limitation; Integrity and confidentiality; Accountability.

Each beneficiary is reminded that under the General Data Protection Regulation 2016/679, the data controllers and processors are fully accountable for the data processing operations, which means that every beneficiary is ultimately responsible for their data collection and processing.

9. Allocation of resources

University of Leeds researchers can upload up to 1 TB of data to the LIDA 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, they can use the Cooperation Tool for the project.

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

Work Package	Responsible Partner
WP2	Stichting Deltares (Deltares)
WP3	Lankhorst Engineered Products BV (LEP)
WP4	Technische Universiteit Delft (TUD)
WP5	Evopro Innovation KFT (EVOPRO)
WP6	University of Leeds (UNIVLEEDS)
WP7	Union des Industries Ferroviares Européenes (UNIFE)

10. 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.

Cooperation Tool will be used for temporal data storage and for data sharing among different partners coordinated by UNIVLEEDS (Project Co-ordinator/Manager).

11. 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 Engineering, Mathematical & Physical Sciences Faculty Research Ethics Committee at the UNIVLEEDS (at MEECResearchEthics@leeds.ac.uk).