HORIZON EUROPE
Digital and emerging technologies for competitiveness and fit for the Green Deal

HYPERIMAGE

A universal spectral imaging sensor platform for industry, agriculture, and autonomous driving.

Starting date of the project: 01/12/2023
Duration: 42 months

= Deliverable D12.4 =
Data Management Plan

Due date of deliverable: 31/05/2024
Actual submission date: 05/12/2024

Responsible WP: Alexander Kabardiadi-Virkovski (Fraunhofer IWS)
Responsible TL: Marina de Souza Faria (ABIMI)
Version: V1

Dissemination level		
PU	Public	x
SE	SENSITIVE – limited under the conditions of the Grant Agreement	

AUTHOR

Author	Institution	Contact (e-mail, phone)
Marina de Souza Faria	ABIMI	souzafaria@amires.eu

DOCUMENT CONTROL

Document version	Date	Change
V0.1	30/06/2024	Partners inputs
V0.2	04/12/2024	First review
V1	05/12/2024	Final Review

VALIDATION

Reviewers		Validation date
Work Package Leader	Alexander Kabardiadi-Virkovski	05/12/2024
Project Manager	Marina de Souza Faria	05/12/2024
Project Coordinator	Alexander Kabardiadi-Virkovski	05/12/2024

DOCUMENT DATA

Keywords	Data, metadata, storage, zenodo.
Point of Contact	Name: Marina de Souza Faria Partner: ABIMI Address: Na okraj 42, Praha 6, 162 00 Phone: +420 730 890 817 E-mail: souzafaria@amires.eu
Delivery date	05/12/2024

DISCLAIMER

Executive Summary

This deliverable represents the 1st version of the HyperImage Data Management Plan prepared in M6. This document aims to describe the data management life cycle for the data sets that will be collected, processed, or generated by the HyperImage project. It specifies what data will be generated and what methodology and standards will be followed, whether and how the data will be exploited and/or be accessible for verification and re-use, and how they will be archived and preserved.

In HyperImage, a consortium-wide accessible workspace in MS Teams implemented since the beginning of the Project M2 which allows partners to share different types of data. The server is hosted by AMIRES and provides access to different project documents and folders.

The HyperImage data can be divided into two categories: data collected & re-used, and data generated. The collected data can be differentiated between data collected inside the project (from partners) and data collected from external sources. The data generated in the project are mainly linked to research activities.

The openly accessible data in the HyperImage project will be the publishable test results that will be used for the development of promotional materials (i.e., Publishable summary, Cordis website, HyperImage website, PowerPoint presentations, leaflets, etc.) and the research data reported in peer-reviewed open access scientific publications and published in the open-access repositories Zenodo.

Table of Contents

1. INTRODUCTION

The Deliverable D12.4 Data Management Plan is part of Task 12.3 (Operational management, quality assurance Period A).

The delay in delivering this document was due to its sensitive nature, as it contained confidential content provided by the project partners. This issue was thoroughly discussed between the project coordinator and the management team to determine the appropriate dissemination level. We had to await a formal decision on whether the document could be made publicly available. Once the coordination team confirmed that the document could be shared openly, the necessary arrangements were promptly made to proceed with its release.

The HyperImage Data Management Plan, as part of making research data findable, accessible, interoperable, and reusable, includes information on: what data will be collected, processed, and/or generated; which methodology & standards will be applied; the handling of research data during & after the end of the project; whether data will be shared/made open access and how data will be curated & preserved (including after the end of the project).

The HyperImage Data Management Plan was developed within the first six months of the project and will be updated throughout the project. It is in line with Article 17 of the Horizon Europe Grant Agreement. As stated in Article 17 of the Grant Agreement the beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles, and by taking all of the following actions:

- Establish a data management plan ('DMP') (and regularly update it).

- As soon as possible and within the deadlines set out in the DMP, deposit the data in a trusted repository; if required in the call conditions, this repository must be federated in the EOSC in compliance with EOSC requirements.

- As soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC0) or a license with equivalent rights, following the principle 'as open as possible as closed as necessary, unless providing open access would in particular:

  o Be against the beneficiary's legitimate interests, including regarding commercial exploitation, or

  o be contrary to any other constraints, in particular, the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP.

- Provide information via the repository about any research output or any other tools and instruments needed to re-use or validate the data.

It is necessary to mention that the disclosure of the project's data should never jeopardize the project's main objective and the potential protection of generated intellectual property (e.g., patent, product design) and further industrial application. The confidentiality obligations, the security obligations, and the obligations to protect personal data still apply. If there is any conflict, the data will not be made openly accessible. All partners contribute to defining the data that will be generated in the project and assess which data can be made public. In case some research data will not be made openly accessible, the Data Management Plan will explain it (IPR issues, exploitation, etc.). Overall, access to the data and research results will always follow the rule "as open as possible, as closed as necessary".

2. HYPERIMAGE DATA MANAGEMENT PLAN

The HyperImage data can be divided into two categories: data collected & re-used, and data generated:

- Collected data can be differentiated between data collected inside the project (from partners) and data collected from external sources. Examples of data collected from partners are previous data on the use cases (LCA, performance, etc.). Both data will remain confidential.

- The data generated in the project are mainly linked to research activities, such as experimental data from material testing technology development, etc.

Each set/type of data will be closely analyzed, and a decision will be made whether the data will be protected/exploited, or whether it will be made publicly available by depositing data without a publication in case the HyperImage partners would find this option appropriate. All partners shall receive prior notice of any planned publication or planned data deposition at least 45 calendar days before the intended date of publication in accordance with the Consortium Agreement and Article 17 of the Grant Agreement.

The scheme (**Figure 1**) summarizes the HyperImage Data Management Plan.
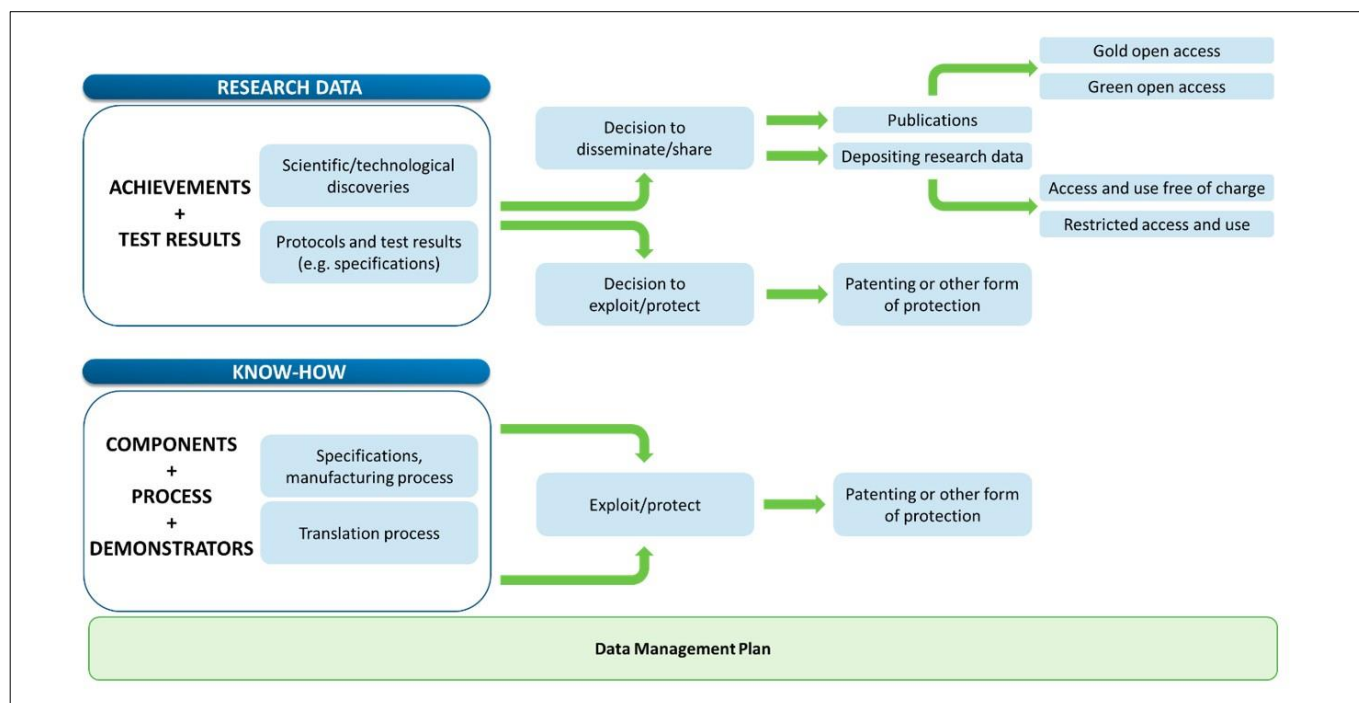
Figure 1: HyperImage data in the wider context of dissemination and exploitation.

2.1. Data Summary

Some partners will make use of existing internal data which is related to research in HyperImage. This data refers to data performance and sustainability of existing use cases which will serve as a baseline for the comparison with the improved versions of the products developed in the project. All such data is considered confidential and will not be shared publicly.

Research data generated by the project is a direct outcome of the research activities carried out within HyperImage by the project partners. Such data will be generated because of project activities, tests, studies, and analyses in partner's laboratories and offices (in case of software use). Experimental data produced in-house will be used to develop new technologies and processes to analyze and improve mainly spectral imaging data generated as well as related ground truth and metadata. Due to the nature of the spectral imaging, data files will be large. These datasets are complemented by experimental and observational data from material analysis, technical system qualifications, and others.

More specifically, this data will be used to:

- Define the format and standards for managing and processing the large amounts of spectral data generated by spectral imaging systems. This includes specifying various properties of the data (such as volume, velocity, and variety), their structure, and how they relate to each other.

  - Defining the spectral data format for storage, machine learning, data analysis, and visualization.

  - Establishing ground truth data formats for different Use Cases.

  - Setting standards for metadata related to the spectral data.

  - Developing concepts for handling data from various spectral imaging sources, ensuring effective and accurate processing.

  - Creating strategies for efficiently managing large volumes of spectral data for the smooth operation of the digital platform.

  - The results will be used to influence design aspects of the cloud-based platform, including data storage, retrieval, and the selection of suitable technologies.

Some of this data may be considered confidential. Such data will not be shared in respect of IP protection and the commercial interest of partners. It will be managed by the respective partners as per their internal policies on IP protection, storage, and data management and security. Each partner has a designated internal repository/data server which is used for storing confidential data.

On the other hand, some data may be considered non-confidential and, as such, will be made publicly accessible. During the project, partners will continuously assess whether the generated data can be disclosed or not.

Research data generated in the project is mostly of interest to the scientific community and other stakeholders (industry, technology developers, etc.) active in the domain of printed electronics. Moreover, producers of products that include printed electronics, such as consumer electronics, smart packaging, and others, could also find the research data of interest.

Dataset Description:

Data summary collected by partner	Fraunhofer IWS
Will you re-use any existing data and what will you re-use it for? What is the origin of the re-used data? State the reasons if re-use of any existing data has been considered but discarded.	NO
What types and formats of data will the project generate or reuse?	It will be generated and use the hyperspectral data primarily in the ENVI-Format. The origin of this data will vary depending on the work packages and use cases. The majority of data generation will take place in WP5 at the test stands, WP7 during the system manufacturing, and in WP8 within the validation phase.
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	The experimental data will be analyzed based on the corresponding statistical (AI) or deterministic models. The parameters of interest in this model vary depending on the use cases. For all planned use cases it will look for the features, that are distributed in the spatial and spectral domain, e.g. homogeneity of the metal layer, materials around off-road vehicles, etc.

What is the expected size of the data that you intend to generate or re-use?	The expected data volume of one single acquisition is typically very high (100s of MBs up to GBs) and can in summary achieve hundreds of terabytes.
What is the origin/provenance of the data, either generated or re-used?	All data will be generated as a result of the project activities in our or the partner's laboratories during the development work and in the application environment for the final validation.
To whom might your data be useful ('data utility'), outside your project?	The data and models can be used outside of the project for autonomous driving, in-line and at-line quality control for different production processes with a project-similar production background, etc.

Data summary collected by partner	WUR
What types and formats of data will the project generate or reuse?	Envi images: Multispectral image for calibration of manufactured snapshot multispectral sensor for WP3 and multispectral image for WP1
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Experimental data will be used to analyze the plant health and properties of leaf chemistry during the growth of the plant in the vertical farm to optimize fertigation and water use.
What is the origin/provenance of the data, either generated or re-used?	All generated data will be from the WUR vertical farm at Bleiswijk, NL.

Data summary collected by partner	NEO
What types and formats of data will the project generate or reuse?	*Simulations data from optical software (ZEMAX) simulating tunable lens. Hyperspectral imaging VNIR and SWIR laboratory data comprising one raw. hyspex and associated .hdr file.*
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	*Assessment of optical quality using tunable lenses for varying object positions. Tests of optical resolutions (using phase shifter) and stabilization of gimbal-free systems*
What is the expected size of the data that you intend to generate or re-use?	*Hyperspectral imaging data size depends on the length of the acquired flight line, typically 2-8 (SWIR) and 10-20 GB (VNIR) for the raw data.*
What is the origin/provenance of the data, either generated or re-used?	*All data are generated for the different work package activities.*
To whom might your data be useful ('data utility'), outside your project?	*Customers*

Data summary collected by partner	SILIOS
What types and formats of data will the project generate or reuse?	Envi images: Multispectral image for calibration of manufactured snapshot multispectral sensor for WP3 and multispectral image for WP1
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Calibration of the manufactured cameras. Wavelength test to refine wavelength specification of snapshot multispectral cameras for the project.
What is the expected size of the data that you intend to generate or re-use?	Not yet defined and can be changed but in preliminary approximation: about 2GB by sensor so about 2x8= 16GB for calibration, and about 80GB for preliminary test
What is the origin/provenance of the data, either generated or re-used?	All data will be generated as a result of the project activities at Silios facilities

Data summary collected by partner	NetCompany (INTRA)
Will you re-use any existing data and what will you re-use it for? What is the origin of the re-used data? State the reasons if re-use of any existing data has been considered but discarded.	INTRA is not expected to provide or generate any datasets for the HyperImage project. However, INTRA will act as the provider of the cloud-based Digital Data Handling Platform that will be used for the collection of the various data from the hyperspectral cameras and sensors deployed across the pilots. Subsequently, the collected data will be made available to the technology providers that will be testing their solutions in the Digital Data Handling Platform in order to analyze the prototype's performance and/or run remote image analysis models and solutions. As a result, we expect to provide data storage during the lifetime of the project, either short- or long-term depending on the specific data usage needs. The Digital Data Handling Platform will be hosted on a public cloud provider (Hetzner) under direct contract with INTRA.
What types and formats of data will the project generate or reuse?	The data types depend on the image formats being provided by spectral imaging sensor systems (e.g. ENVI). The data formats should be compatible with Apache Kafka, e.g., JSON, Avro, etc., and/or the platform's storage component.
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	The purpose of the data collection is to make spectral data available to the solution providers for the performance analysis of their solutions and/or enable them to run remotely spectral data processing and analysis models and solutions.
What is the expected size of the data that you intend to generate or re-use?	This depends on the size of the various data that will be provided as input from the spectral imaging sensor systems deployed on the pilot sites.
What is the origin/provenance of the data, either generated or re-used?	Spectral imaging sensors and systems
To whom might your data be useful ('data utility'), outside your project?	The collected data are expected to be used mainly by the solution providers for the evaluation of their prototypes.

Data summary collected by partner	DIVE
What types and formats of data will the project generate or reuse?	HSI data of bipolar modules and components, Reflectance spectra= $f(x,y,\lambda)$, Format: envi
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Based on the experimental HSI data, the surface quality (e.g. potential contamination) of bipolar modules and components is to be reliably evaluated. The data will be used to create models for reliable objective predictions of the surface condition. In the long term, this should replace time-consuming X-ray examinations.
What is the expected size of the data that you intend to generate or re-use?	100 Gb to 1 Tb
What is the origin/provenance of the data, either generated or re-used?	Data will be generated at Fraunhofer IWS labs, DIVE labs, and on the IFBIP production side.

Data summary collected by partner	Infineon (IFBIP)
Will you re-use any existing data and what will you re-use it for? What is the origin of the re-used data? State the reasons if re-use of any existing data has been considered but discarded.	No re-use of existing data, all data will be generated during the Project.
What types and formats of data will the project generate or reuse?	Optical inspection of surfaces of DUTs including yield, and surface strength compared to X-ray pictures (reference from the established process) HSI data of bipolar modules and components, Reflectance spectra= $f(x,y,l)$, Format: envi
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	The data will be used for machine learning and training of the algorithms by DIVE. Based on the experimental HSI data, the surface quality (e.g. potential contamination) of bipolar modules and components is to be reliably evaluated. The data will be used to create models for reliable objective predictions of the surface condition. In the long term, this should replace time-consuming X-ray examinations.
What is the expected size of the data that you intend to generate or re-use?	defined by DIVE, 100 Gb to 1 Tb
What is the origin/provenance of the data, either generated or re-used?	Both - first generation for training, Re-used for quality inspection (algorithm). Data will be generated at Fraunhofer IWS labs, DIVE labs, and on the IFBIP production side.
To whom might your data be useful ('data utility'), outside your project?	Partners and competitors in the field of semiconductor production

Data summary collected by partner	GROWY
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Getting inside the growth process to get yield gain
What is the origin/provenance of the data, either generated or re-used?	All data will be generated as a result of the project activities at the WUR and/or our vertical AMS3 in Amsterdam.

Data summary collected by partner	ROB
What types and formats of data will the project generate or reuse?	Experimental data from field tests for Autonomous Driving Use Case scenario (i.e. odometry, GPS coordinates, images from the environment, point clouds…)
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Experimental data from field tests will be used to improve the new iterations of the autonomous driving algorithm inside the use case.
What is the origin/provenance of the data, either generated or re-used?	All data will be generated from the UGV sensors onboard during the trial tests.

Data summary collected by partner	KET MARKET
Will you re-use any existing data and what will you re-use it for? What is the origin of the re-used data? State the reasons if re-use of any existing data has been considered but discarded.	Yes, KETMarket will reuse publicly available research project reports (e.g. stored on CORDIS or Techn. Informationsbibliotek), to support future commercial Exploitation of HyperImage Results with SME and industry customers. Such existing data will be used to identify potential users and complementary technologies.
What types and formats of data will the project generate or reuse?	The project will generate information about available imaging and sensor technologies in the market, as well as data about technologies and innovations implemented in companies and commercial products, that have a relationship to the HyperImage Technologies and Solutions.
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	To support SMEs in finding the best-suited technology solution for their product development, to support HyperImage partners in creating market-suited technology offers related to the application cases in HyperImage (autonomous driving, agriculture, quality control in production) and other potential applications.
What is the expected size of the data that you intend to generate or re-use?	several Gigabytes

What is the origin/provenance of the data, either generated or re-used?	Origin of the Data is from public sources (e.g. company and research institutes's websites or CORDIS or the KETMarket digital Marketplace), from use case and pilot case partners, and HyperImage partners and external stakeholders with interest in HyperImage Technology. We are "only" operating the database (the "KETMarket Innovation Marketplace"), while the data is provided by the consortium partners and external stakeholders.
To whom might your data be useful ('data utility'), outside your project?	(1) SMEs looking for a sensor technology solution for their product/process development; (2) Researchers seeking information about who has expertise and solutions in image sensing technologies - for partner search for projects.

Data summary collected by partner	OPTO
What types and formats of data will the project generate or reuse?	The main experimental data will be related to the optical performance of our tunable lenses, beam shifters, and steering mirrors. For what concerns the beam shifter, new antireflective coatings or materials for the window will be used to make the product suitable for its integration into NEO's hyperspectral camera, and their performance will be evaluated. This information might be re-used in the future. The same applies to the liquid lens and steering mirrors. Test reports for the electrical lens modules will be generated once the first product samples are manufactured. This will be done in accordance with the Optotune standard process to generate such tests. Examples can be found at https://www.optotune.com/elm-f-series
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	The experimental data will be used to generate specifications for our devices and to monitor device improvements during the project.
What is the expected size of the data that you intend to generate or re-use?	Few GBs
What is the origin/provenance of the data, either generated or re-used?	All data will be generated as a result of the project activities in Optotune's laboratories.
To whom might your data be useful ('data utility'), outside your project?	Potential customers in the future.

Data summary collected by partner	4K-MEMS
Will you re-use any existing data and what will you re-use it for? What is the origin of the re-used data? State the reasons if re-use of any existing data has been considered but discarded.	No. Except for internally generated confidential data.
What types and formats of data will the project generate or reuse?	The main experimental data will be related to the optical performance of our devices in terms of current, voltage, and modulation speed. Lifetime data will also be taken. This experimental data will be considered confidential

What is the purpose of the data generation or re-use and its relation to the objectives of the project?	The experimental data will be used to generate specifications for our devices and to monitor device improvements during the project.
What is the expected size of the data that you intend to generate or re-use?	On the order of several Tb.
What is the origin/provenance of the data, either generated or re-used?	All data will be generated as a result of the project activities in 4K-MEMS SA laboratories.
To whom might your data be useful ('data utility'), outside your project?	Unknown

2.2. Fair Data

2.2.1. Making data findable, including provisions for metadata

To make data findable, it must be identifiable, locatable, and actionable. All non-confidential data sets in HyperImage will be assigned a unique persistent identifier (e.g., DOI) and metadata will be provided. Such metadata will include at least publication/data author(s), publication title, date of publication, publication venue; statement on Horizon Europe funding; grant project name, acronym, and number; and licensing terms. The data will generally be open access under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC0) or a license with equivalent rights unless exceptions apply. Keywords will be applied as applicable to optimize the possibility for discovery and potential re-use.

2.2.2. Making data accessible

Non-confidential data of HyperImage will be stored in trusted repositories: Zenodo by CERN (https://zenodo.org/) The project Manager, Marina de Souza Faria, set up a community page for the project at Zenodo repository available in Figure **2**. All data sets will be assigned a unique persistent identifier (DOI) and comply with the rules of Horizon Europe.
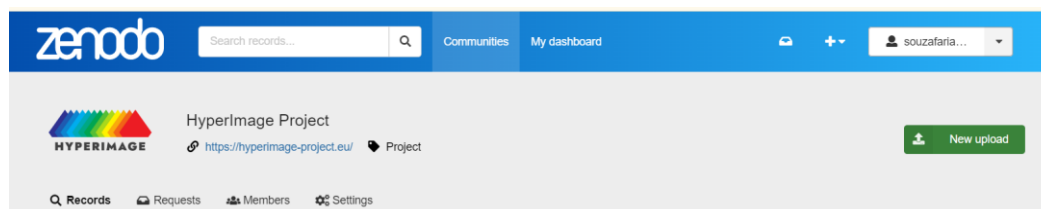

Figure 2: Community for HyperImage Project at Zenodo repository

No embargo period is foreseen for accessing non-confidential data.

Metadata associated with non-confidential data will be made openly available under the standard license (CC0). Public data and associated metadata deposited on the before-mentioned repositories will remain available and findable as long as the repositories exist. It is not foreseen that special software will be required to access the data. However, in case this happens, information on the necessary software to access and read the data will be included in the metadata.

2.2.3. Making data interoperable

Community-endorsed interoperability best practices will be applied to ensure the accessibility and interoperability of the data. Moreover, standard scientific units and abbreviations will be used to make the data interoperable and allow data exchange and reuse within and across disciplines.

In case it is unavoidable to use uncommon or project-specific ontologies or vocabularies, mappings to more commonly used ontologies and publication of the generated ontologies or vocabularies will be considered.

2.3. Increase Data Re-use

Documentation needed for validation of data analysis and facilitation of its re-use (e.g., readme files with information on methodology, codebooks, analyses, variable definitions, units of measurement, etc.) will be based on the internal Quality Management Systems of the project partners. Documentation will be provided as readme files, attached to the data sets published on repositories.

Classified information may not be disclosed to any third party without prior explicit written approval from the owner of the information and the granting authority. All open-access data may be used by third parties during and after the project.

To document data appropriately, files and folders will be named according to local customs in clearly marked HyperImage folders. Datasets will be named with date and version.

Data quality will be ensured by following some best practices such as:

- rigorous data profiling and control of incoming data;

- careful data pipeline design to avoid duplicate data;

- accurate gathering of data requirements; enforcement of data integrity;

- integration of data lineage traceability into the data pipelines.

Therefore, research data will be stored in open standard formats to ensure reusability also in the future. Making available spectral imaging data to industrial users (reference data for training AI algorithms) is a key part of the HyperImage business and exploitation strategy. Therefore, suited license models for data access and re-use will be developed during the project runtime, always following fair and reasonable access and pricing models. Whereas not access to the community-provided data but primarily maintenance and continuous update of the platform backend infrastructure will require financial compensation after the end of the project. In any case, free license models are foreseen for non-commercial access to the dataset and algorithms. They will be complemented with a commercial license with full access to all data especially industrial-adapted AI models for production. The software platform will be accessible publicly via the internet, with the mentioned differences in use for non-profit commercial licenses.

Data formats or storage requirements may change during data acquisition. For this, the Data Management Plan will be updated during the project.

Further to the FAIR principles, the Data Management Plan should also address research outputs other than data. HyperImage Data Management Plan will consider aspects related to the allocation of resources, data security, and ethical aspects.

2.3.1. Other research outputs

Other research outputs that will be generated might include information on materials, components, parts, tools, and instruments necessary for the implementation of the project (e.g., equipment for the laser soldering process, heterogeneous packaging solutions, microfluidic cartridges, or read-out instrumentation). For some partners, digital outputs are not expected to be produced. These outputs will be documented in the form of deliverables.

Samples and prototypes will be carefully tracked and ensured that at the end of the project they either will be destroyed or returned to the owner.

2.3.2. Allocation of resources

Confidential data will be stored on internal data servers. Costs for the management and maintenance of these datasets will be covered by the internal resources of the partners (e.g., as part of central security and data storage procedures or budgets allocated for IPR and dissemination/publications), as set in Article 6.2 (other goods, works and services) and Annex 2 of the Grant Agreement. Costs associated with depositing non-confidential data on the repositories (i.e., personnel costs) and their maintenance during the project duration will be charged to the project. However, Zenodo Research Data Repository is free, therefore it will not require any maintenance cost.

Where appropriate, grant funding will be used to make the project results publicly accessible. Each partner has an appointed person responsible for data management generated by each organization. Central data management and coordination of the project will be overseen by the Project Coordinator and the Project Manager. Data deposited in the Zenodo repository will be retained for the lifetime of the repository.

2.3.3. Data security

All data, confidential and non-confidential, will be stored on the internal data servers of the partners with appropriate safety measures put in place for safe storage and potential data recovery. Storage and archiving of data on internal data servers will guarantee their long-term preservation and curation. The consortium-wide accessible data storage (OwnCloud) servers are physically located in Prague, Czech Republic, and hosted by AMIRES. Access is provided through credentials customized for each partner to ensure security. Moreover, INTRA will also set up a cloud for the partners to store and access hyperspectral images.

Non-confidential data will be stored in open-access repositories with appropriate security measures put in place, such as Zenodo and the HyperImage website.

2.4. Ethics

The applicants will consider appropriately the implications of the autonomous decision-making by the AI-embedded system in the autonomous vehicle test case and the adversarial, critical, or damaging effects in case of risks or threats, such as design or technical faults, defects, outages, attacks, misuse, inappropriate use and apply the appropriate detection and response mechanisms for these undesirable adverse effects.

All partners will ensure that all ethical issues related to activities in the project are addressed in compliance with ethical principles, and the applicable international and national law. The project activities will pay special attention to the EU's AI High-Level Expert Group's Ethics Guidelines for Trustworthy Artificial Intelligence.

This includes the provision of the informed consent forms, which outline what data will be collected, for what reason, how and where it will be stored and used, and when and how it will be finally destroyed (if applicable) or reused. This form also contains contact details of the designated Data Protection Officer and will be provided in the language and terms intelligible to the participants.

All data management procedures will follow regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons concerning the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, GDPR). Protecting the identity of participants is guaranteed by anonymization or, in case of evolving data in time, pseudonymization. All personal information will be handled following the GPDR regulations and guidelines of the respective ethical committee and national legislation. Following the GPDR guidelines, information will be given to

participants of the data handling and sharing, and register owning. Participants will have the possibility to deny the anonymization and sharing of their data. No personal data will be transferred from the EU to a non-EU country or international organization.

Research data will be stored at partners' internal network servers and secured and protected with appropriate identification procedures, as security measures. All this data will be pseudonymized at the data collection phase and anonymized after the data collection phase.

After the research, research data will be retained for further study. The data will be used for algorithm and application research and development. Provisionally, data will be saved for 5 years, and after this period the need for extension will be reviewed yearly. The nominated person will take care of destroying the data when there is no research needed anymore. Manually processed personal data will be protected by restricting the number of persons who have access to handle the data. If a list of ID numbers linked to personal data is created, it will be stored in a closed cabin/place with restricted access. Data exchange will utilize only anonymized data and protected data transfer services and will follow GDPR.

3. CONCLUSIONS

The purpose of the Data Management Plan is to identify the data sets that will be collected, processed, or generated by HyperImage. It specifies which data will be generated, which methodology and standards will be followed, whether and how the data will be exploited and/or made accessible for verification and re-use, and how the data will be curated and preserved. For all those data which will not be made openly accessible, an explanation was provided. The reason for not making the data accessible is mostly since such data will be subject to exploitation/protection of generated intellectual property.

Non-confidential research data will be made public by depositing it in an open-access repositories Zenodo open-access repository. These data sets will be accompanied by the relevant metadata and assigned a persistent unique identifier.

4. DEGREE OF PROGRESS

The deliverable is 100% fulfilled. The Data Management Plan will be updated during the entire lifetime of the project.

5. DISSEMINATION LEVEL

This deliverable is public.