Improved Robotic Platform to perform Maintenance and Upgrading Roadworks: The HERON Approach

Grant Agreement Number: 955356

D1.2: Data Management Plan (first version)

Work package	WP1: Project Coordination and Management
Activity	Task 1.4: Knowledge and information management
Deliverable	D1.2: Data Management Plan (first version)
Authors	Iason Katsamenis, Eftychios Protopapadakis, Anastasios Doulamis, Nikolaos Doulamis, Matthaios Bimpas, Dimitris Kalogeras
Status	Final (F)
Version	1.0
Dissemination Level	Public (PU)
Document date	29/11/2021
Delivery due date	30/11/2021
Actual delivery date	30/11/2021
Internal Reviewers	Carlos Martin-Portugues Montoliu (ACCI), Yannis Handanos (OLO)
	This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 955356.

Document Control Sheet

Version history table			
Version	**Date**	**Modification reason**	**Modifier**
0.1	15/09/2021	Initial Table of Contents and basic structure of the deliverable	Iason Katsamenis
0.2	04/10/2021	Introduction	Iason Katsamenis
0.3	18/10/2021	General Objective of the Data Management Plan	Iason Katsamenis, Eftychios Protopapadakis
0.4	08/11/2021	Findable, Accessible, Interoperable and Re-usable	Iason Katsamenis, Eftychios Protopapadakis
0.5	24/11/2021	HERON datasets	All partners
0.6	25/11/2021	Ethics and Security	Iason Katsamenis, Eftychios Protopapadakis
0.7	26/11/2021	Conlusions	Iason Katsamenis
1.0	30/11/2021	Final version ready for submission	Anastasios Doulamis

Legal Disclaimer

Table of Contents

List of Tables

List of Figures

Abbreviation Lists

Table 1: Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
CA	Consortium Agreement
CCTV	Closed-Circuit Television
CUD	Chaussée Urbaine Démontable (Demountable Urban Roadway)
DMP	Data Management Plan
DOI	Digital Object Identifier
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable and Re-usable
GA	Grant Agreement
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
NDTI	National Department of Transport Infrastructure
OA	Open Access
POPD	Protection of Personal Data
RC	Reinforced Concrete
RI	Road Infrastructure
TC	Technical Committee
UAV	Unmanned Aerial Vehicle
URL	Uniform Resource Locator
WP	Work Package
YOLO	You Only Look Once

Table 2: Abbreviations of the Partners' names

Short name	Participant organization name
ICCS	Institute of Communications and Computer Systems
ACCI	Acciona Construcción S.A.
OLO	Olympia Odos Operation S.A.
UGE	Université Gustave Eiffel
ETHZ	Eidgenössische Technische Hochschule Zürich
ROB	Robotnik Automation
CORTE	Confederation of Organisations in Road Transport Enforcement

STWS	SATWAYS - Proionta Kai Ypiresies Tilematikis Diktyakon Kai Tilepikinoniakon Efarmogon Etairia Periorismenis Efthinis EPE
RISA	RisaSicherheitsanalysen Gmbh
INAC	InnovActs
IKH	Ainoouchaou Pliroforiki SA -IKnowHow-
RG	Resilience Guard Gmbh

Glossary of Terms

Table 3: Glossary of terms

Term	Explanation
Creative Commons	Licenses that allow creators to communicate which rights they reserve, and which rights they waive for the benefit of recipients or other creators.
Data cluster	A group of data that share similar characteristics.
Dataset	A structured collection of data generally associated with a unique body of work.
Green Open Access	Route to open access where the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.
Gold Open Access	Route to open access where an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers.
Metadata	Data that provides information about other data. They serve to provide information on the data produced, collected, or handled.
Open access	The practice of providing online access to scientific information that is free of charge to the reader.
Pseudonymization	The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.
Version	A variation of a dataset, metadata, or deliverable, i.e., an update, edit or change from an earlier version.

Executive Summary

This deliverable is written in the framework of WP1 – Project Coordination and Management of the HERON project under Grant Agreement No. 955356. Deliverable 1.2, namely "Data Management Plan", provides a roadmap for the data generation, collection, storage, and processing processes to be followed throughout the entire HERON project. To this end, it presents the archiving strategies to be followed within the project for the information provided by the data owners during the project and the management of the content collected.

This deliverable report defines an outline of the data that need to be generated/collected and managed within HERON. The objective of the Data Management Plan is that all types of data useful to the project are clearly identified, FAIR (easily Findable, openly Accessible, Interoperable, and Reusable), and finally don't raise security or ethical concerns.

It is underlined that the data will be stored in databases hosted by secured servers that are regularly backed up, so as to ensure that access is only granted to authorized users and that the data are protected from illegal/ inappropriate exposure and loss.

In parallel, it is noted HERON presents no ethical issues. The data collected will be related to traffic flows. For any reason, personal content will not be collected and/or gathered by the consortium. No humans' will be tracked and no humans' opinions will be collected and gathered. Any data will be given to the consortium members only fully respecting national and European legislation and regulations. HERON care does not lay on the personal data content but on the content referring to the traffic flows and the infrastructure. Nevertheless, it is underlined that in case that any personal data, such as persons' identity, is needed to be collected, the information will be treated securely.

Informed consent from the participants regarding the processing of their data will be requested prior to their involvement in project-related research activities, to guarantee that all potential participants "do agree" in their participation in the project execution phase. Moreover, the participants will be given the option to easily withdraw their consent and consequently exercise their rights stemming from the General Data Protection Regulation (GDPR, 2016/679), such as the right to be forgotten as well as the right to resign from the study/research at any time and for any reason.

1 Introduction

1.1 Purpose of the Document

The main purpose of the Data Management Plan deliverable (DMP) is to provide information on the data that will be generated/collected, processed, and used by all the partners of the HERON project. The project aims to contribute to the automated inspection and maintenance of Road Infrastructures (RIs) using robotics-based and machine learning algorithms. DMPs are generally not to be perceived as one-off documents, but instead are likely to evolve during the lifespan of the project, if significant changes occur. DMPs also provide an ideal opportunity to create best practices with regards to metadata standards, file formats, storage, and risk management practices, leading to greater sustainability and longevity of data as well as higher quality standards.

This DMP aims, through the GDPR and FAIR principles, to guide HERON partners in defining:

- The types of data to be created to serve the core objectives of the HERON project.
- How the data will be collected, created, or generated.
- How the data will be stored, processed, and preserved.
- How and which data will be made available to third parties during and after the project's completion.

In a nutshell, the DMP is an official project deliverable (D1.2) of the HERON project and is formulated in compliance with the EC guidelines on FAIR data management in H2020 (EC, 2016). Lastly, the dissemination level of the deliverable is public.

The remainder of this document is organized as follows:

Initially, Section 2 briefly discusses the general objective of the DMP. Section 3 describes the adopted strategies making the data FAIR (Findable Accessible Interoperable and Reusable) according to the EC guidelines[1]. Subsequently, Section 4 provides the details of the identified HERON datasets and their corresponding responsible partners. Section 5 summarises the security or ethical concerns. Lastly, Section 6 concludes this deliverable report.

1.2 Intended Audience

The document's target is the HERON consortium partners, as it describes how all archiving strategies are to be followed within the project for the information provided by the data owners during the project and the management of the content collected.

1.3 Interrelations

This deliverable interacts with all other project activities, as it provides information on the data that will be generated/collected, processed, and used by all the partners of the HERON project.

[1] "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020", Version 3.2, 21 March 2017, available here: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

2 General Objective of the Data Management Plan

The purpose of the specific deliverable is to present a Data Management Plan (DMP) for the HERON project, that:

- Fully describes the project data management life-cycle (see Figure 1)
- Explicitly complies with the principles of the Open Research Data Pilot (EC, 2017).

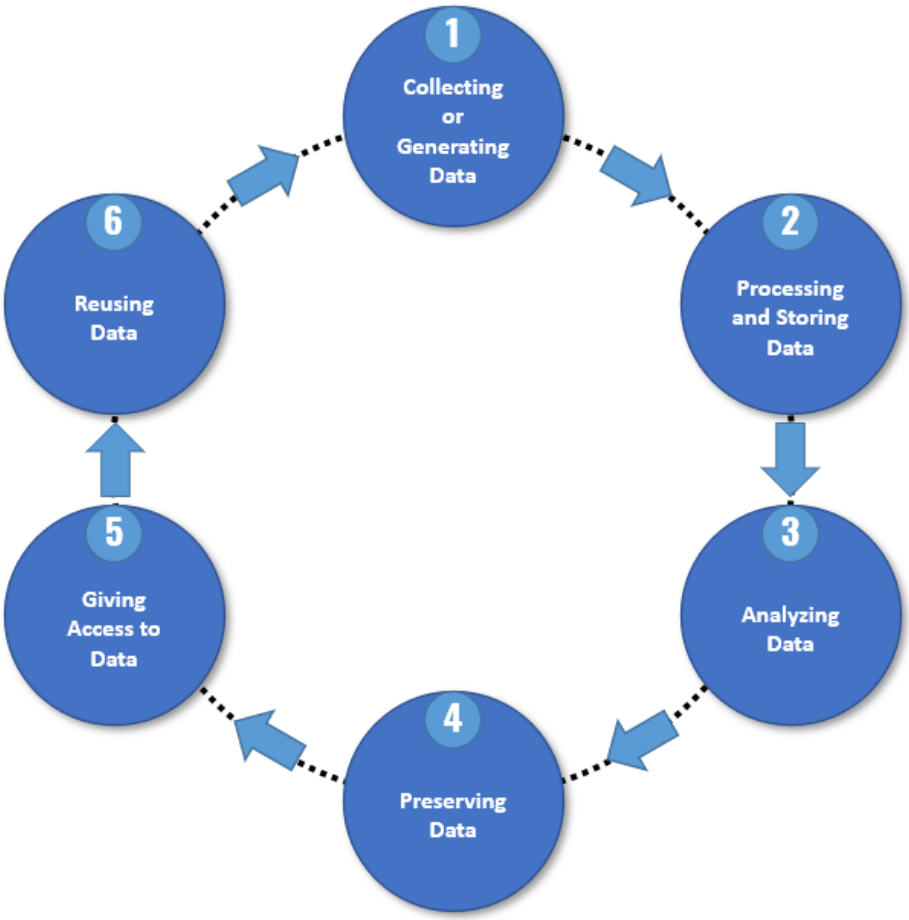

Figure 1: The HERON research data life-cycle phases.

To this end, the project endeavors to offer public access to:

- Findings reported in scientific publications (Self-archiving / 'green' OA and Open access publishing / 'gold' OA) and public deliverables.
- Research data needed to validate the results presented in the deposited scientific publications.
- Research data characterized as “open” that will be generated/collected throughout the project's entire duration.
- Semantic/metadata descriptions and other technical/scientific information.

Research data are planned to be archived at a European open data repository (e.g. Zenodo[2]), as well as a dedicated repository at the HERON project's website. The ultimate goal is to maximize and improve the re-usage and accessibility of the scientific outcomes generated by the HERON project, for the benefit of not only the pilot sites but also for activating the exploitation of the project's findings.

From the same point of view, this DMP, as will be further analyzed in Section 3, attempts to present a management strategy for the data collected, created, or generated during the HERON project that renders them more easily Findable, Accessible, Interoperable, and Re-usable (i.e. compliance with the FAIR principles).

In HERON, the data collected will be related to traffic flows and thus do not present ethical issues. Nevertheless, it is underlined that in case that any personal data, such as persons' identity, is needed to be collected, the information will be treated securely and the collection will be in line with the Protection of Personal Data (POPD) requirements established by the EC and the national regulations. In general, D1.2 outlines the type of data that will be collected, generated, and processed as well as how and why the specific data should be generated/collected, processed, stored, maintained, documented, shared, and reused.

2.1 Intellectual Property Rights (IPR) Management

The HERON project aims at developing technological and scientific results of major importance and interest to the scientific community and industry. The project consortium includes partners whose technologies and technical data may be subjected to Intellectual Property Rights (IPRs). Hence, the consortium will apply appropriate data protection measures and will consequently crosscheck with the concerned partners before data publication. The Consortium Agreement (CA) along with the Grand Agreement (GA, Article 26) address issues relevant to the decision making, the risk management strategies, the legal aspects, the trademarks and the right of each partner to exploit the project findings.

The partners have also agreed on the access rights to Background (GA, Article 25). The term Background is defined in Article 24 of the GA as,

"any data, know-how or information — whatever its form or nature (tangible or intangible), including any rights such as intellectual property rights — that:

(a) is held by the beneficiaries before they acceded to the Agreement, and

(b) is needed to implement the action or exploit the results."

The agreement foresees that any access rights to the results and Background (as specified for each partner in CA, Appendix 1) needed by a partner for performing a research and/or work that is strictly relative to the HERON project, shall be granted on a royalty-free basis. In Attachment 1 of the CA, each partner outlines its own pre-existing know-how along with any specific limitations and/or conditions for implementation and exploitation.

Moreover, since the data collected/processed/generated through the project's activities are of high value, appropriate measures are foreseen to prevent them from being falsified. All data repositories used by the project will comply with specific security and data protection

[2] https://zenodo.org/

measures. In addition, protective measures against infiltration will be taken as well as physical protection of core system parts and access control measures.

2.2 Data life-cycle

The data life-cycle for the sets of data to be generated/collected as well as processed by the HERON project can be described as a six-phase procedure, even though the latter may not always strictly follow the sequence of steps depicted in Figure 1.

In principle, however, the data life-cycle consists of the six phases that are presented below:

- Collecting or Generating Data: This procedure concerns the generation and/or collection of data. Data may be collected by means of questionnaires, surveys, interviews, online repositories, in-situ (e.g., optical) sensors, online applications, etc., or generated following processing (see phase 2) of the collected data or via various simulations. At the least, the collected/generated data should be delivered for the project needs in an appropriate digitized format suitable for further processing.
- Processing and Storing data: The collected/monitored data needs to be initially processed for further exploitation. Under the scope of the present deliverable, the processing is not considered an identical procedure to digitization (even though digitization can be considered a kind of processing), since the former, depending on the data type, refers to validation and data cleaning and pseudonymization, as well as possible data conversion to a convenient format for further analysis. Regarding storing, both processed and raw data have to be stored, but different access protocols apply to processed and unprocessed datasets, with the unprocessed ones strictly restricted for use only by specific consortium partners. It is noted that storing is closely related to phase 5 of the life-cycle (i.e., Giving Access to Data).
- Analyzing Data: This is a crucial and important phase as it includes the various actions that have to be taken for analyzing the data and delivering useful results to serve the key objectives of the project as well as effectively evaluate the set Key Performance Indicators (KPIs).
- Preserving Data: The specific phase is related to the actions that must be followed to maintain the data during and after the completion of the project. Moreover, issues relevant to data recovery measures, secure storage, as well as the transfer of sensitive data are all relevant to the specific stage.
- Giving Access to Data: The specific procedure contains the actions that have to be taken either for sharing the data among the HERON consortium partners or outside the consortium provided that the given dataset is marked as “open”. Therefore, for each of the datasets, it must be determined whether its access will be widely open or restricted to specific groups/end-users.
- Re-using Data: The specific process refers to the that allow the reuse of project research data by future research projects as well as follow-up applications/studies, in order to maximize the impact of the HERON project.

3 Findable, Accessible, Interoperable and Re-usable (FAIR)

HERON complies with the EC-template for DMP and hence it intends to generate data that is FAIR (Findable, Accessible, Interoperable and Re-usable). To accomplish the aforementioned goal, appropriate measures are being taken in line with the EC H2020 pertinent guidelines (EC, 2017). Generally, the HERON project will participate in the Open Research Data Pilot (ORDP), but specific datasets that are identified as "restricted" (see the dataset description and in particular the "Data Sharing" category, in Tables 6 and 17-20 of Section 4) will be excluded from this process.

Making the research data FAIR aims to ease knowledge discovery and innovation, as well as to allow data and knowledge integration and re-use. Compliance of the HERON project with the FAIR principles, but more importantly with its overall concept, ensures that the most vital components addressed in the life-cycle data management are covered.

More specifically, the HERON project is determined to:

- make the research findings and the project innovations openly accessible and findable to the extent possible,
- disseminate and communicate the research findings not only through public deliverables but also, with other supplementary means, such as training and demo events, the project's social media channels and website, publications in peer-reviewed journals, annual magazines, conference presentations, monthly newsletters, etc.

The following sections below present the key considerations towards following the FAIR data policies, that support the utilization of the HERON-produced data.

3.1 Making data findable

The HERON project is focused on securing that the produced data will be identifiable and easily discoverable by the interested stakeholders and beneficiaries. To serve this scope, the project employs a public deliverable policy for the intended deliverables that are marked as "public". This policy requires the public deliverables to be universally accessible, to everyone interested, through the project website and from the EU commission portal[3]. It is worth pointing out herein that, since 2017, it is an obligation that all public deliverables should be also published on the Commission's website.

Furthermore, all the collected/generated datasets will be accompanied by metadata for facilitating their findability. The term "metadata" refers to the informative data provided for the generated/collected research data. In principle, metadata enables project partners, stakeholders and current or potential beneficiaries to find the data that is made available in a repository (either internal for "restricted" data or online for "open" data). The most common definition of metadata is "information about data". Metadata includes in-formation on technical and business procedures, data rules and data constraints as well as logical and physical data structures. Metadata helps organizations understand their data, work-flows and systems. Without reliable metadata, an organization does not have knowledge of what sort of data it processes, what that data represents, how that data flows through the system, and by who these data could be accessed. Some indicative examples of metadata are presented below:

[3] https://cordis.europa.eu/

- Traffic data, including information about the type, format, time, duration, and protocol used for routing.
- Location data, which may refer to the latitude, longitude, and altitude of the user's equipment, direction, identification of the network cell in which the user's device is located, and time.
- Subscribers' data, which may refer to the name, contact details, email, phone number, maybe address of a given subscriber.

This step is deemed to be a vital one since it is very closely related to the data management within the consortium as well as to both the reusability of the dataset and the long-lasting exploitation of the project findings. In HERON, the metadata (e.g., a short description, keywords, data origin) that will accompany the generated datasets, will follow a versatile format (tailored to the dataset) and will be uploaded as a separate accompanied file to the raw or processed data. Indicative metadata and/or short description may include information on the location of the sensors, the date of the measurements, the sensors set up, keywords, sample size, file types, legal restrictions etc. If relevant, blank copies of used questionnaires will be also stored along with the pertinent data.

An overview of the data types and datasets that will be handled within the context of this project is provided in Section 4. Regarding documents/deliverables and dissemination material, clear and harmonized naming conventions will be used.

Regarding the adopted clear versioning strategy, each set of data produced (dataset, deliverables, etc.) will be named uniformly and will include a table with version control. The table (i.e., data history) for controlling the version will include the following fields:

- Version: Version number (starting from 0.1)
- Version Date: The date which the specific version became available
- Description: Brief description of the current version
- Reason for Change: The main reason for producing the new version
- By: Who prepared that version [author(s) and organization(s)]

In order to be able to distinguish and easily identify each dataset they will be assigned with a unique name. This name can also be used as the identifier of the dataset (see Section 4 and in particular Table 5 for more details).

3.2 Making project data openly accessible

The HERON consortium has agreed to follow an open access approach, to the extent possible depending on the specific type of the dataset, that entails to share the project outcomes within but also beyond the project consortium. To this end, selected processed data and findings will be shared with the scientific community, stakeholders and beneficiaries, utilizing both in-paper, online and web-based means such as: publications in open-access platforms and journals, datasets in open access repositories, workshops, the project's website and social networks, a dedicated annually released magazine, press releases, newsletters, etc.

In the HERON project, the data collected will be related to traffic flows. For any reason, personal content will not be collected and/or gathered by the consortium. However, it is noted that in case that any personal data, such as persons' identity, is needed to be collected, the information will be treated securely. To this end, in such a case, the pseudonymous personal data will be stored by the relevant responsible partners in dedicated databases that will be accessible

only to certain authorized users. Processed data that will be shared externally to the consortium will be deposited in the dedicated project website repository and in most cases also to an open European repository (e.g. Zenodo). The uploads in Zenodo, or any other similar open repository are assigned with a Digital Object Identifier (DOI), to make them citable and trackable. Archiving the datasets at an open repository guarantees a long-term and secure preservation of the data at no additional cost to the project. In view of the precautions for protection of personal data, it is explicitly confirmed that any data will become publicly available, only after care is taken with regard to rules of confidentiality, anonymity and protection.

Moreover, HERON will follow the Open Access practice of providing online access to its scientific research articles. Publications and research data that will be made available to third parties will not contain any personal information. For "restricted" data, interested external users must contact the Technical Committee (TC) of the project, which is responsible for the IPR management. The TC will also inform the dataset owner, in order to potentially grant access to the pertinent data. The data should, at a minimum, be anonymized and contain no personal data. It is important to note that open access does not entail an absolute obligation to publish all data. In fact, it is up to the researchers and the EU to decide whether data is suitable and ethical to be published or not.

To protect the copyright of the project knowledge, an adequate licensing scheme may be adopted in the case of open access data (e.g. Creative Commons License). Furthermore, HERON adopts a public deliverable policy, for the overwhelming majority of them.

Table 4: Project's strategy to make the project data openly accessible.

Open datasets	Open datasets and project's public outcomes will be uploaded to the HERON website[4] and the majority of them also to an open European repository. Links to the open shared datasets will be available in the HERON website.
Publicity of the deliverables	All project deliverables will be available to authorized consortium members through the internal online project management tool (Microsoft SharePoint). The public project deliverables and only the executive summaries of the non-public deliverables will be published in the HERON project website[4] as well as in the commission' website.

3.3 Making project data interoperable

The interoperability of the generated/collected project data is closely related to the project impact. In fact, interoperability is the data property that ensures data exchange and re-use between researchers, institutions, companies and other stakeholders. Hence, interoperability aspects have been thoroughly considered in HERON, aiming to enable the maximization of the data value through the utilization of common systems for transmitting and/or exchanging information.

In general, the HERON approach, that foresees the utilization of a uniform Data Management System, amounts to the usage of a few well-established interfaces, limiting the complexity that often stems from the utilization of multiple interface typologies. Interoperability of data will be enabled through the use of standardized data. Hence, the project data will be made available

[4] http://www.heron-h2020.eu/

in formats that are compatible with standard commercial or open software. Overall, the target is to distribute the generated/collected data in popular file formats (e.g., plain text, open spreadsheets, JSON files). The data will be also accompanied by appropriate metadata.

Any script developed aiming to externally offer data will be open and thoroughly documented to enable and encourage re-use from every third-party application without forcing any dependencies and keep the technological barriers low. All security mechanisms will rely on information security industry standard format (e.g. for malicious traffic signatures or incident reporting).

3.4 Enhance project data re-usability

The HERON project intends to make as much scientific data, generated by the project, as possible re-usable to third parties, in order not only to maximize the impact of the project, but also to save resources and to limit duplication of data collection. However, the data classified as "restricted" will not be shared due to security, privacy or other concerns explicitly identified by the data owner whereas the data classified as "embargo" will only be shared following the end of the embargo period that it often stems, at least in the case of journal publications, from the journal's publication policy (in the case of Green OA journals). The research data classified as "open" that is also associated with a published research paper, will be made available only after the manuscript is accepted for publication and always in accordance with the journal policies. For datasets deposited on an open data repository the access is unlimited.

All datasets will be cleared of bad records by means of applying data cleaning algorithms to ensure data integrity and detection of irregularities and patterns. Furthermore, a quality evaluation process will be exercised for assessing the quality of the project generated data and the project process (specifically addressed in Task 1.3 of the DoA, namely "Quality and Ethics Assurance").

4 HERON datasets

In order to specify (a) the HERON datasets, (b) their main features, and (c) the consortium partners that are responsible for these datasets, a questionnaire form was distributed to the consortium members. In particular, it comprises of a set of points/questions related to the datasets that were initially specified by the HERON partners. Thereby, in order to effectively identify the architecture of the project, consortium partners that will generate, collect, analyze or process data, were asked to:

- Specify the datasets that they are in charge of.
- Provide relative feedback, information, and comments at a certain level of detail.

It is underlined that the questionnaire form classifies the datasets into five major categories (data clusters), which are listed below:

- Questionnaire or survey data
- On-site collected data
- Collected data via online repositories
- Generated (simulated) data
- Other

Based on the five aforementioned categories, 15 datasets (see Table 5) have been finally specified by the HERON partners, by completing the distributed dataset questionnaire form (see Table 6). Lastly, the 15 aforementioned identified datasets by the HERON partners (see Table 5) as well as their main characteristics are presented and further described in detail from Table 7 to Table 21.

Table 5: List of the identified datasets that will be utilized in the HERON project.

No.	Responsible partner	Dataset Name	Data URL (given that the dataset is open)
1	ICCS	***Road damage detection dataset*** (See Table 7 for more details)	https://github.com/sekilab/RoadDamageDetector/
2	ICCS	***Real-time traffic cones detection dataset*** (See Table 8 for more details)	https://github.com/MarkDana/RealtimeConeDetection
3	ICCS	***Road damage general detection dataset*** (See Table 9 for more details)	To be provided
4	ICCS	***Image Dataset for Road Surface Markings*** (See Table 10 for more details)	http://display.sbu.ac.ir/databases
5	ICCS	***Obstacle detection dataset*** (See Table 11 for more details)	To be provided

6	ICCS	***UAV Road Images Dataset*** (See Table 12 for more details)	To be provided
7	IKH	***Cracks and Potholes in Road Images*** (See Table 13 for more details)	https://github.com/biankatpas/Cracks-and-Potholes-in-Road-Images-Dataset
8	IKH	***Asphalt Crack Dataset*** (See Table 14 for more details)	https://data.mendeley.com/datasets/xnzhj3x8v4/2
9	IKH	***Pothole-600*** (See Table 15 for more details)	https://sites.google.com/view/pothole-600/dataset?authuser=0
10	IKH	***Annotated Potholes Image Dataset*** (See Table 16 for more details)	https://public.roboflow.com/object-detection/pothole
11	UGE	***CUD images*** (See Table 17 for more details)	N/A
12	UGE	***RC cracks images*** (See Table 18 for more details)	N/A
13	ACCI	***A2 Road technical and inventorying data*** (See Table 19 for more details)	N/A
14	ACCI	***A2 Pavement status & Meteo Data*** (See Table 20 for more details)	N/A
15	OLO	***Road Design, Traffic & Meteo Data*** (See Table 21 for more details)	N/A

Table 6: Dataset form, which was distributed to the HERON project partners.

Dataset information	Description
Dataset Name	Provide a name for the dataset.
Responsible Partner	Identify the partner responsible for this dataset.
Related WP/Task	Specify the WP(s) and Task(s) related to the dataset.
Dataset category	Select in which of the categories below the dataset belongs: • Questionnaire or survey data • On-site collected data • Collected data via online repositories • Generated (simulated) data • Other: **Please specify**
File type	Specify the file type of the data (e.g., JPG, MP4, AVI, XLSX, CSV, PDF, etc.)

Dataset description	• Provide a short description of the dataset. Make sure to include, at least: the points below:
	  ○ Specify its origin: when, where, and how the data have been retrieved/generated.
	  ○ How the data serves the project's goals.
	  ○ Identify which partners, as well as any potential secondary users, might utilize the data and why.

Data size	Specify the size of the dataset. If you do not know exactly or you are not sure about it, please provide an estimation.

Data sharing	• Select if the dataset is:
	  ○ Open: Open for public use.
	  ○ Embargoed: It will become public when the embargo period that is specified by the publisher is over. If the data is classified as embargo, provide also the end date of the embargo period, in format DD/MM/YYYY.
	  ○ Restricted: For internal project use only.
	• Specify how the data will be shared.
	• Specify whether the dataset needs to be licensed and/or credited.
	• Determine what software tools or methods are needed in order to access the data.
	• Specify if the dataset contains personal data and how this issue is handled.

Data URL	• If the data is classified as "Open" (see "Data sharing" above) specify the URL link to the dataset. If the URL link is not yet known, please write down "To be provided". • Otherwise, please write down "N/A".
Archiving & Preservation	• Determine how the data will be stored and retained during and after the project period (e.g., databases, public and local institutional repositories, etc.). • Specify how long the data will be preserved.
Ethics	Identify any ethical or legal issues that may have an impact on data sharing.

Table 7: Road damage detection dataset (ICCS)

Dataset information	Description
Dataset Name	Road damage detection dataset
Responsible Partner	ICCS
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	Collected data via online repositories
File type	• RGB road images: JPG • Annotations: XML
Dataset description	The dataset contains more than 26,000 annotated road images with defects from Japan, India, and the Czech Republic, and thereby is large-scale and heterogeneous. The structure of the data is the same as the Pascal VOC format, and hence the ground truth annotations are represented as bounding boxes. The images were collected using a vehicle-mounted Smartphone (installed on a car's dashboard), similarly to the advanced RGB sensors to be installed to the HERON automated vehicle. Also, the dataset covers two types of deterioration, which are crucial for the detection and maintenance tasks of the HERON system, namely pavement deterioration (e.g., potholes, linear, and alligator cracks) and road marking deterioration (e.g., white line and crosswalk blur). The specific data will be utilized by ICCS and IKH to train and validate the AI algorithms for detecting the various types of deterioration on the road.
Data size	Roughly 1.5GB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • Images on this dataset are available under the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0). • The JPG files are accessible with any image viewer or editing software. The XML files are accessible with any text editor. • Considering issues with privacy, based on visual inspection, when a person's face or a car license plate is clearly reflected in the image, they are blurred out, thus not entailing personal data.
Data URL	https://github.com/sekilab/RoadDamageDetector/
Archiving & Preservation	The data is stored and preserved on GitHub servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 8: Real-time traffic cones detection dataset (ICCS)

Dataset information	Description
Dataset Name	Real-time traffic cones detection dataset
Responsible Partner	ICCS
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	Collected data via online repositories
File type	• RGB road images: PNG • Annotations: XML
Dataset description	The dataset contains 260 annotated road images with traffic cones. The structure of the data is available in PASCAL VOC and YOLOv3 format. The images were collected using a Smartphone and were labelled manually using the annotation tool LabelImg. The specific data will be utilized by ICCS and IKH to train and validate the AI algorithms for detecting traffic cones in order for the HERON system to be able to distribute and remove them in an automated and controlled manner.
Data size	25MB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • The PNG files are accessible with any image viewer or editing software. The XML files are accessible with any text editor. • Images do not contain personal data.
Data URL	https://github.com/MarkDana/RealtimeConeDetection
Archiving & Preservation	The data is stored and preserved on GitHub servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 9: Road damage general detection dataset (ICCS)

Dataset information	Description
Dataset Name	Road damage general detection dataset
Responsible Partner	ICCS
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	Collected data via online repositories
File type	• RGB road images: PNG, JPG • Annotations: XML
Dataset description	The dataset will contain more than 1000 annotated road images with various defects (e.g., cracks, potholes, cracks, edge failure, ageing, rutting, bumps, markings wear, etc.). The structure of the data will be available in PASCAL VOC and YOLOv3 format. The images will have been obtained from heterogeneous sources (e.g., DSLR cameras, UAVs, cell phones, etc.) and will be labelled manually using the annotation tool LabelImg. The specific data will be utilized by ICCS and IKH to train and validate a general detection algorithm (e.g., YOLO) for identifying and localizing various road damages.
Data size	Roughly 500MB
Data sharing	• Open: The data will be openly accessible to the public. • Shareable via URL links. • Images on this dataset will be available under the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0). • The image files are accessible with any image viewer or editing software. The XML files are accessible with any text editor. • Considering issues with privacy, based on visual inspection, when a person's face or a car license plate is clearly reflected in the image, they will be blurred out, thus not entailing personal data.
Data URL	To be provided
Archiving & Preservation	The data will be stored and preserved on GitHub servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 10: Image Dataset for Road Surface Markings (ICCS)

Dataset information	Description
Dataset Name	Image Dataset for Road Surface Markings
Responsible Partner	ICCS
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	Collected data via online repositories
File type	RGB road marking images: JPG
Dataset description	The dataset consists of over 68,000 labelled images of road markings in 18 popular classes [4]. In particular the categories are: (1) Caution Symbol, (2) Caution Text, (3) Crosswalk Caution Symbol, (4) Crosswalk Caution Text, (5) Crosswalk, (6) Forward and Turn Left, (7) Forward and Turn Right, (8) Forward, (9) School, (10) Slow, (11) Speed Bump, (12) Speed Limit, (13) Stop, (14) Stopline, (15) Strain Speed, (16) Turn Left, (17) Turn Right, (18) Yield Line. Furthermore, this dataset contains images of marking signs in three different qualities, excellent, fair and poor. The poor images have serious occlusion, motion blur or are depreciated. It also contains road surface markings under various daylight conditions. The specific data will be utilized by ICCS and IKH to train and validate the AI algorithms for detecting road markings in an RGB image, as well as their status and their category.
Data size	972MB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • The JPG files are accessible with any image viewer or editing software. • Images do not contain personal data.
Data URL	http://display.sbu.ac.ir/databases
Archiving & Preservation	The data is stored and preserved on DiSPLaY Group servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 11: Obstacle detection dataset

Dataset information	Description
Dataset Name	Obstacle detection dataset
Responsible Partner	ICCS
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	Collected data via online repositories
File type	• RGB road images: PNG, JPG • Annotations: XML
Dataset description	The dataset will contain more than 2000 annotated road images with various obstacles (e.g., debris, tires, animals, rubbish, etc.). Additionally, through this dataset, the state of the road edges will be able to be identified (e.g., vegetation, debris, rubbish, etc.). The structure of the data will be available in PASCAL VOC and YOLOv3 format. The images will have been obtained from heterogeneous sources (e.g., DSLR cameras, UAVs, cell phones, etc.) and will be labelled manually using the annotation tool LabelImg. The specific data will be utilized by ICCS and IKH to train and validate AI algorithms for recognizing various obstacles as well as the state of the road edges.
Data size	Roughly 1.0GB
Data sharing	• Open: The data will be openly accessible to the public. • Shareable via URL links. • Images on this dataset will be available under the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0). • The image files are accessible with any image viewer or editing software. The XML files are accessible with any text editor. • Considering issues with privacy, based on visual inspection, when a person's face or a car license plate is clearly reflected in the image, they will be blurred out, thus not entailing personal data.
Data URL	To be provided
Archiving & Preservation	The data will be stored and preserved on GitHub servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 12: UAV Road Images Dataset (ICCS)

Dataset information	Description
Dataset Name	UAV Road Images Dataset
Responsible Partner	ICCS
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Task 5.2)
Dataset category	Collected data via online repositories
File type	PNG, JPG
Dataset description	The dataset will contain more than 100 annotated images of roads, taken from drones, or general UAVs, equipped with cameras. The specific data will be utilized by ICCS and IKH to train and validate the AI algorithms for detecting objects of predefined categories according to the specified system requirements and use cases (e.g., road defects, obstacles, cones, etc.) from individual images taken from drones. Within the pilots, if necessary, we could collect additional RGB images taken from drones, in order to improve the generalization capabilities of the obtained deep learning models, always in compliance with the local regulations at the corresponding time period.
Data size	The dataset will gather at least 100 images. This number can be increased with the necessities of the project.
Data sharing	• Open: The data will be openly accessible to the public. • Shareable via URL links. • Images on this dataset are available under the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0). • The image files will be accessible with any image viewer or editing software. • Considering issues with privacy, based on visual inspection, when a person's face or a car license plate is clearly reflected in the image, they will be blurred out, thus not entailing personal data.
Data URL	To be provided
Archiving & Preservation	The data will be stored and preserved on GitHub servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 13: Cracks and Potholes in Road Images (IKH)

Dataset information	Description
Dataset Name	Cracks and Potholes in Road Images
Responsible Partner	IKH
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	Collected data via online repositories
File type	• RGB road images: JPG • Annotations: PNG masks
Dataset description	The dataset was developed using images made available by Brazilian National Department of Transport Infrastructure (NDTI), through the Access to Information Law - Protocol 50650.003556/2017-28. The images are from highways in the states of Espírito Santo, Rio Grande do Sul and the Federal District. 2235 images were selected manually, following criteria such as not showing signs of vehicles and people, as well as not having image defects. This work consists of 2235 samples of roads where each image has 3 masks that delimit the vehicle's path and crack and pothole defects. These images will be used to increase the precision of the HERON defect detection system. The specific data will be utilized by ICCS and IKH to train and validate the AI algorithms for detecting the various types of deterioration on the road.
Data size	Roughly 200MB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • Images on this dataset are available under the MIT License. • The JPG and PNG files are accessible with any image viewer or editing software. Considering issues with privacy, this dataset does not contain any sensitive or personal data.
Data URL	https://github.com/biankatpas/Cracks-and-Potholes-in-Road-Images-Dataset
Archiving & Preservation	The data is stored and preserved on GitHub servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 14: Asphalt Crack Dataset (IKH)

Dataset information	Description
Dataset Name	Asphalt Crack Dataset
Responsible Partner	IKH
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	Collected data via online repositories
File type	• RGB pavement images: JPG
Dataset description	400 over-the-top images of Asphalt cracks and Non-crack images that are captured using cameras, by Jayanth Balaji A et al. (DOI: 10.17632/xnzhj3x8v4.2) These images will be used for classification purpose, to increase the precision of the HERON defect detection system. The specific data will be utilized by ICCS and IKH to train and validate the AI algorithms for detecting the various types of deterioration on the road.
Data size	Roughly 200MB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • Images on this dataset are available under the Creative Commons Attribution 4.0 International license. • The JPG files are accessible with any image viewer or editing software. • Considering issues with privacy, this dataset does not contain any sensitive or personal data.
Data URL	https://data.mendeley.com/datasets/xnzhj3x8v4/2
Archiving & Preservation	The data is stored and preserved on Mendeley servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 15: Pothole-600 (IKH)

Dataset information	Description
Dataset Name	Pothole-600
Responsible Partner	IKH
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	Collected data via online repositories
File type	• Greyscale road images: PNG • Annotations: PNG masks, PNG heatmaps
Dataset description	This dataset contains 600 over-the-top images of pavement potholes which are annotated with masks, and also contain depth information in the format of PNG heatmap. It was collected using a ZED stereo camera. The road disparity images were estimated using PT-SRP [1]; the disparity transformation algorithm was first introduced in [2], and an advanced version was presented later on in [3]. These images will be used to increase the precision of the HERON defect detection system. The specific data will be utilized by ICCS and IKH to train and validate the AI algorithms for detecting the various types of deterioration on the road.
Data size	212MB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • The PNG files are accessible with any image viewer or editing software. Considering issues with privacy, this dataset does not contain any sensitive or personal data.
Data URL	https://sites.google.com/view/pothole-600/dataset?authuser=0
Archiving & Preservation	The data is stored and preserved on Google Sites servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 16: Annotated Potholes Image Dataset (IKH)

Dataset information	Description
Dataset Name	Annotated Potholes Image Dataset
Responsible Partner	IKH
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	Collected data via online repositories
File type	• RGB road images: JPG • Annotations: TXT bounding boxes
Dataset description	Fully bounding-box annotated image dataset of potholes and damaged roads. This is a collection of 665 images of roads with the potholes labelled. The dataset was created and shared by Atikur Rahman Chitholian as part of his undergraduate thesis and was originally shared on Kaggle. These images will be used to increase the precision of the HERON defect detection system. The specific data will be utilized by ICCS and IKH to train and validate the AI algorithms for detecting the various types of deterioration on the road.
Data size	46MB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • The JPG files are accessible with any image viewer or editing software and the TXT files with any text editor. Considering issues with privacy, this dataset does not contain any sensitive or personal data.
Data URL	https://public.roboflow.com/object-detection/pothole
Archiving & Preservation	The data is stored and preserved on Roboflow and Kaggle servers with unlimited preservation time.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 17: CUD images (UGE)

Dataset information	Description
Dataset Name	CUD images
Responsible Partner	UGE
Related WP/Task	WP2 and WP3
Dataset category	On-site collected data
File type	JPG, PNG
Dataset description	• Images were taken during the timeframe of the HERON project (2021 – 2025), • Taken at the CUD locations (Université Gustave Eiffel, Nantes), • Images were taken with hand-held cameras, • This dataset will be used to detect and inspect CUD slabs. The detection and inspection algorithms will be fitted on this dataset. • This dataset will be used by UGE and ICCS (WP3) for the inspection task. Other partners, like ETH and ROB, may use this dataset to write the algorithms for the trajectories of the robot or the robotic arm itself.
Data size	The dataset will gather at least 50 images. This number can be increased with the necessities of the project.
Data sharing	• The dataset is restricted (for internal project use), • The dataset will be shared by email (with sharing link). • The dataset will have to be credited, as the CUD slabs are an outcome of several research projects involving many partners (not only UGE). • No specific software tool or method is needed in order to access the data. • There is no personal data in this dataset.
Data URL	• N/A
Archiving & Preservation	• The data will be stored and retained during and after the project on personal computers at partner UGE. There will be no sharing through a repository. • The data will be preserved at least until the end of project HERON, and possibly until there is a significant evolution of the design of the CUD slabs (geometry or materials).
Ethics	The CUD slabs have been designed with several partners, which makes it not possible to share this database with entities outside the HERON project.

Table 18: RC cracks images (UGE)

Dataset information	Description
Dataset Name	RC cracks images
Responsible Partner	UGE
Related WP/Task	WP2 and WP3
Dataset category	On-site collected data
File type	JPG, PNG
Dataset description	• Images taken during the timeframe of the HERON project (2021 – 2025), • Taken at the reinforced concrete structures (bridges, buildings and retaining walls), • Images taken with hand-held cameras, • This dataset will be used to detect, assess and repair RC cracks. The detection and assessment algorithms will be fitted on this dataset. • This dataset will be used by UGE and ICCS (WP3) for the inspection task. Other partners, like ETH and ROB, may use this dataset to write the algorithms for the trajectories of the robot or the robotic arm itself.
Data size	The dataset will gather at least 50 images. This number can be increased with the necessities of the project.
Data sharing	• The dataset is restricted (for internal project use), • The dataset will be shared by email (with sharing link). • The dataset will have to be credited, at by default there is no current information on which structures will supply these images. • No specific software tool or method is needed in order to access the data. • There is no personal data in this dataset.
Data URL	• N/A
Archiving & Preservation	• The data will be stored and retained during and after the project on personal computers at partner UGE. There will be no sharing through repository. • The data will be preserved at least until the end of project HERON.
Ethics	There is no ethical or legal issue.

Table 19: A2 Road technical and inventorying data (ACCI)

Dataset information	Description
Dataset Name	A2 Road technical and inventorying data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Tasks 2.3) • WP7 (Task 7.4)
Dataset category	Other: Baseline information (inhouse developed)
File type	Pdf, jpg, dwg, txt and xls formats
Dataset description	The dataset contains the information about: inventorying (traffic signs, bolards, safety elements, etc.), actual design of the motorway and details of the sections that will be used during the pilot activities in A2 (Spanish pilot). Open-source datasets on traffic signs (e.g. Kaggle: https://www.kaggle.com/andrewmvd/road-sign-detection) and road elements to train and validate AI algorithms for detecting diverse types of traffic signs and road elements (technical partners involved in these tasks: IKH and ICCS).
Data size	Roughly 2.0GB
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of HERON Project and the asset owner (Spanish Ministry of Transport) • Shareable via URL links.
Data URL	N/A
Archiving & Preservation	Available at ACCIONA's servers and the internal network of the A2 Torija Traffic Control Centre
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

Table 20: A2 Pavement status & Meteo Data (ACCI)

Dataset information	Description
Dataset Name	A2 Pavement status & Meteo Data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Tasks 2.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	Txt, jpg, mp4 and xls format
Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025). The dataset contains meteorological data and information of the pilot sections (characteristics, location/coordinates, pictures, videos etc.). gathered during the pilot activities. These data will support the inspection and repair activities as well as the drone flights. • Data sources: A2 traffic control centre repositories, RGB cameras, CCTV records, meteo stations at A2. • Main technical partners involved: IKH and ICCS on training and validation of the AI algorithms for detecting the various types of deterioration on the road. ROB and ETH for the UGV functioning
Data size	Roughly 1.0GB
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of HERON Project and the asset owner (Spanish Ministry of Transport) • Shareable via URL links.
Data URL	N/A
Archiving & Preservation	Available at ACCIONA's servers and the internal network of the A2 Torija Traffic Control Centre
Ethics	Considering issues with privacy, based on visual inspection captured using both drones or UGV, when a person's face or a car license plate is clearly reflected in the image, they are blurred out, thus not entailing personal data. The images and video gathered will count on the same ethics consideration than the CCTV images from the A2 motorway.

Table 21: Road Design, Traffic & Meteo Data (OLO)

Dataset information	Description
Dataset Name	Road Design, Traffic & Meteo Data
Responsible Partner	OLO
Related WP/Task	• WP2 (Tasks 2.3) • WP7 (Task 7.4)
Dataset category	Available from the design studies and real-time sensors at the road
File type	• dwg, txt and xls format
Dataset description	The dataset contains the final design studies of the road section which will be used during the pilot study, as well as the available traffic and meteo data.
Data size	Roughly 1.0GB
Data sharing	• Data will be available upon request to all interested partners of the HERON Project
Data URL	N/A
Archiving & Preservation	Available at OLO's servers.
Ethics	No ethical or legal issues are expected for the collected data that may affect data sharing.

5 Ethics and Security

HERON presents no ethical issues. The data collected will be related to traffic flows. For any reason, personal content will not be collected and/or gathered by the consortium. No humans' will be tracked and no humans' opinions will be collected and gathered. Any data will be given to the consortium members only fully respecting national and European legislation and regulations. HERON care does not lay on the personal data content but on the content referring to the traffic flows and the infrastructure.

No personal data will be collected in the HERON project. Any information will be related to the maintenance procedures and it will be scientific in terms of civil engineering construction. Additionally, no personal views and opinions will be collected. Lastly, no individualized opinions will be gathered per subject. Instead, statistical data on the maintenance procedures, statistics of the data will be collected.

However, it is noted that in case that any personal data, such as persons' identity, is needed to be collected, the information will be treated securely. Participants will be provided information in the form of a participant information sheet on how their data will be processed, who is the data controller, and the possible risks or benefits of participation. It will also include a description of the purpose of the research, who is organizing and funding the research, and an explanation about what will happen to the results of the research. It will be made clear that their data is stored and processed only for the purposes of the project. Each participant will be asked to sign an informed consent form to document and verify the information they have been given. Participants will have access – before, during, and after the events – to staff who can advise them on any issues, questions, doubts, and comments they may have.

For the data handling, the project will fully respect:

- Directive 95/46/EC of the European Parliament and the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
- The Charter of Fundamental Rights of the EU, specifically the article concerning the protection of personal data.
- The opinions of the European Group on Ethics in Science and New Technologies in their report "Citizens Rights and New Technologies: A European Challenge" on the Charter on Fundamental Rights related to technological innovation.
- General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679).

6 Conclusions

The present document outlines the basic guidelines and principles for managing the data that will be collected, generated, or processed during the HERON project. In parallel, the deliverable describes all archiving strategies that are to be followed within the project for the information provided by the data owners during the project and the management of the content collected. In order to determine the key information related to the expected datasets and adopted FAIR compliant practices, a questionnaire form was distributed among all relevant HERON partners, requesting their feedback. Through the specific template/survey, a total of 15 datasets have been initially identified by the partners. It is noted that in case it is necessary to add some additional datasets as the project progresses, they will fall under the scope of the identified data clusters and, therefore, the relevant guidelines will apply.

For potential partners, which in contrast to the current phase of the project, at later stages need to collect, generate or somehow process data, this document should serve as a reference point. All project partners must adopt the FAIR principles the guidelines presented in this document to ensure that the appropriate methods are used to collect, store, maintain, use and reuse the research data generated, during as well after the completion of the HERON project. Thereby, compliance with the described guidelines of this document is mandatory for all project partners.

References

[1] Fan, R., Ai, X. and Dahnoun, N., 2018. Road surface 3D reconstruction based on dense subpixel disparity map estimation. IEEE Transactions on Image Processing, 27(6), pp.3025-3035.

[2] Fan, R., Ozgunalp, U., Hosking, B., Liu, M. and Pitas, I., 2019. Pothole detection based on disparity transformation and road surface modeling. IEEE Transactions on Image Processing, 29, pp.897-908.

[3] Fan, R. and Liu, M., 2019. Road damage detection based on unsupervised disparity map segmentation. IEEE Transactions on Intelligent Transportation Systems.

[4] Seyed Hamid Safavi, Mohammad Eslami, Aliasghar Sharifi Najafabadi, Amirhosein Hajihoseini, Mohammadreza Riahi, Maryam Rekabi, Sadaf Sarafan, Rahman Zarnoosheh, Ehsan Khodapanah Aghdam, Sahar Barzegari Banadkoki, Seyed Mohammad Seyedin Navadeh, Farah Torkamani-Azar, "Image Dataset for Persian Road Surface Markings", 10th Iranian Conference on Machine Vision and Image Processing (MVIP), Accepted for Publication, 2017.