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

Grant Agreement Number: 955356

D1.11: Data Management Plan (second version)

Work package	WP1: Project Coordination and Management
Activity	Task 1.4: Knowledge and information management
Deliverable	D1.11: Data Management Plan (second version)
Authors	Iason Katsamenis, Nikolaos Bakalos, Eftychios Protopapadakis, Anastasios Doulamis, Nikolaos Doulamis, Dimitris Kalogeras
Status	Final (F)
Version	1.0
Dissemination Level	Public (PU)
Document date	30/11/2022
Delivery due date	30/11/2022
Actual delivery date	30/11/2022
Internal Reviewers	Dimitrios Bilionis (RISA), Franziska Schmidt (UGE)
	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	19/09/2022	Initial Table of Contents and basic structure of the deliverable	Iason Katsamenis
0.2	03/10/2022	Introduction	Iason Katsamenis
0.3	21/11/2022	HERON datasets	All partners
0.4	25/11/2022	Updated chapters	Iason Katsamenis, Eftychios Protopapadakis
0.5	28/11/2022	Conlusions	Nikolaos Bakalos
1.0	30/11/2022	Final version ready for submission	Anastasios Doulamis, Nikolaos Bakalos

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 1.11, namely "Data Management Plan (second version)", provides an update of the roadmap of processes to be followed throughout the entire HERON project that was described in D 1.2, "Data Management Plan (first version)".

This second version reports on additional datasets captured since the initial submission of the DMP and restates the project's adherence to FAIR principles and ethical and data security regulations in the generation, capturing, storage, and analysis of project-related data.

In parallel, it is noted HERON presents no ethical issues. The data collected will be related to traffic flows. 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 any personal data, such as a person's identity, is needed to be collected, the information will be treated securely.

1 Introduction

1.1 Purpose of the Document

The main purpose of this update of the Data Management Plan deliverable (DMP) is to report on the data that will be generated/collected, processed, and used by all the partners of the HERON project.

This second version of the DMP, through the GDPR and FAIR principles (see Figure 1) described in the first version (D1.2), reports on:

- 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.
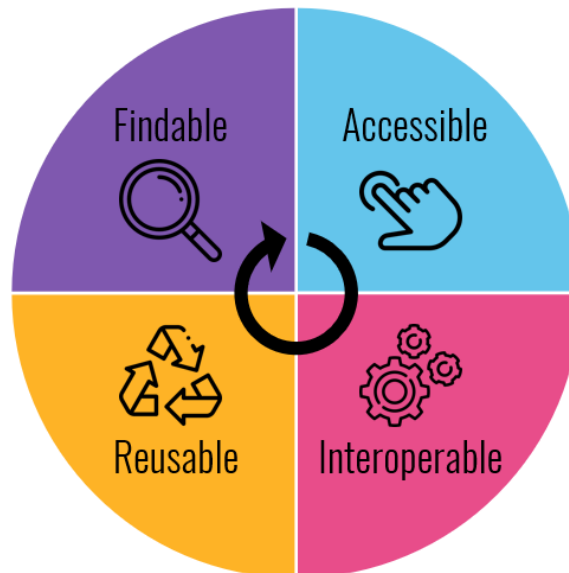
Figure 1: FAIR data principles (Findable, Accessible, Interoperable, and Reusable).

Similarly to D1.2, this update of the DMP is an official project deliverable (D1.11) 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 provides the details of additional HERON datasets and their corresponding responsible partners. Section 3 summarises the security or ethical concerns. Lastly, Section 4 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.

2   HERON datasets

In this section, we report on additional datasets that have been gathered by the HERON partners. The questionnaire used in D1.2 was again utilized for reporting new datasets from project partners. As a result, the following five major categories of datasets are identified:

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

Based on the five aforementioned categories, 20 datasets (see Table 4) have been finally specified by the HERON partners, by completing the distributed dataset questionnaire. Lastly, the 20 aforementioned identified datasets by the HERON partners (see Table 4) as well as their main characteristics are presented and further described in detail from Table 5 to Table 24.

Table 4: 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	***Traffic cone detection dataset*** (See Table 5 for more details)	https://github.com/ ikatsamenis/Cone-Detection
2	ICCS	***UAV defect recognition dataset*** (See Table 6 for more details)	https://github.com/ luisaugustos/Pothole-Recognition
3	ICCS	***Crack semantic segmentation train dataset*** (See Table 7 for more details)	https://github.com/ KangchengLiu/Crack-Detection-and-Segmentation-Dataset-for-UAV-Inspection
4	ICCS	***Crack semantic segmentation test dataset*** (See Table 8 for more details)	To be provided
5	ACCI	***A2 Acciona alligator crack validation data*** (See Table 9 for more details)	N/A
6	ACCI	***A2 Acciona crack validation data*** (See Table 10 for more details)	N/A
7	ACCI	***A2 Acciona distraction validation data*** (See Table 11 for more details)	N/A

8	ACCI	*A2 Acciona guardrail validation data* (See Table 12 for more details)	N/A
9	ACCI	*A2 Acciona incident validation data* (See Table 13 for more details)	N/A
10	ACCI	*A2 Acciona inventorying validation data* (See Table 14 for more details)	N/A
11	ACCI	*A2 Acciona obstacle validation data* (See Table 15 for more details)	N/A
12	ACCI	*A2 Acciona pothole validation data* (See Table 16 for more details)	N/A
13	ACCI	*A2 Acciona faded road marking validation data* (See Table 17 for more details)	N/A
14	ACCI	*A2 Acciona rutting validation data* (See Table 18 for more details)	N/A
15	ACCI	*A2 Acciona signalling validation data* (See Table 19 for more details)	N/A
16	ACCI	*A2 Acciona surface defect validation data* (See Table 20 for more details)	N/A
17	ACCI	*A2 Acciona vegetation validation data* (See Table 21 for more details)	N/A
18	OLO	*Olympia crack validation data* (See Table 22 for more details)	N/A
19	OLO	*Olympia pothole validation data* (See Table 23 for more details)	N/A
20	OLO	*Olympia faded road marking validation data* (See Table 24 for more details)	N/A

Table 5: Traffic cone detection dataset (ICCS)

Dataset information	Description
Dataset Name	Traffic cone 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, PNG • Annotations: TXT
Dataset description	The dataset [1] contains RGB data from heterogeneous sources and sensors (e.g., DSLR cameras, smartphones, UAVs). Furthermore, the images vary in terms of illumination conditions (e.g., overexposure, underexposure), environmental landscapes (e.g., highways, bridges, cities, countrysides), and weather conditions (e.g., cold, hot, sunny, windy, cloudy, rainy, and snowy). In parallel, several images include various types of occlusions, thus making the traffic cone detection task more challenging. Each image has a corresponding .txt file with the bounding box information of the traffic cones (YOLO annotation format: object-class, center_x, center_y, width, height). The total number of RGB images in the dataset is 540 with various resolutions ranging from 114×170 to $2{,}100 \times 1{,}400$. It is underlined that the total number of traffic cones in the entire dataset is 947. From the images of the whole dataset, 92.5% were used for training the deep models, and 7.5% for testing their effectiveness. Among the training data, 80% of them were used for training and the remaining 20% for validation. Thereby, the specific data have been utilized by ICCS and IKH to train and validate the AI algorithms for detecting and localizing traffic cones in order for the HERON system to be able to distribute and remove them in an automated and controlled manner. More details on the use of the specific dataset are presented in Section 4.4.2 of D3.1.
Data size	121MB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • The image files are accessible with any image viewer or editing software. The TXT files are accessible with any text editor. • Images do not contain personal data.
Data URL	https://github.com/ikatsamenis/Cone-Detection
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 6: UAV defect recognition dataset (ICCS)

Dataset information	Description
Dataset Name	UAV defect recognition 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	• UAV images: JPG • Annotations: TXT
Dataset description	The dataset [2] consists of 568 labelled road images, with a resolution of 3840×2160 pixels, from RGB sensors mounted on a UAV. After the pre-processing process, the total number of labelled images in the dataset was 1,362 images. After the pre-processing procedure among the 1,362 UAV images, 70% were used for training (1,191 images), 20% for validation (114 images), and 10% for testing (57 images) the detection capabilities of the trained deep models. More details on the use of the specific dataset are presented in Section 4.6.2 of D3.1.
Data size	Roughly 500MB.
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. The TXT files are accessible with any text editor. • Images do not contain personal data.
Data URL	https://github.com/luisaugustos/Pothole-Recognition
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 7: Crack semantic segmentation train dataset (ICCS)

Dataset information	Description
Dataset Name	Crack semantic segmentation train 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: PNG masks
Dataset description	The specific dataset, which is utilized for the training procedure of the segmentation models, consists of 4,717 RGB images that depict crack defects. More specifically, the datasets that were utilized are: (i) CrackForest Dataset (CFD) with 118 image samples, (ii) CRACK500 with 3,363 image samples, (iii) Cracktree200 with 206 image samples, (iv) DeepCrack with 521 image samples, and (v) GAPS384 with 509 image samples. In particular, the crack semantic segmentation train dataset consists of 4,717 annotated images. It is also noted that during the data pre-processing step, the input RGB images were resized to a resolution of 256×256 pixels. Lastly, 80% of the data is used for training the deep models (3,774 images), while the rest 20% is used for validation (943 images). More details on the use of the specific dataset are presented in Section 4.1.1 of D3.2.
Data size	Roughly 520MB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • The image files are accessible with any image viewer or editing software. • Images do not contain personal data.
Data URL	https://github.com/KangchengLiu/Crack-Detection-and-Segmentation-Dataset-for-UAV-Inspection
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: Crack semantic segmentation test dataset (ICCS)

Dataset information	Description
Dataset Name	Crack semantic segmentation test dataset
Responsible Partner	ICCS
Related WP/Task	• WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4)
Dataset category	On-site collected data
File type	• RGB road images: JPG • Annotations: PNG masks
Dataset description	For the data collection procedure of the test dataset, a GoPro HERO9 Black was used. During the data acquisition process, the optical sensor was mounted on an inspection vehicle. It is emphasized that the acquired data are RGB images with an aspect ratio of 4:3 and, in particular, with a pixel resolution of 5,184×3,888. Moreover, the RGB sensor was set to shoot at a high frame rate and, more specifically, at 50 frames per second in order to ensure sufficient data acquisition regarding both the positive (road surface with cracks) and negative (non-deteriorated road surface) events. It is also highlighted that in order to deal with the severe class imbalance problem from the acquired RGB data, image patches with a resolution of 256×256 were manually extracted and then annotated. The RGB image patches were segmented and verified by engineer experts. More specifically, crack areas were first delineated by a junior computer vision engineer with 4 years of experience and then refined by two engineers with 5 to 15 years of experience. Lastly, all annotated masks were further refined and verified by two senior engineers with more than 20 years of experience in the specific or related field. The crack semantic segmentation test dataset contains 100 annotated images with a resolution of 256×256 in order to evaluate the comparative models that perform the crack segmentation task. More details on the use of the specific dataset are presented in Section 4.1.2 of D3.2.
Data size	Roughly 2.5GB
Data sharing	• Open: The data is openly accessible to the public. • Shareable via URL links. • The image files are accessible with any image viewer or editing software. • Images do not contain 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 9: A2 Acciona alligator crack validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona alligator crack validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains alligator crack data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	  ○ ICCS and IKH to validate the AI algorithms for detecting and localizing the alligator cracks on the road surface.
	  ○ ROB and ETH for the UGV functioning.

Data size	Roughly 4MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 10: A2 Acciona crack validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona crack validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains crack data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	  o ICCS and IKH to validate the AI algorithms for detecting and localizing the cracks on the road surface.
	  o ROB and ETH for the UGV functioning.

Data size	Roughly 6MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 11: A2 Acciona distraction validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona distraction validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains distraction data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	  o ICCS and IKH to validate the AI algorithms for detecting and localizing the distractions in the road infrastructure.
	  o ROB and ETH for the UGV functioning.

Data size	Roughly 0.5MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 12: A2 Acciona guardrail validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona guardrail validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains guardrail data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	  o ICCS and IKH to validate the AI algorithms for detecting and localizing the damaged guardrails in the road infrastructure.
	  o ROB and ETH for the UGV functioning.

Data size	Roughly 6MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 13: A2 Acciona incident validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona incident validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	XLSX and PDF

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains incident data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	  o ICCS and IKH to validate the AI algorithms for detecting incidents and identifying the status of the road edges.
	  o ROB and ETH for the UGV functioning.

Data size	Roughly 2MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 14: A2 Acciona inventorying validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona inventory validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	DOCX and XLSX

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains inventorying data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	  ○ ICCS and IKH to validate the AI algorithms for detecting and localizing various inventorying and road elements (e.g., traffic signs, reflectors, bollards, signals, safety elements, etc.).
	  ○ ROB and ETH for the UGV functioning.

Data size	Roughly 23MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 15: A2 Acciona obstacle validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona obstacle validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains obstacle data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	  o ICCS and IKH to validate the AI algorithms for detecting and localizing the obstacles on the road surface.
	  o ROB and ETH for the UGV functioning.

Data size	Roughly 4MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 16: A2 Acciona pothole validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona pothole validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains pothole data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	   ○ ICCS and IKH to validate the AI algorithms for detecting and localizing the potholes on the road surface.
	   ○ ROB and ETH for the UGV functioning.

Data size	Roughly 7MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 17: A2 Acciona faded road marking validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona faded road marking validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains faded road marking data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	   ○ ICCS and IKH to validate the AI algorithms for detecting and localizing the faded road markings on the road surface.
	   ○ ROB and ETH for the UGV functioning.

Data size	Roughly 6MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 18: A2 Acciona rutting validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona rutting validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains rutting data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	   o ICCS and IKH to validate the AI algorithms for detecting and localizing rutting on the road surface.
	   o ROB and ETH for the UGV functioning.

Data size	Roughly 1MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 19: A2 Acciona signalling validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona signalling validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains signalling data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	  o ICCS and IKH to validate the AI algorithms for detecting and localizing signalling in the road infrastructure.
	  o ROB and ETH for the UGV functioning.

Data size	Roughly 10MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 Acciona surface defect validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona surface defect validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains surface defect data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	   o ICCS and IKH to validate the AI algorithms for detecting and localizing the surface defects.
	   o ROB and ETH for the UGV functioning.

Data size	Roughly 6MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 21: A2 Acciona vegetation validation data (ACCI)

Dataset information	Description
Dataset Name	A2 Acciona alligator crack validation data
Responsible Partner	ACCI
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• Images and information generated during the timeframe of the HERON project (2021 – 2025).
	• The dataset contains data that depict vegetation covering various road elements on the edges of the infrastructure, such as safety elements or signals. The dataset will support the inspection and repair activities as well as the UAV flights.
	• Data sources: A2 traffic control centre repositories, RGB cameras, and CCTV records at A2.
	• Main technical partners involved:
	  o ICCS and IKH to validate the AI algorithms for detecting and localizing vegetation on the road edges that cover safety elements or signals.
	  o ROB and ETH for the UGV functioning.

Data size	Roughly 4MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the 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 22: Olympia crack validation data (OLO)

Dataset information	Description
Dataset Name	Olympia crack validation data
Responsible Partner	OLO
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• The dataset contains crack (flexible and rigid) data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: OLO repositories, RGB cameras, and CCTV records at Olympia Odos.
	• Main technical partners involved:
	   o ICCS and IKH to validate the AI algorithms for detecting and localizing the cracks on the road surface.
	   o ROB and ETH for the UGV functioning.

Data size	Roughly 37MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the HERON Project • Shareable via URL links.
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.

Table 23: Olympia pothole validation data (OLO)

Dataset information	Description
Dataset Name	Olympia pothole validation data
Responsible Partner	OLO
Related WP/Task	• WP2 (Task 2.3) • WP3 (Tasks 3.1, 3.2, 3.3) • WP4 (Task 4.4) • WP5 (Tasks 5.2, 5.3) • WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG

Dataset description	• The dataset contains pothole data and will support the inspection and repair activities as well as the UAV flights.
	• Data sources: OLO repositories, RGB cameras, and CCTV records at Olympia Odos.
	• Main technical partners involved:
	  o ICCS and IKH to validate the AI algorithms for detecting and localizing the potholes on the road surface.
	  o ROB and ETH for the UGV functioning.

Data size	Roughly 3MB This number can be increased with the necessities of the project.
Data sharing	• Restricted: For internal project use only: Data will be available upon request to all interested partners of the HERON Project • Shareable via URL links.
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.

Table 24: Olympia faded road marking validation data (OLO)

Dataset information	Description
Dataset Name	Olympia faded road marking validation data
Responsible Partner	OLO
Related WP/Task	- WP2 (Task 2.3) - WP3 (Tasks 3.1, 3.2, 3.3) - WP4 (Task 4.4) - WP5 (Tasks 5.2, 5.3) - WP7 (Task 7.4)
Dataset category	On-site collected data
File type	JPG
Dataset description	- The dataset contains faded road marking data and will support the inspection and repair activities as well as the UAV flights. - Data sources: OLO repositories, RGB cameras, and CCTV records at Olympia Odos. - Main technical partners involved: - ICCS and IKH to validate the AI algorithms for detecting and localizing the blurred road markings on the road surface. - ROB and ETH for the UGV functioning.
Data size	Roughly 19MB This number can be increased with the necessities of the project.
Data sharing	- Restricted: For internal project use only: Data will be available upon request to all interested partners of the HERON Project - Shareable via URL links.
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.

3   Ethics and Security

For data handling, HERON will fully adhere to:

- 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).

However, similarly to what was reported in D1.2, the data collected by the project are related to traffic flows and infrastructure readings. Personal data will not be collected by the consortium and no humans will be tracked. HERON care does not lay on the personal data content but on the content referring to the traffic flows and the infrastructure.

Moreover, as it was reported in D1.2 in case any personal data, such as a person's identity, is needed, the information will be treated securely in the manner reported and following all ethical and data security regulations and legislation.

4   Conclusions

The present document outlines updates on the principles for managing the data that will be collected, generated, or processed during the HERON project. Through the specific template/survey, a total of 20 additional datasets have been 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 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 and 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 in this document is mandatory for all project partners.

References

[1]   Katsamenis, I., Karolou, E. E., Davradou, A., Protopapadakis, E., Doulamis, A., Doulamis, N., & Kalogeras, D. (2023). TraCon: A novel dataset for real-time traffic cones detection using deep learning. In Novel & Intelligent Digital Systems Conferences (pp. 382-391). Springer, Cham.

[2]   Silva, L. A., Sanchez San Blas, H., Peral García, D., Sales Mendes, A., & Villarubia González, G. (2020). An architectural multi-agent system for a pavement monitoring system with pothole recognition in UAV images. Sensors, 20(21), 6205.