Ref. Ares(2020)7150039 - 27/11/2020

Drones4Safety

Research & Innovation Action (RIA)

Inspection Drones for Ensuring Safety in Transport Infrastructures

D4S Data Management Plan
D1.2

Due date of deliverable: 30.11.2020

Start date of project: June 1st, 2020

Type: Deliverable
WP number: WP1

Responsible institution:
The University of Southern Denmark

Editor and editor's address:
Annika Lindberg (akl@tek.sdu.dk)
Emad Samuel Malki Ebeid (esme@mmmmi.sdu.dk)

Version 1.0 (M6)
Release Date: June 29, 2020

Project funded by the European Commission within the Horizon 2020 Programme		
Dissemination Level		
PU	Public	✓
CO	Confidential, only for members of the consortium (including the Commission Services)	

Change Log

Rev.	Date	Who	Site	Change
1	02/10/2020	Annika Lindberg	SDU	Created initial version.
2.	12/11/2020	Annika Lindberg	SDU	First version for the internal project review. Materials received from the partners Delair, Aric, SDU IMADA, EUCENTRE, EUROCONTROL
3.	20/11/2020	Annika Lindberg	SDU	Updates based on the internal review and comments.

1 Executive Summary

The purpose of the deliverable (D1.2) is to describe the Data Management Plan for the Drones4Safety project. The Data Management Plan (DMP) is a key element for good data management in the project. The DMP describes the project data management life cycle for the data to be collected, processed and/or generated in the project. The plan discusses what kind of data will be collected and processed, which methodology and standards will be applied during data collection and handling, elaborates procedures for sharing and open access to the Drones4Safety data. Furthermore, procedures in relation to the General Data Protection Regulation (GDPR) are defined and how Drones4Safety ensures the protection of the involved companies' data, information and privacy rights.

The Commission is running a flexible pilot under Horizon 2020 called the Open Research Data Pilot (ORD pilot). The ORD pilot aims to improve and maximize access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions. The diagram1 on the following page describes the principles for the Open data pilot program.

As part of making research data findable, accessible, interoperable and re-usable (FAIR), this DMP includes information on:

- the handling of research data during & after the end of the project
- what data will be collected, processed and/or generated
- which methodology & standards will be applied
- whether data will be shared/made open access and
- how data will be curated & preserved (including after the end of the project).

The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. As a minimum, the DMP will be checked and updated in the context of the periodic evaluation/assessment of the project and in time for the final review at the latest.

The data management plan has been submitted and will be collaboratively maintained in the DMP Online tool.
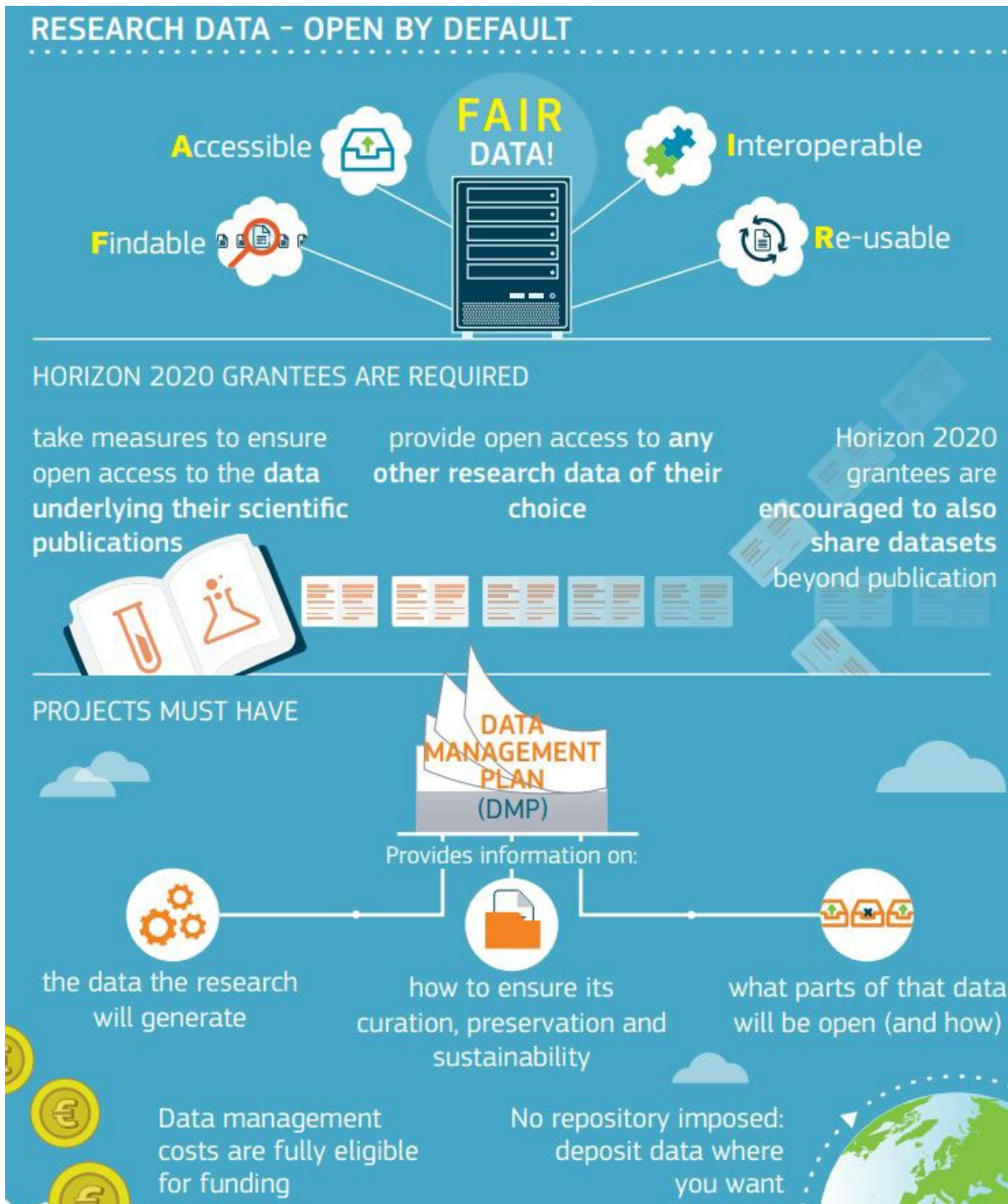
Diagram 1: Research Data - Open by default (Source:
https://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf#view=fit&pagemode=none)

For more information and requirements for data management in Horizon2020 project, see the Horizon 2020 Online Manual, here.

Contents

Acronyms

Acronym	Description
DMP	Data Management Plan
D4S	Drones4Safety
EC	European Commission
EU	European Union
FAIR	Findable, accessible, interoperable and re-usable
GDPR	General Data Protection Regulation
ORD	Open Research Data
IPR	Intellectual Property Rights
WP	Work Package

Explanations for H2020 terms can be found in the funding and tenders glossary:
https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/glossary

2 Introduction

The overarching aim of the D4S project is to develop a system of autonomous, self-charging, and collaborative drones that can inspect a big portion of transportation infrastructures in a continuous operation. The proposed solution harvests energy from overhead power or rail lines in the proximity of the desired infrastructure to be inspected to operate its drones for a longer time. The project gets information about the applicable transport infrastructure to be inspected from open maps and satellite data and forwards that information to its drones to conduct their autonomous mission. This will be achieved through the following objectives:
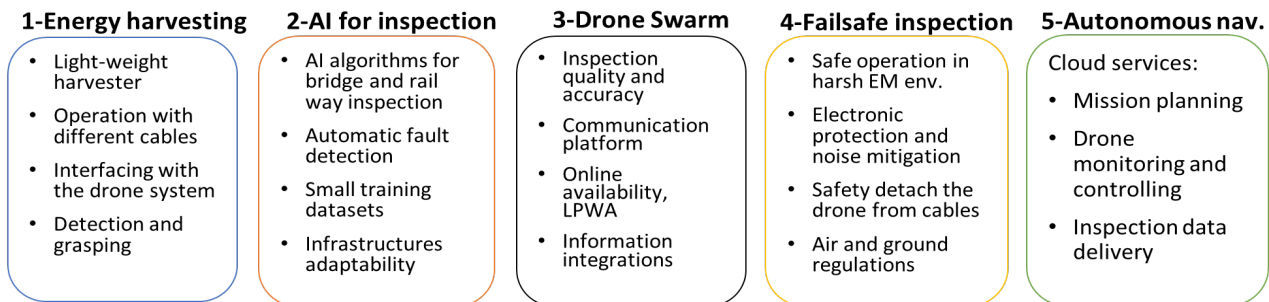
1-Energy harvesting	2-AI for inspection	3-Drone Swarm	4-Failsafe inspection	5-Autonomous nav.
• Light-weight harvester • Operation with different cables • Interfacing with the drone system • Detection and grasping	• AI algorithms for bridge and rail way inspection • Automatic fault detection • Small training datasets • Infrastructures adaptability	• Inspection quality and accuracy • Communication platform • Online availability, LPWA • Information integrations	• Safe operation in harsh EM env. • Electronic protection and noise mitigation • Safety detach the drone from cables • Air and ground regulations	Cloud services: • Mission planning • Drone monitoring and controlling • Inspection data delivery

Figure 1: Drones4Safety Objectives

This D4S Data Management Plan gives an overview of the data and information collected throughout the project and shows the interaction and interrelation of the data collecting activities within and between the work packages. The DMP will also link these activities to the D4E partners and discuss their responsibilities with respect to all aspects of data handling. Furthermore, the D4S DMP will lay out the procedure for data collection, consent procedure, storage, protection, retention and destruction of data, and confirmation that they comply with national and EU legislation. This DMP aims at providing an effective framework to ensure comprehensive collecting and handling of the data used in the project. Thereby and wherever trade secrets of the participating companies and industries are not violated, D4S strives to comply with the open access policy of Horizon 2020.

The DMP is using of the HORIZON 2020 FAIR DATA MANAGEMENT PLAN TEMPLATE and has been written with reference to the Guidelines to FAIR data management in Horizon 2020[1] and the GDPR (Regulation (EU) 2016/679).

The Horizon 2020 fair data management plan template includes a set of questions that will be answered with a level of detail appropriate to the project. It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project. Rather, the DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

The data management plan has been submitted and will be collaboratively maintained in the DMP Online tool. The following chapters are subtracted and will be maintained in the DMP online tool.

---

[1] European Commission, Participant Portal H2020 Online Manual
https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

3 Drones4Safety Data Summary

According to the Horizon 2020 fair data management plan guidelines and the DMP online tool, this section will give answer the following questions:

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

The main purpose of the data collection is:

- The research in the Drones4Safety project involves the collection of data from bridges and railways. This data includes data about defects and anomalies for bridges and railway infrastructure based on imagery and sensor data.

- The project also gets information about the applicable transport infrastructure to be inspected from open maps and satellite data and forwards that information to its drones to conduct their autonomous mission.

Additionally, the Drones4Safety project will in the relation to the dissemination of the project and for promoting the solution generate and collect diverse data outputs, including data on:

- Stakeholder contacts
- Projected economics and solution feasibility
- Public acceptance
- Validation of innovative tools/solutions for the assessment of the service safety and quality
- Benefit-risk assessments

As the project progresses and initial data is identified and collected, further information on the data collected shall be outlined in the data management plan.

Relation to the objectives of the project:

- Data collected is fully in scope and in line with the project objectives. The objective of the Drones4Safety project is to increase the safety of the European civil transport system by developing a cooperative, autonomous and continuously operating drone system to inspect accurately railways and bridges. The drones run onboard AI algorithms that are trained to detect faults on the transportation infrastructure (railways/bridges). Once the fault is detected, the drone reports it to the drone operator via its cloud service system that offers inspection services such as inspection mission control, fault detection reports, and swarm fleet management. This will contribute to increase safety and security of the overall civil transport system and to enhance safe and seamless mobility of cargo and passengers.

Specification for type, formats, re-use, origin of the data.

The data collection and selection of most convenient format will be the responsibility of the relevant task leader and will be integrated in a database hosted on by the project.

- For **Mission control and Navigation** (WP6) is there are two types of data:
    - Open data and derived versions for autonomous navigation incl. map data in OSM XML format, no-fly zones in GeoJSON format, as well as weather data and satellite images from open data sources. These data will be processed into internal representations such as graphs of selected entities from map data etc. The tools for doing so will be made available as open source such that the data collection and processing can be replicated by anyone. The data itself will, thus, not be made available as it would only provide a redundant and soon-to-be out-of-date snapshot of the open data. The data is expected to fill several hundred giga bytes.
    - Inspection data in the form of telemetry data and images collected during test missions and stored in the backend until handed over to the data analysis and management in WP4. The telemetry data will consist of positions and flight information in the format provided by the MAVLink protocol. The images will be geo-tagged and likely be in JPG format. This data is expected to fill several giga bytes. This data will only be collected for internal testing and will not be made available (at least not as part of WP6).

- Data will also be collected in the framework of the **validation activities**:
    - These data will allow the organisation of the study as well as collecting feedback from the various study participants (e.g. pilots, engineers, observers).
    - The data collection will be made in order to feed observation, reporting and analysis of the results. It is anticipated that qualitative data will be collected by all players participating in the validation activities, while quantitative data and associated metrics will be collected through the various technical components of the validation activities. Quantitative results will be used to support the project qualitative analysis considering their statistical relevance.
    - It is not expected to re-use existing data.
    - The data are collected through questionnaires of the participants to the experiments and various technical components of the validation activities.

- Concerning **bridges field application (WP4 and WP7)**, collected data will consist of pictures and movies used to create the digital twin of bridges and to identify structural damages and elements.

- **ARIC** will collect all sensor data which can be used for the positioning solution in Drones4Safety. This will mainly be the sensor data from GNSS, IMU, camera and Lidar. Furthermore, ARIC will store the map information which is used for the path planning.

- **Delair (Alteia)** will collect RGB and IR pictures for bridges and railways with the purpose of providing the required info for photogrammetry tools used to build the 3D model of assets and AI tools able to detect defects on assets.
    - The types and formats of data generated/collected are; Tiff and Jpeg for RGB and Jpeg plus eventual radiometric formats for IR
    - In relation to re-using existing data, some public datasets will be used for the purpose of training AI algorithms. For Bridges an example of public dataset is represented by the CODEBRIM public dataset available at link https://doi.org/10.5281/zenodo.2620293

Size of the data.

- The expected size of the data is difficult to estimate at this stage. The size estimate will be finalized later in the project when the project has progressed, and the first data is identified and collected.

  - ° On average to train properly AI on a specific detection of a defect we need between 500 to 1000 pictures
  - ° For photogrammetry applications the number of pictures is related to the size of the analyzed structure.
  - ° The data for Mission control and Navigation is expected to fill several hundred giga bytes.
  - ° Inspection data in the form of telemetry data and images collected during test missions is expected to fill several giga bytes.

The data utility: to whom will it be useful

The collected research and data will be useful for the partner organizations participating in the project and the scientific and the UAS Stakeholders community outside the project. This target audience can be further split into several stakeholders' segments:

a. Railways, train, power lines and bridges provider, operators and safety experts interested in issues related to safety. They are potential users of the Drones4Safety system and have a key role in justifying its validity and utility for the safe infrastructure management.

b. Drone operators, ICT industry, Large and SMEs specialized in AI, Cloud & Big Data; key innovators in safety, data management and software developers, interested in issues related to automated drone operations. This segment can include also drone engineers, designers, mechanics, drone technicians, pilots, robotic technicians, etc.

c. Industries and industrial associations interested in acquiring a better overview on available experimental innovations.

d. Academic and scientific community: developing research projects in the domains involved in the project (railways, power line, bridges, inspection, energy harvesting, automated drone operations, robotic or satellite navigation, AI, IaaS, Big Data). This segment can include universities, scientific organizations, students, research establishments, professors, etc. Sharing knowledge with scientific community supports the research and fosters future work programmes.

e. Other EC projects and Coordination and Support Actions (CSAs), as well as large national and international initiatives working in similar areas where synergies can be found. Activities with JArus and EUROCAE, Shift2Rail, SESAR and H2020 projects are some examples. To this target the Drones4Safety achievements could give a better and more efficient contribution.

4 Fair Data in the Drones4Safety project

This section is providing a description for how the project and the project partners are making research data **findable, accessible, interoperable** and **re-usable** (FAIR).

4.1 Making data findable, including provisions for metadata

According to the Horizon 2020 fair data management plan guidelines and the DMP online tool, this section will give answer the following questions:

- **Outline the discoverability of data (metadata provision).**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used.**
- **Outline the approach towards search keyword.**
- **Outline the approach for clear versioning.**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**.

The Drones4Safety consortium should work to ensure that its data will be findable, accessible, interoperable and reusable, according to the points below:

It is expected to deposit the generated and collected data in an open online research data repository. Any chosen online repository needs to facilitate identification of data and refer to standard identification mechanisms, which should be outlined ensuring that research outputs and datasets are cross-referencing each other.

All the data sets will be stored in each of the participant entities databases or in the internal project data and document repository. Microsoft Teams/SharePoint is used for project documentation and Gitlab for technical documentation and code. In addition, those categorized as open will be publicly shared through the project website and ZENODO (https://zenodo.org/).

The details for naming conventions used, approach towards search keywords, clear versioning and standards for metadata creation are under construction and will be added in the next version of the data management plan.

4.2 Making data openly accessible

According to the Horizon 2020 fair data management plan guidelines and the DMP online tool, this section will give answer the following questions:

- **Specify which data will be made openly available. If some data is kept closed provide rationale for doing so.**
- **Specify how the data will be made available.**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**

- **Specify where the data and associated metadata, documentation and code are deposited.**
- **Specify how access will be provided in case there are any restrictions.**

In order to maximise the impact of D4S data, the project will facilitate sharing of results and deliverables within and beyond the consortium. Selected data and results will be shared with the scientific community and other stakeholders through publications in scientific journals and presentations at conferences, as well as through open access data repositories.

The Drones4Safety project data is first stored and organized in a database by the data owners and on the project database. All data are made available for verification and re-use, unless the task leader can justify why data cannot be made openly accessible.

The openly available data will be accessible by:

- D4S project web site (https://drones4safety.eu/)
- ZENODO (https://zenodo.org/communities/drones4safety)
- Open access journals

All the relevant research data in the Drones4Safety project will be made available through Zenodo (www.zenodo.org). This repository has been funded by the European Union with the specific objective of storing the research data that will be generated by Framework 7 and Horizon 2020 research projects.

Using this repository has a number of advantages over the self-hosting or use of an institutional repository.

- It provides a more visible and more centralised location for accessing research data, covering a number of EU projects and other sources of research data, rather than a single project and institution. This improves dissemination of the data.

- Data in the repository can be accessed free of charge. A large number of standard licences are supported.

- Data in the repository can easily be searched, mined and otherwise exploited.

- The portal automatically generates a DOI and metadata for each dataset that is submitted. The DOI provides a location-independent reference point that is intended to ensure long-term accessibility of the data. This can be referred to directly from research publications.

- The portal automatically links data to funded EU projects.

- Because it is dedicated to this purpose, the repository provides a long-term storage solution, ensuring the longevity and relevance of the research data. In particular, the data that is generated is guaranteed to be stored and curated long after the project has finished. Without continual intervention, this would difficult to achieve using local and/or institutional repositories.

The metadata that is recorded for each dataset will include the date of submission, the owner of the data, a description of the data content, and a link to the D4S project. Publications from D4S will use Zenodo DOI references to link to relevant datasets, where possible. Any local storage of data will be according to the GDPR guidelines of the respective local entities.

The project coordinator (SDU) will be responsible for continuous evaluation of D4S procedures for data collection, production of metadata, quality of data, as well as policies for storage, backup, archiving, and sharing of data. It should always be ensured that published data are allowed by the owner.

4.3   Making data interoperable

According to the Horizon 2020 fair data management plan guidelines and the DMP online tool, this section will give answer the following questions:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability. If not, will you provide mapping to more commonly used ontologies?**

Partners will observe OpenAIRE guidelines for online interoperability. These guidelines can be found at: https://guidelines.openaire.eu/en/latest/.

As the project progresses and data is identified and collected, further information on making data interoperable shall be outlined in the DMP. In specific, information on data and metadata vocabularies, standards or methodology to follow to facilitate interoperability and whether the project uses standard vocabulary for all data types present to allow interdisciplinary interoperability.

4.4   Increase data re-use (through clarifying licences)

According to the Horizon 2020 fair data management plan guidelines and the DMP online tool, this section will give answer the following questions:

- **Specify how the data will be licensed to permit the widest reuse possible.**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.**
- **Specify whether the data produced and/or used in the project is usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.**
- **Describe data quality assurance processes.**
- **Specify the length of time for which the data will remain re-usable.**

The Drones4Safety project is expected to produce a substantial volume of novel data and knowledge through experimental activities that will be presented to the scientific community, industry, policy-makers and society at large through a carefully designed portfolio of dissemination actions. Datasets uploaded in the project repository will be freely accessible. Potential users are expected to adhere with the Drones4Safety Terms of Use and will be subject to scrutiny by the Drones4Safety team.

As the project progresses and data is identified and collected, further information on increasing data re-use will be outlined in the DMP.

5   Allocation of Resources

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

Data management in the Drones4Safety project will be done as part of the WP1 and SDU, as project coordinator, will be responsible for data management. SDU has allocated a part of the overall WP1 budget and person months to these activities. Costs related to open access to research data are eligible as part of the Horizon 2020 grant. Resources for long term preservation, associated costs and potential value, as well as how data will be kept beyond the project and how long, will be discussed by the whole consortium during the annual project meetings.

Costs cannot be claimed retrospectively. Project beneficiaries will be responsible for applying for reimbursement for costs related to making data accessible to others beyond the consortium.

6   Data Security

According to the Horizon 2020 fair data management plan guidelines and the DMP online tool, this section will give answer the following questions:

- **Address data recovery as well as secure storage and transfer of sensitive data.**

For the duration of the project, datasets will be stored on the responsible partner's storage system. Every partner is responsible to ensure that the data are stored safely and securely and in full compliance with European Union data protection laws. After the completion of the project, all the responsibilities concerning data recovery and secure storage will go to the repository storing the dataset.

The research data in the Drones4Safety project will be made available through Zenodo also after the project end. (www.zenodo.org).

All research data underpinning publications will be made available for verification and re-use unless there are justified reasons for keeping specific datasets confidential. The main elements when considering confidentiality of datasets are:

- Protection of intellectual property regarding new processes, products and technologies where the data could be used to derive sensitive information that would impact the competitive advantage of the consortium or its members,

- Commercial agreements as part of the procurements of components or materials that might foresee the confidentiality of data

- Personal data that might have been collected in the project where sharing them is not allowed by the national and European legislation. As it is not a purpose for the project to collec personal data, this unintended content will either be blurred or permanently removed.

7   Ethical Aspects

According to the Horizon 2020 fair data management plan guidelines and the DMP online tool, this section will give answer the following questions:

- **To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former.**

As also stated in the deliverable D9.1, POPD Requirement No. 1, the research in the Drones4Safety project involves the collection of data from bridges and railways and does not involve tracking, profiling, or any other activity that may infringe the privacy of third persons. No human subjects will be involved in the Drones4Energy research, and the project will not intentionally collect any personal data. In order to limit the amount of unintended content in the drones' recordings, the filming will occur in times when there is a minimum number of persons in the bridges/railways.

The Drones4Safety project will be handling all drone recordings as if they contain personal data until they are screened for unintended content. In that case:

a) the unintended content will either be blurred or the frames will be permanently removed from the recordings. For example: if inspection images should show persons, license plates or similar objects, these areas will be grayed out. We are aiming to automate this process by using state-of-the-art object detection in the form of YoloV4 models.

b) the drones and their recordings will be handled by a limited number of persons,

c) the data will be transferred from the drone to secure storage as soon as possible and deleted from the drone and the drones will be kept in a safe location that locks.

The Drones4Safety project has not entered any ethics issues in the ethical issue table in the administrative proposal forms but will continuously monitor if ethics issues arise – especially concerning drone regulations that are prone to on-going developments in the EU and the individual Member States.

The Coordinator, SDU has a Research Ethics Committee as well as an appointed Data Protection Officer that will be consulted for advice if issues arise.

All partners will assure that the EU standards regarding ethics and data management are fulfilled. D4S partners must comply with the ethical principles and confidentiality.

8 Other issues

According to the Horizon 2020 fair data management plan guidelines and the DMP online tool, this section will give answer the following questions:

- **Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any).**

No additional procedures identified.