Enhancing Copernicus Security Services – EU governmental crisis management hub for forced population displacement

Intermediate Data Management Plan (DMP), D1.5

WP1 – Project Management 1 - 1st period

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under GA 101190051. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Lead Contributor	Liza Panagiotopoulou (GSH)
Contributors	-
Reviewers	Jose Santos (SATCEN)

Due Date	28/02/2026
Delivery Date	27/02/2026
Type	DMP — Data Management Plan
Dissemination Level	PU - Public
Keywords	Data management, FAIR principles, Open Science, Data security, Ethical Compliance, GDPR

Document History

Version	Date	Description	Author	Description/ Action	Validation
0.1	15/02/2026	First Draft	L. Panagiotopoulou (GSH)	Initial draft of the updated document	Feedback from consortium partners collected and incorporated.
0.2	20/02/2026	Second Draft	L. Panagiotopoulou (GSH)	Internal review process	SATCEN review completed. Feedback incorporated
0.3	23/02/2026	Final Draft	L. Panagiotopoulou (GSH)	Finalisation of document	SAB security check completed. Feedback incorporated.
1.0	27/02/2026	Final Version	L. Panagiotopoulou (GSH)	Submission to EC	
1.1	26/03/2026	Updated Final version	L. Panagiotopoulou (GSH)	Inclusion of missing inputs of C3I	
1.1	02/04/2026	Updated Final version	L. Panagiotopoulou (GSH)	Submission to EC	

Legal Disclaimer

This document reflects only the views of the author(s). Neither the European Commission nor the Granting Authority (European Health and Digital Executive Agency) is in any way responsible for any use that may be made of the information it contains.

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The above referenced consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

Contents

Executive Summary

The present deliverable, D1.5 “Intermediate Data Management Plan (DMP), is associated with Task T1.3, “Initial Data Management Plan”, within WP1 “Project Management 1 – 1st period”, coordinated by GSH, the project coordinator. The project’s DMP was first submitted at M4 as D1.3 “Initial Data Management Plan”, with this intermediate update scheduled for M15 as Deliverable D1.5.

The initial DMP described the procedures for data collection, processing, storage, sharing, and preservation throughout the THEIA project lifecycle, ensuring effective and secure data management in compliance with applicable regulations. This intermediate version further refines the data management approach based on the project’s progress and aims to provide the most detailed overview possible at this stage. Any newly identified datasets or changes not previously foreseen will be documented in subsequent DMP updates.

The Consortium strongly supports Open Science principles and recognises the value of data reuse for the European innovation ecosystem and economy. Research data associated with exploitable results will not be openly shared where this could compromise commercialisation, data protection or confidentiality requirements. All other research data will be made publicly available through appropriate open-access repositories.

The document opens with an overview of THEIA’s objectives and the DMP framework, followed by a summary of the data and a detailed description of the application of FAIR principles. It also addresses resource allocation, data security, legal aspects, intellectual property rights and ethical considerations.

List of Tables

List of Figures

List of Acronyms / Abbreviations

Table 1. List of Acronyms/Abbreviations

Acronym / Abbreviation	Explanation
CA	Consortium Agreement
CSS	Copernicus Security Services
DMP	Data Management Plan
DPO	Data Protection Officer
DoA	Description of the Action
EC	European Commission
ECHR	European Convention on Human Rights
EU	European Union
FRONTEX	European Border and Coast Guard Agency
GA	Grant Agreement
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
JCA	Joint Controllers Agreement
KERs	Key Exploitable Results
MoM	Minutes of Meeting
NATO	North Atlantic Treaty Organisation
ParDPO	Partner Data Protection Officer
PSO	Project Security Officer
SAB	Security Advisory Board
TFEU	Treaty on the Functioning of the European Union
WP	Work Package

1. Introduction

Addressing critical challenges such as population displacement due to conflicts, exacerbated by factors like climate change, extreme weather events, food shortages, and poverty, remains paramount. The implementation of THEIA, integrating data fusion, processing, and analysis, particularly leveraging Geospatial Artificial Intelligence (GeoAI) and Machine Learning, is poised to enhance the efficacy of existing services significantly. Through the amalgamation of multi-temporal data and diverse datasets, THEIA empowers better decision-making and adapts to evolving policy and user needs. This technological advancement, bolstered by GeoAI, augments detection capabilities and ensures timely access to crucial information, bridging the gap between capabilities and stringent security demands.

By integrating non-space data and end-user intelligence, THEIA's supply chains add value not only at the operational level but also at regional and local levels, facilitating improved coordination. Furthermore, THEIA catalyzes fostering EU-independent capabilities and technologies, thereby bolstering the European space ecosystem's consolidation and ensuring the sustainable coexistence of legacy and New-Space solutions. Its services cater to a wide array of end-users, including EU entities such as SATCEN and Frontex, Member State Ministries of Defence, Intelligence Agencies, Police Forces, NATO and potentially Extra-EU National and Supranational Entities such as the United Nations.

This document is the intermediate version of the Data Management Plan (DMP). It updates the initial DMP submitted at M4 and describes how data are generated, collected and managed.

The WP1 consists of the following Tasks:

- Task 1.1: "1st-period project management and coordination towards objectives" [M1-M15]
- Task 1.2: "Initiation of Quality Assurance and Risk Management Framework" [M1-M15]
- **Task 1.3: "Initial Data Management Plan" [M1-M15]**

This document is the second output of Task 1.3 and represents the sixth deliverable of WP1.

The following sub-sections present the scope and objectives, as well as the structure of the document.

1.1 Purpose and scope of the deliverable

The purpose of this intermediate version of DMP is to outline and update how data are collected, processed, stored, shared and preserved throughout the THEIA project lifecycle. The Plan ensures on the one hand that data is managed effectively and securely and on the other hand in compliance with relevant regulations. The scope of the deliverable is to present the DMP, a Plan

to define the strategies, policies and procedures for handling data throughout the project's lifecycle.

1.2 Structure of the deliverable

This document consists of the following chapters:

- The executive summary of the deliverable.
- **Chapter 1** which provides a short description of THEIA project objectives along with the purpose, the scope and the structure of the deliverable.
- **Chapter 2** which provides an overview of THEIA Data Management Plan.
- **Chapter 3** which elaborates Data Summary.
- **Chapter 4** which describes the FAIR Data.
- **Chapters 5, 6 7, 8, 9, 10 and 11** which outline the Allocation of Resources, Data Security, Legal Framework, Personal Data Management, IPR, Ethics and Conclusions, respectively.

1.3 References

- Project GA with No. 101190051
- THEIA Partners CA
- D1.2 - Risk Identification Management and Quality Assurance Plan
- D1.3 - Initial Data Management Plan (DMP)
- D3.1 - THEIA Ethics version 1 (1st period)
- D1.4 - Intermediate report on Risk Identification Management & Quality Assurance Plan
- Draft D3.2 - THEIA Ethics version 2 (1st period)
- General Project Review Consolidated Report of the Interim Technical Review
- Draft JCA
- Information about Open Science principles in EU-funded projects is based on official EU policies and guidelines, including: Open Science Policy by the European Commission, FAIR Data Principles etc.

2. THEIA Data Management Plan

The THEIA DMP outlines the project's general data management policy and approach. It is a living document that will evolve throughout the project's duration refining policies and providing further details on generated and collected datasets.

This is the intermediate version of the DMP, with one more version to follow: a final version of DMP at project's end (M30), which will also cover the post-project period. The DMP will be regularly updated to reflect project progress and evolving datasets, including significant changes such as new data, consortium policy updates, changes in consortium composition, or other relevant external factors.

The first version of the DMP aimed to provide an initial report on the data to be collected during the project, ensuring compliance with FAIR data principles (Findable, Accessible, Interoperable, Reusable) and detailing how the data are stored, in which repository and how they are preserved. The current version builds upon the initial DMP, adding updates based on data and experiences gathered during the project and follows, as also the initial version, the guidelines of the EU Horizon Europe programme.

This intermediate DMP includes the information specified in the GA (e.g., Data Types, Formats, Ethics and Intellectual Property, Short- and Long-term Data Storage, etc.), following a slightly different template than initially planned, in order to streamline the presentation of available information.

2.1 Introduction

It is well known that the effective data management always facilitates knowledge discovery and at the same time drives innovation by enabling the integration and reuse of data and knowledge. In Horizon EU projects, the DMP plays a crucial role in ensuring structured and efficient data management.

The research conducted in THEIA focuses on enhancing CSS, specifically supporting the EU governmental crisis management hub for forced population displacement, which may pose security challenges. A key priority for the consortium is therefore to safeguard the data collected or generated throughout this research. To address this, THEIA's data security policies will be comprehensively outlined in the DMP.

At the same time, Horizon Europe aims to accelerate research by promoting data that is Findable, Accessible, Interoperable, and Reusable (FAIR) which contribute in making effective data management essential. In this framework and by integrating best practices in data handling, THEIA ensures compliance with FAIR principles while securing sensitive research materials.

The main goal of the DMP is to define the data management life cycle for all data collected, processed or generated throughout the project. The DMP also aims to produce data that other researchers may benefit from. It will outline the types of data produced, how they will be utilised or made accessible for verification and reuse and the methods for their curation and preservation. This task will facilitate the ongoing identification, monitoring, and qualification of data produced by participants in the THEIA project, ensuring proper usage and compliance with established data security policies throughout the entire process.

An almost standard data cycle includes the following stages: Collecting or Creating Data, Processing Data, Analysing Data, Preserving Data, Sharing Data etc. as presented in the following Diagram (Figure 1).

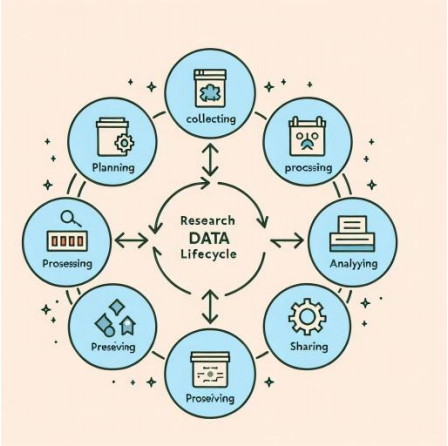

Figure 1. Data cycle

It has to be mentioned that at each stage of the data management cycle, it is imperative to fully comply to IPR and contractual clauses. Additionally, it is important to define where the data will be stored and for how long along with the access rights.

2.2 Purpose of the DMP

The purpose of this second version of the DMP is to:

1. Present updated information about the datasets that will be generated, collected and processed during the THEIA project.
2. Detail how the data will be shared, made open, curated and preserved.
3. Provide an overview and an update of the necessary measures for ensuring satisfactory data management from an ethical and security perspective.

The key objectives of the THEIA DMP are:

- To ensure full respect and observance of security and ethical issues throughout the project's implementation.
- To align the THEIA project with the EU's data management policy, which promotes open access to data funded by the EU.

These objectives are fully compliant with the existing Horizon Europe guidelines.

2.3 GA Provisions

The THEIA DMP is an internal, evolving document aligned with the GA and EU-funded project guidelines. It adapts to dataset updates and changes in Consortium policies throughout the project.

The GA also outlines provisions for DMP content and the project's open access policy.

Specifically, the DMP is discussed in GA the following paragraphs:

- ANNEX 1 (DoA): T1.3 Initial Data Management Plan (page 69)
- ANNEX 5: Open access policy (page 11)
- ANNEX I (PART B): Open access policy in “1.2.10 Open sciences practices” and “1.2.11 Research Data Management and Management of other Research Outputs” (page 31)

2.4 General principles of THEIA DMP

As already mentioned above a DMP outlines how data will be handled throughout a project's lifecycle by ensuring compliance with FAIR principles (Findable, Accessible, Interoperable, Reusable). In the DMP all data processes such as data collection, processing, storage, sharing and long-term preservation while at the same time addressing ethical, legal and security considerations are defined. A DMP ensures transparency, reproducibility and open access, aligning with funder policies, such as the EU’s Open Science guidelines.

2.4.1 Open Science principles

The Open Science principles of EU-funded projects aim to promote the principles of transparency, accessibility and collaboration in research which are embedded in Horizon Europe and previous EU research frameworks, ensuring that publicly funded research benefits society.

A large amount of data from the THEIA project will be provided as Open Access research data, meaning it can be accessed and reused in accordance with the terms and conditions outlined in the GA. Openly available data will be freely accessible to both experts and non-experts, allowing for exploitation, reproduction and dissemination at no cost to external users.

2.4.1.1 Open Science: Open Access to Scientific Publications

In line with the principles, as also described in the GA of the THEIA project, the consortium partners are required to ensure open access to all peer-reviewed scientific publications which are related to their results from the project. For this reason, a "green" open access model will be adopted, making the papers available through the project website, and additionally through a publication repository.

More specifically, the following practices shall be considered:

- at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication will be deposited in a trusted repository for scientific publications which for THEIA is Zenodo[1].
- immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or a license with equivalent rights. This license enables re-users to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. CC BY includes the following elements:

 BY: credit must be given to the creator.

Figure 2. CC BY license elements

- information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.

Metadata of deposited publications must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machineactionable) and provide information at least about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and if possible, for their organizations and the grant. Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication.

[1] https://zenodo.org/communities/eu/

2.4.1.2 Open Science: Research Data Management

The THEIA research data management is in line with the FAIR principles by taking all of the following actions:

- establish a DMP which will be regularly updated (versions on M4, M15 and M30).
- as soon as possible deposit the data in a trusted repository
- as soon as possible 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 (CC 0) or a license/dedication with equivalent rights, following the principle "as open as possible as closed as necessary".

As explained also above, a Zenodo.org community, that is curated by GSH, was set up for openly sharing project data, including research outputs such as datasets to allow others to build on our work, while new data can be added even after the project ends.

2.4.2 Alignment to FAIR

FAIR stands for Findable, Accessible, Interoperable, and Reusable. According to FAIR data principles, research data should be easily discoverable, include clear access information, be compatible with other datasets, and be reusable.

The THEIA DMP ensures that:

- Research data will be managed according to FAIR principles to enhance reusability.
- The project outlines in this comprehensive DMP how data will be handled, stored and shared.
- Open access to research data will be adopted (see above), with exceptions for security, privacy or commercial interests (IPR issues).

During the implementation of the project, the template recommended for the Horizon Europe DMP, which consists of a series of questions primarily focused on compliance with FAIR principles, are addressed and answered by the consortium to ensure THEIA data complies to these principles. This template is provided in Annex I of this document.

2.4.3 Data Governance of THEIA

Beyond the compliance to FAIR principles, THEIA project establishes a solid data governance framework which prioritises data protection and security policies to ensure compliance and safeguard sensitive information. Additionally, THEIA consortium partners consider the data quality important, as the project relies on diverse datasets from multiple sources. So, ensuring

high data quality throughout the entire data lifecycle and at the same time maintaining data security are key governance priorities for the project.

The project is committed to delivering the highest-quality information and analysis, with a quality control process in place (see Deliverables D1.2 and D1.4). Furthermore, THEIA's data governance framework ensures the proper management of critical and sensitive data, including information collected from sensors and hardware components of THEIA services. For the handling of this data all appropriate methods will be applied such as anonymisation, encryption, if necessary, etc. for mitigating risks associated with third-party access.

This data governance strategy aims to maximise the value of collected and utilised data, enhancing information sharing, collaboration and innovation in business models. Data governance in THEIA is exercised and monitored at multiple levels, both at the central and partner levels, ensuring effective oversight and compliance.

At the central level, THEIA's data governance structure includes the THEIA Data Manager, the Ethical Manager and the Project Security Officer (PSO). These roles are assigned as follows:

- Data Manager: Ms. L. Panagiotopoulou (GSH – Project Coordinator)
- Ethical Manager: Prof. Dr. V. Papakonstantinou (MPL – Ethical Manager)
- Project Security Officer (PSO): Mr. C. Kontopoulos (GSH)

Each consortium partner has designated a Partner Data Protection Officer (ParDPO) or a responsible contact to ensure compliance with data protection regulations. THEIA relies on these contacts to uphold personal data protection.

The following Table lists the names and contact details of the DPOs or designated data protection contacts provided by the partners to handle compliance, respond to data subject requests, and liaise with authorities. Each partner is responsible of their own processing of personal data within the project research. These DPOs and data protection contacts must handle data subject requests and other ethics issues, duly informing the appointed professionals in THEIA's data governance structure, and considering the internal policies for ethics in THEIA established in deliverable D3.2 – THEIA Ethics Version 2 (1st period)".

Table 2. THEIA ParDPO or responsible contact to ensure compliance with data protection regulations

	THEIA partner	Name	Email
1	GSH	Flora Chalakatevaki	flora@geosystems-hellas.gr
2	SATCEN	Joana Gonçalves	dpo@satcen.europa.eu
3	OHB DS	Dr. Dominik Röhr	dominik.roehr@ohb-ds.de
4	LUXSPACE	Peter Wagner	peter.wagner@luxspace.lu
5	ICCS	Dr. Dimitrios Kalogeras	dpo@iccs.gr

6	SPACE-SI	Dr. Hubert Fröhlich	Hubert.froehlich@space.si
7	AIT	Ines Burgstaller (Project's contact person)	ines.burgstaller@ait.ac.at
		DPO	dpo@ait.ac.at
8	ED Luxembourg	George Boultadakis	george.boultadakis@eurodyn.com
9	CREO	Inspektor Ochrony Danych/Data Protection Officer	iod@creotech.pl
10	C3I	Panagiotis Pierettis	p.pierettis@c3i.com.cy
11	WTOC	Panagiotis Raptis	piraptis@wtoc.eu
12	MPL	Javier Lopez-Guzman (designated data protection contact)	info@mplegal.services

Regarding the Data Protection Officer development, the General Project Review Consolidated Report (He) issued on 02/10/2025 established a specific Expert opinion on deliverables in its Annex 1. Amongst those elements, the reviewers recommend: "Additionally, it would be advisable to provide a clearer description of the Ethical Manager's role and responsibilities, encourage close alignment with the DPOs, and consider appointing a certified consortium-level DPO to support consistent application of data protection rules across partners."

After careful assessment of the personal data processing operations in THEIA, and the information collection amongst the partners, the consortium has decided not to appoint a consortium-level DPO. A consortium-level DPO would add administrative complexity and cost without providing significant additional benefit, given that the project's processing of personal data is limited and already managed at the partner level. THEIA project will rely on the network of existing DPOs and designated data protection contacts for the handling of data subjects' requests. This network is coordinated by THEIA's data governance structure, including its Ethical Manager and Data Manager. This decision is in compliance with the relevant legal applicable provisions to the project. A detailed legal assessment is offered in section 5.2.1 Data Protection Officer assessment in D3.2 THEIA Ethics version 2 (1st period).

The main responsibilities of the THEIA Data Manager include close alignment with DPOs and designated data protection contacts:

- Developing and maintaining the DMP, outlining how data will be collected, stored, processed, shared and preserved.
- Ensuring compliance with Open Access and FAIR principles.
- Coordinating the work for defining standards, formats and metadata for data collection.
- Tracking data use and ensuring adherence to the DMP.

- Reporting data-related progress and issues to the consortium and funding bodies.

As mentioned above, to effectively address ethical concerns within the project, an Ethical Manager has been appointed and is encouraged to work closely with the ParDPOs. His role and responsibilities are described in detail in Section 6. Ethics by design in THEIA, 6.1 Ethics manager of "D3.2 – THEIA Ethics Version 2 (1st period)".

Besides, an Ethics Advisor was appointed who is not part of the SAB, help ensure the project's full compliance with ethical standards and reporting periodically to HaDEA. His responsibilities described in section 7. THEIA Ethics advisor of "D3.2 – THEIA Ethics Version 2 (1st period)". Finally, the PSO's primary role, as outlined in the GA Security Section (p. 5), is to ensure compliance with EU classified information handling rules and applicable security procedures. The PSO is supported by the SAB in these duties.

3. Data Summary

This section provides a description of the data utilised in the THEIA project. Various types and formats of data will be presented, including their purpose, estimated size (if known) and source/origin.

As part of the intermediate DMP update, partners updated/ revised the types of data to be collected and generated throughout the project, including the rationale and necessity for their collection.

The project coordinator conducted an update of the initial data survey within the consortium using an Excel file titled "*Inventory of Existing and New Datasets*".

One sheet, "*Existing Datasets per WP and per Task*", includes information on:

- Existing data acquired from public/open sources or other sources that will be used,
- Type of data and their sources,
- Nature of the data,
- Format and expected size,
- How the data will be made available internally within the project,
- Whether the data will be accessible to external actors,
- And whether any personal data will be used for project purposes.

The second sheet, "*New Datasets per WP and per Task*", collects details on:

- Data that will be generated within the project,
- Nature of the data,
- Format and expected size,
- And whether data, products, or services generated in the project will be made available.

The information presented below is a summary of the research outcome described.

The Inventory is regularly updated as the project progresses (e.g., to reflect operational needs arising from the use cases, new data sources relevant to the use cases or technical partners, new data sources upon request etc.) and is included in Annex II of the current version of the DMP

3.1 Purpose of the data collection/generation and relation to the objectives of THEIA

Data collection, generation and processing are integral to THEIA's objectives. The purpose is outlined as follows:

- Research data: Collected for deliverable creation across almost all WPs through desktop research, literature reviews, interviews and workshops, in the form of recordings, written notes and transcripts.
- End-user requirement data: Gathered in WP5 to identify and analyse end-user needs (dataset).
- Experimental Data: Encompasses methodologies for processing and integration across multiple WPs, with WPs 6 and 7 developing Earth Observation data methodologies, WP8 implementing ground data integration, WP9 establishing data federation and cybersecurity frameworks, WP10 advancing AI and machine learning techniques and WP11 focusing on THEIA platform integration.
- Demonstration data: Collected from pilot activities in WP12 (datasets).
- Business development data: Used for the development of the business plan in WP13 and WP14 (document).
- Geospatial data: To support the development of the THEIA components, data from various sources, including ground cameras, satellites, and other relevant sensors, will be used.

Most of these datasets will be openly shared with the scientific community, ensuring accessibility and contributing to broader research efforts. These datasets will be included in a later version of the DMP.

3.2 Types and formats of data that THEIA will generate/collect

In order to fulfil the purpose of the data collection/generation, the THEIA project will collect and generate the following types and formats of data:

Table 3. THEIA Types and formats of data

Data/Data Source	Data type	Data format	Data origin
Surveys, questionnaires, workshops data, validation data	Electronic document	Word document (.doc, .docx) Excel document (.xls/.xlsx) Pdf document	WP3, WP4, WP5, WP8, WP12, WP13, WP14
Deliverables	Electronic document	Word document (.doc, .docx) Excel document (.xls/.xlsx) Pdf document	All WPs
Video files	Electronic document	.mov, .mpeg, .avi, .mp4, etc.	WP5, WP7, WP10, WP12, WP13, WP14

Audio files	Electronic document	.mp3, .wav, etc.	WP5, WP7, WP12, WP13, WP14
Images	Electronic document	.jpg, .png, .gif, etc.	All WPs
Software	Source Code	Source Code	WP6, WP7, WP8, WP9, WP10, WP11
Signed documents (e.g. Consent forms, information sheets, attendance lists, Consortium Agreement, etc.)	Electronic document	Word document (.doc, .docx) Excel document (.xls/.xlsx) Pdf document	WP1, WP2, WP3, WP4, WP5, WP12, WP13, WP14
Presentations	Electronic document	Powerpoint document	All WPs
Geospatial data	Vector and Raster	Shapefile, JSON, GeoJSON, tiff, GeoTiff, JPG2000 Geo-located detection events in .json format, NetCDF	WP9, WP10, WP11

3.3 Origin of the data, reuse of existing ones and expected size of the data

To support the project deliverables, data are collected, evaluated and aggregated from end-users and external stakeholders engaged in the project activities, acquired through desktop research, workshops and surveys. Potentially existing data will be exploited in the form of published materials relating to CSS and serve as a reference in order to analyse operational or other incidents that may occur.

To support the development of the THEIA components, data from various sensors (cameras, satellites, etc.) will be used.

Information that are expected to be used as existing and/ or be generated from the project collected in the initial data survey within the consortium titled "*Inventory of Existing and New Data*".

The expected size of the data collected/generated/processed under the THEIA project ranges from kilobytes to gigabytes, even terabytes, depending on the period of retention of the data, the sensor data etc.

3.4 Information types in the THEIA project

According to project's GA the project involves sensitive information requiring limited dissemination due to security reasons and for this reason two types of deliverables are foreseen: Public and Sensitive.

3.4.1. Public Deliverables of THEIA

The following Table present THEIA deliverables, classified as PUBLIC, updated in accordance with the Amendment currently in progress.

Table 4. THEIA Public Deliverables

No	Title
D1.2	Risk identification Management & Quality assurance plan
D1.3	Initial Data Management Plan (DMP)
D1.4	Intermediate report on Risk Identification Management & Quality Assurance Plan
D1.5	Intermediate Data Management Plan (DMP)
D2.1	Final report on Risk Identification Management & Quality Assurance Plan
D2.2	Final version of the Data Management Plan (DMP)
D3.1	THEIA Ethics version 1 (1st period)
D3.2	THEIA Ethics version 2 (1st period)
D4.1	THEIA Ethics (2nd period)
D5.4	EO Data Processing Pathway Selection Guide
D6.1	Delivery of the tailored Very-High-Resolution Earth Observation and Space-based Video Data Tools
D7.3	Assessment of UAS and Terrestrial Sensing
D9.3	Implementation of National Constellation Data Federation and Optimized API Access
D11.1	Deployment and Integration of the THEIA Platform
D13.1	Website and Project Logo
D13.2	Dissemination and Communication Plan (Version 1/1st-period)
D13.3	Exploitation plan (Version 1)
D13.4	Dissemination and Communication Report (Version 1/1st-period results)
D13.5	Exploitation plan (Version 2)
D14.1	Dissemination and Communication Plan (Version 2/2nd-period)
D14.2	Exploitation plan (Version 3/2nd-period results)
D14.3	Dissemination and Communication Report (Version 2/2nd-period results)

3.4.2. Sensitive Deliverables of THEIA

In the Table below THEIA deliverables classified as deliverables with “Sensitive information with security recommendation” - SENSITIVE are presented and have been updated in accordance with the Amendment currently in progress.

Table 5. THEIA Sensitive Deliverables

No	Title
D1.1	MoM of Kick-off Meeting
D1.6	Initial Progress Report
D2.3	Progress Report
D5.1	Stakeholders engagement plan, CSS Gap Analysis Report and User Feedback Summary
D5.2	Enhanced Service Features
D5.3	Use Case Definition Report
D5.5	THEIA Architecture Concept
D6.2	Selection and optimisation of the methodology of appropriate space data
D7.1	Assessment of current and planned EO (Optical & SAR) micro and pico-satellites and RF satellites mission
D7.2	Comprehensive Assessment and Evaluation of Current and Planned EO/RF Missions Against Stakeholder Requirements and Identification of Gaps and Recommendations
D8.1	Delivery of the crowdsourcing tool
D8.2	Compilation and Analysis of Statistical, Economic, and Demographic Data for AOIs
D8.3	Delivery of high-velocity transnational data
D9.1	Delivery of the data models and cyber-secure data exchange framework
D9.2	Cyber-Secure Data Exchange Framework and Service for THEIA
D10.1	Advanced ML Techniques for Satellite Image Processing and Georeferencing
D10.2	AI-Based Processing and Georeferencing for UAS and Terrestrial Sensors
D10.3	Intelligence Gathering and Georeferencing from Open and Closed Sources
D10.4	Fusion and Geospatial AI (GeoAI) Module Development
D12.1	Report on THEIA Use Cases and Demonstration Requirements
D12.2	Technology Validation Preparatory Report
D12.3	Comprehensive Report on THEIA Demonstration Activities
D12.4	Performance and Impact Assessment Report
D15.1	OEI - Requirement No. 1

3.5 Usefulness of collected data

The project's data will benefit a wide range of stakeholders, including policymakers, research institutions and technology providers. It can support evidence-based decisions, drive research and innovation, improve Copernicus Security Services (civil security, law enforcement, and crisis management), and assist designated end users.

4. FAIR Data

This section provides an initial overview of the necessary measures that will be adopted by THEIA to ensure that the data collected or generated during the project comply with the FAIR principles:

- **Findability**: Includes identifiers, keywords, metadata standards and other best practices to enhance data discoverability and facilitate reuse by third parties.
- **Accessibility**: Presents the repository where data will be stored, access conditions (open access, access protocols and restrictions) and the availability of metadata.
- **Interoperability**: Discusses the vocabularies, standards, formats and methodologies that will be applied to ensure seamless data exchange, reuse and interoperability.
- **Reusability**: Describes the expected documentation, including methodology explanations, codebooks (if applicable), variables and other relevant details to support data reuse.

4.1 Making Data Findable

Data storage, processing, and sharing among project participants is facilitated through the dedicated project data-sharing platform/ repository (https://www.dropbox.com/work/THEIA%20project) where all THEIA team members have access, while interaction with the broader public will take place via the official project website and dedicated repository (Zenodo) as explained below.

A file naming system has been set in D1.2 and D1.4 for easy identification of files stored in the project data-sharing repository/ Dropbox. A common naming system has also been set up for data findability, as well as keywords in the datasets/deliverables.

THEIA will be fully compatible with the principle of "Findability" as identifiers, metadata and search keywords will be used for facilitating data findability.

Analytically, the publications will be made available adopting the Open Science principle as described above, and for this, a persistent identifier will be adopted (please see next Section). Metadata will be used which will include among others: title, data types/formats and software, etc. as described above. In the later DMP versions, metadata schemas and repositories will be detailed (e.g. CKAN or whatever will be used). For published articles, journals assign a Digital Object Identifier (DOI), a unique and permanent code that identifies the article. For conference contributions, bibliographic references will be provided.

Additionally, it would be examined the possibility of developing and sharing among THEIA partners a comprehensive Glossary in order to enhance the findability of (meta)data.

Furthermore, the project will incorporate geometadata standards such as ISO 19115 and formats such as GeoJSON, which are useful for ensuring interoperability and consistency across geospatial datasets. These standards will enable integration and usage of data by providing detailed metadata descriptions and a widely accepted format for geographic data.

4.2 Making Data Accessible

Research data generated in the THEIA project will comply with all requirements of the GA (i.e. Article 16, Article 17 etc.).

To facilitate and ensure maximum information sharing and accessibility with the consortium, a document repository has been made available to consortium partners from the project's start. Further details about this repository can be found in D1.2 “Risk Identification, Management & Quality Assurance Plan”.

Regarding THEIA's shareable data, a dedicated repository has been established on Zenodo.org to publish and host all relevant datasets. The Zenodo community serves as a centralised hub for data generated by EU projects aiming to enhance discoverability and accessibility for all interested parties. Zenodo enhances the discoverability of research outputs by assigning a Digital Object Identifier (DOI) to each upload.

The THEIA community, as illustrated in the next Figure has been established by GSH to facilitate open sharing of project data, including potential research outputs, enabling others to build upon the work produced. The repository, named “*THEIA eu project”*, features its logo in the top left corner and can be accessed via the following link:

https://zenodo.org/uploads/new?community=theia-project

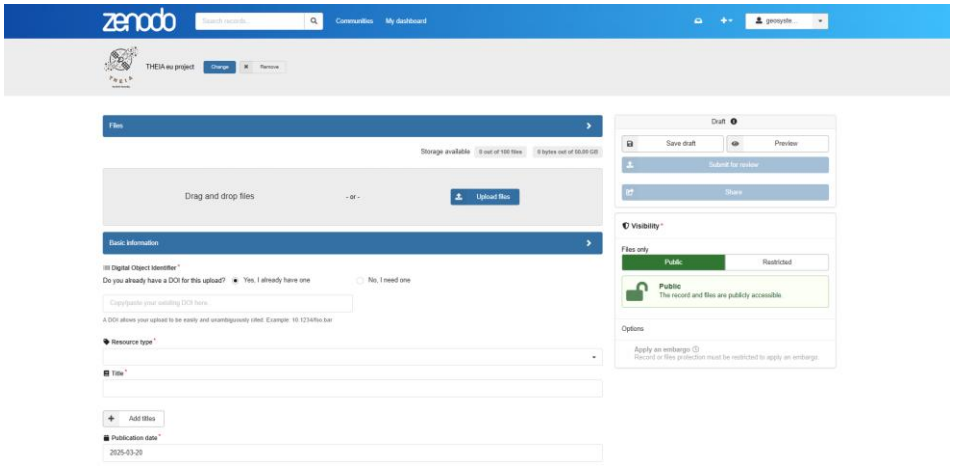

Figure 3. THEIA community in Zenodo

The THEIA consortium will ensure that all available data remains easily retrievable for all interested parties. Data will be formatted according to its type and will be made accessible along with relevant links to any necessary software tools for proper utilization.

All publications will be made available on Zenodo as already mentioned above.

Additionally, Public deliverables, dissemination material and information related to the project will be made available to the public through project's website (https://theia-project.eu).

Which THEIA data will be shareable will be determined throughout the project's progress while carefully considering IPR, legal requirements and ethical guidelines. Furthermore, interview data—including recordings, questionnaires etc.—collected from users will not be openly shared or published as primary data due to privacy and security considerations.

4.3 Making Data Interoperable

The concept of interoperability requires that both data and metadata be machine-readable and that a consistent terminology be applied. To achieve this, a standardised document template should be used to define the structure of the exchanged information.

In most cases, a common vocabulary will be adopted for all data types, ensuring consistency across different contexts. This standardised terminology will align with widely accepted conventions in the business creation ecosystem, eliminating any barriers to data interoperability and reuse.

4.4 Increase Data Re-use

The re-use of data, if necessary, will be limited to research purposes as specified by the license. Anonymous data may be used for scientific publications, but data cannot be copied or distributed and must be properly referenced when included in publications. The collected data will integrate information from multiple sources, each governed by its own policies.

To promote wide reusability, the use of Creative Commons licenses, with CC-BY as the default, will be encouraged for research articles. This license allows for copying, distribution and transmission of the work while ensuring that essential author rights remain protected.

Partners' initial plans include storing data for future use, sharing and exploiting key results, using the data for internal development, and making publication decisions at the Consortium level.

5. Allocation of Resources

The costs associated with making data FAIR have been allocated and covered by the THEIA project budget, ensuring that no additional expenses are anticipated. Dissemination materials, such as scientific publications, will be made available as previously mentioned on Zenodo and project's web site.

The Data Manager is generally in charge of data management and compliance for the THEIA project, along with the contact points and DPOs or persons responsible for data protection compliance of each organisation involved. Data management will be handled as part of WP1 (Project Management). Currently, GSH, as the project coordinator, along with contributions from all consortium partners, is responsible for data management in the THEIA project.

6. Data Security

For the THEIA project, data security remains highly relevant and is part of an ongoing process of continuous improvement. THEIA will take measures to ensure Confidentiality, Integrity and Availability by protecting data against unauthorised access, use and distribution. Access controls are also considered among the THEIA consortium.

All generated and collected data will be securely managed, in the duration of the THEIA project. The purpose is (a) to protect against unauthorised access and (b) ensure there is no loss or leak of information. Additionally, an incident response plan will be developed and delivered in a later version of the DMP to ensure effective action in the event of data breach or other security incidents. This plan will:

- outline procedures for identifying, reporting, and mitigating incidents
- designate roles and responsibilities within the team and
- include communication strategies to inform relevant stakeholders

The data collected and/or generated throughout the whole period of the project's execution will be held in the CDO module.

THEIA-system data are stored in:

- A database running on the CREOTECH VM

- A virtual filesystem of the CREOTECH VM

- An S3 bucket exposed by CREOTECH, and running on Amazon infrastructure.

Concerning the first one, recovery of the PostgreSQL data can be achieved by using a combination of a file-system backup together with WAL files. By this way, the database state is gotten from the file system, and the logs maintained in the WAL files are applied, beginning from the timestamp to which the file-system backup refers. In this way, the present state of the database is recreated.

Concerning the second one, the VM itself can be backed up using several tools available freely or in the market. This must be decided with CREOTHECH.

Concerning the third one, security is ensured due to the fact that backups are most probably part of the S3 Amazon subscription. For certainty, CREOTECH must be asked about this.

Finally, the project will establish a data retention and disposal policy to ensure that data is retained only for as long as necessary and securely disposed of once it is no longer needed. Additionally, the project will explore the adoption of a data classification scheme to categorize generated data (e.g. public, private, sensitive). While deliverables are already classified, certain

in-progress data may also require classification to ensure appropriate security measures are applied to each data type.

7. Legal framework and Guidelines

The protection of personal data is a fundamental right, ensuring that individuals are safeguarded against unlawful processing of their information

According to THEIA Deliverable D3.1 (p. 21) within the EU, primary legislation such as the Charter of Fundamental Rights and the TFEU lays the constitutional foundation for data protection. The GDPR[2] serves as the main secondary legislation regulating data processing.

With respect to all data processing activities of the project as they are described above in detail (Section 2.4.3), constant guidance will be provided by the Data Manager and the Ethical Manager appointed for the Project at central level along with the Data Protection Officer of each partner or the person responsible for data protection compliance (ParDPO) in collaboration with the appointed Ethics Advisor.

The THEIA Consortium commits to the protection of personal data processed during the lifetime of the research project and will implement the appropriate safeguards in order to be compliant with the GDPR provisions.

THEIA D3.1 "THEIA Ethics version 1 (1st period)" already submitted in M3 (February 2025) provides a preliminary examination of ethical considerations, data protection and privacy, ensuring adherence to key legal and ethical principles, as well as the necessary legal documentation.

THEIA D3.2 "THEIA D3.2 Ethics version 2 (1st period)", submitted in M15 (February 2026) deals with specific Ethics developments in the project. Thanks to the engagement of the technical partners in the project in Ethics by design initiatives, they are spotted, analysed and treated with the utmost care, and developed in respect with the right to privacy and data protection.

Besides, as mentioned above a JCA is under preparation for the project which will be signed by all consortium partners.

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119, 4.5.2016, p. 1–88.

8. Personal Data Management

The protection of personal data is a fundamental right, ensuring that individuals are safeguarded against unlawful processing of their information. At the European level, legal protections for data privacy derive from the provisions already exposed in the previous sections on the European Convention on Human Rights (ECHR). Within the European Union, from the EU Charter of Fundamental Rights respectively, which explicitly establish the right of individuals to the protection of their personal data.

While data protection and privacy are closely related, they are recognized as distinct rights in legal frameworks worldwide. Data protection emerged from the right to privacy, and both serve as essential mechanisms to uphold fundamental values such as freedom of expression and freedom of assembly.

The key distinction between these rights lies in their legal scope and implementation. The right to privacy is broadly framed as a prohibition on interference, with exceptions permitted under certain public interest justifications. In contrast, data protection is a proactive and structured right, requiring a system of checks and balances to safeguard individuals when their personal data is processed. This system ensures compliance through independent oversight and the enforcement of data subject rights.

Within the EU, primary legislation such as the Charter of Fundamental Rights and the TFEU lays the constitutional foundation for data protection. The GDPR serves as the main secondary legislation regulating data processing.

This section will detail THEIA activities in relation to the content of the THEIA D3.1 and D3.2 deliverables, specifically addressing: (a) the types of personal data and special categories of data utilised, (b) the processing of personal data and its lawful basis and (c) compliance with the data minimisation principle in these processing activities.

8.1 Core definitions under the GDPR

To ensure legal clarity in the THEIA project, it is crucial to reference the definition key data protection concepts, as they determine the applicable legal framework. The GDPR (Articles 4, 9, 22 and Recital 51) provides the following definitions, as listed in Deliverable D3.1 “THEIA Ethics version 1 (1st period)” section 4.2.1. Core definitions.

8.2 Processing of personal data

The processing of personal data by THEIA consortium partners will be fully compliant with Article 5 of the GDPR, as explicitly outlined in Article 15 of the project's GA.

According to this:

The beneficiaries must process personal data under the Agreement in compliance with the applicable EU, international and national law on data protection (in particular Regulation (EU) 2016/679, Regulation (EU) 2018/1725).

They must ensure that personal data is:

- processed lawfully, fairly and in a transparent manner in relation to the data subjects
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes
- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed
- accurate and, where necessary, kept up to date
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data is processed and
- processed in a manner that ensures appropriate security of the data.

The beneficiaries may grant their personnel access to personal data only if it is strictly necessary for implementing, managing and monitoring the Agreement. The beneficiaries must ensure that the personnel is under a confidentiality obligation.

The beneficiaries must inform the persons whose data are transferred to the Granting Authority and provide them with the Portal Privacy Statement.

Regarding the lawfulness of processing, it is the responsibility of each partner handling personal data to clearly establish the legal basis for processing.

As part of the creation of the JCA for the personal data of the THEIA project, a table will be prepared outlining the legal basis for collecting and processing personal data. For research activities, this is primarily based on Article 6(1)(b) GDPR, as it is necessary for the performance of a contract to which the data subject is a party, along with the consent of participants under Article 6(1)(a) GDPR and Recital 39 GDPR. For dissemination activities, the legal basis is the legitimate interest of the partners, in accordance with Article 6(1)(f) GDPR. This table will provide a task-by-task breakdown of the legal basis for processing personal data and will be included as an annex to the JCA.

An example of the table to be filled in by partners, providing information on Personal data categories, Data subject categories and Processing Purposes per WP and Task, is provided below.

Table 6. THEIA Table for the Personal data processing details

WP/ Task	Personal data types	Individuals' categories	Reasons for Processing	Legal Justification for processing	Transfer of personal data outside the EU or EEA
T1.1-T1.3	Name, surname, email address, phone number, role within the organisation	Consortium Partners Advisory Board	For the purpose of communication, cooperation and information sharing between Consortium Partners. For the purpose of communication.	Article 6(1)(b): performance of a contract: Necessary for project coordination, technical development, or validation under the GA. Art. 6(1)(b): Performance of contract for Advisory Board engagement and communication.	No
T2.1-T2.2	Name, surname, email address, phone number, role within the organisation	Consortium Partners Advisory Board	For the purpose of communication, cooperation and information sharing between Consortium Partners. For the purpose of communication.	Article 6(1)(b): performance of a contract: Necessary for project coordination, technical development, or validation under the GA. Art. 6(1)(b): Performance of contract for Advisory Board engagement and communication.	No
WP3	Identification data (e.g. name, data of birth, age, gender, address, email, phone number) Occupation, employment or professional affiliation	Consortium Partners Ethics Advisor	Generally, personal data of the researchers in the project is processed, such as contact emails, to reach them and pose questions about ethical aspects of the project. Incidentally, access may also be possible to personal data processed by the rest of the partners in the project through the project repository, reports and deliverables of the project.	Article 6(1)(b) GDPR, as it is necessary for the performance of a contract to which the data subject is a party.	No
WP4	Identification data (e.g. name, data of birth, age, gender,	Consortium Partners Ethics Advisor	Generally, personal data of the researchers in the project is processed,	Article 6(1)(b) GDPR, as it is necessary for the performance of a contract	No

	address, email, phone number) Occupation, employment or professional affiliation		such as contact emails, to reach them and pose questions about ethical aspects of the project. Incidentally, access may also be possible to personal data processed by the rest of the partners in the project through the project repository, reports and deliverables of the project.	to which the data subject is a party.	
T5.1	Name, surname, email address, phone number, role within the organisation	Authorities that cooperate with CSS	GSH will engage external stakeholders through interviews or questionnaires to support user requirements and gap analysis of existing CSS. The goal is to capture requirements for a technical system aligned with end-user needs for THEIA services. Interviews will be conducted online and recorded only with the interviewees' consent.	Art. 6(1)(a): Consent: Voluntary participation based on informed consent for research or testing activities	No
T6.1, T6.3	First and last name, email address, and position within the organization.	Consortium Partners and satellite data providers involved in the scope of WP6.	Contact is required to obtain updated mission information. Only professional contact details that were publicly disclosed by the individuals or their organisations are used.	Processing is based on Art. 6(1)(f) (legitimate interests) and Art. 6(1)(b) GDPR (performance of a contract). It is necessary for updating mission information and for coordination and technical activities under the GA. Only minimal professional contact details are processed; communication is professional and voluntary.	No
T6.2	Name, surname, email address	Consortium Partners	For the purpose of communication,	Article 6(1)(b): performance of a contract: Necessary for	No

			cooperation and information sharing between Consortium Partners.	project coordination, technical development, or validation under the GA.	
T71.-T7.3	Name, surname, email address, role within the organisation	Different satellite data providers within the context of WP7 scope	The project needs to contact data providers for updated mission information. Their professional contact details were made publicly available by them / their companies through websites or conference presentations.	Article 6(1)(f) processing necessary for legitimate interests pursued by the controller: data processing is minimal and the purpose of the communication is professional and reasonable (try to verify / update the already publicly available information about the respective mission). Their involvement is voluntary.	No
T7.4	Aerial imagery and video data with no identifiable individuals	Not applicable / No identifiable individuals	Data acquisition and system validation for UAS-based sensing within THEIA	Article 6(1)(f) GDPR – legitimate interest for research and technical validation activities	No
T8.1	User generated content (posts, comments), personal data voluntarily disclosed within such content (names, contact details, locations), public timestamps and content metadata (reaction counts), platform account identifiers (e.g. username, user ID) are accessed but not stored or processed	Online Identities (The representation of a person, organisation, or entity within the digital realm)	User identifiers are only processed for anonymization. Data disclosed within content is used for producing analytic indicators. All personal data processed by the Crowdsourcing tool will comply with applicable data protection regulations (e.g. GDPR).	Article 9(2)(j) the processing is strictly for scientific or statistical research with in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.	No
T8.2–T8.3	Name, surname, email address, phone number, role within the organisation	Consortium Partners Advisory Board	For the purpose of communication, cooperation and information sharing between Consortium Partners. For the purpose of communication.	Article 6(1)(b): performance of a contract: Necessary for project coordination, technical development, or validation under the GA. Art. 6(1)(b): Performance of contract for Advisory Board engagement and communication.	No

T9.1	Name, surname, email address	Consortium Partners	For the purpose of communication, cooperation and information sharing between Consortium Partners.	Article 6(1)(b): performance of a contract: Necessary for project coordination, technical development, or validation under the GA.	No
T9.2	Name, surname, email address	Consortium Partners	For the purpose of communication, cooperation and information sharing between Consortium Partners.	Article 6(1)(b): performance of a contract: Necessary for project coordination, technical development, or validation under the GA.	No
T9.3	Name, surname, email address	Consortium Partners	For the purpose of communication, cooperation and information sharing between Consortium Partners.	Article 6(1)(b): performance of a contract: Necessary for project coordination, technical development, or validation under the GA.	No
T9.4	Name, surname, email address	Consortium Partners & different national datasets owners within the context of WP9 scope	For the purpose of communication, cooperation and information sharing between Consortium Partners and in order to contact selected data providers for national data access & API information using publicly available professional contact details	Article 6(1)(b): performance of a contract: Necessary for project coordination, technical development, or validation under the GA. Data providers involvement on voluntary basis	No
T10.1	Videos / Images	Consortium Partners & Personnel employed by Consortium Partners.	Image data have and will be collected during the training phase of the computer-vision models. Image data will be collected during dedicated recording sessions (e.g. pilot demonstrations) to perform computer-vision inference (e.g. vehicle and vessel detection/tracking).	Art. 6(1)(b): Performance of a contract: Necessary for project coordination, technical development, or validation under the GA.	No
T10.2	Videos / Images of persons	Consortium Partners & Personnel employed	mage data was collected during dedicated recording sessions (e.g. pilot	Art. 6(1)(b): Performance of a contract: Necessary for project coordination,	No

		by Consortium Partners. If public open-source RGB datasets are used for model training or performance evaluations, other persons can be visible in the images.	demonstrations) to develop computer vision algorithms (e.g. vehicle and person detection/tracking) and to support processing within the final THEIA system.	technical development, or validation under the GA.	
T10.3	Name, surname, email address, phone number, role within the organisation	Consortium Partners Advisory Board	For the purpose of communication, cooperation and information sharing between Consortium Partners. For the purpose of communication.	Article 6(1)(b): performance of a contract: Necessary for project coordination, technical development, or validation under the GA. Art. 6(1)(b): Performance of contract for Advisory Board engagement and communication.	No
T10.4	Videos / Images	Consortium Partners & Personnel employed by Consortium Partners	Image data will be recorded in dedicated recording sessions (e.g. in the Pilot Demonstrations) for developing computer vision algorithms (detection/tracking of vehicles, persons etc) and also processing image data in the final BorderForce system.	Art. 6(1)(b): Performance of a contract: Necessary for project coordination, technical development, or validation under the GA.	No
T11.1	Videos / Images / web-scraped data ped data	Consortium Partners & Personnel employed by Consortium Partners	Video, image, RF, and json data will be collected and processed during the (a) development phase of the THEIA system and (b) the demonstration phase of the THEIA system. The outputs of this processing will be presented to the end-users.	Art. 6(1)(b): Performance of a contract Necessary for project coordination, technical development, or validation under the GA.	No
T11.2	Videos / Images / web-scraped data	Consortium Partners & Personnel employed	Video,image, RF, and json data will be collected and	Art. 6(1)(b): Performance of a contract	No

		by Consortium Partners	processed during the (a) development phase of the THEIA system and (b) the demonstration phase of the THEIA system. The outputs of this processing will be presented to the end-users.	Necessary for project coordination, technical development, or validation under the GA.	
T12.2	Name, surname, email address, phone number, role within the organisation, Videos / Images	Consortium Partners & Personnel employed by Consortium Partners	Personal data of project partners may be processed for the development of a roadmap describing the integration process and the partners' responsibilities.	Art. 6(1)(b): Performance of a contract: Necessary for project coordination, technical development, or validation under the GA.	No
T12.3	Name, surname, email address, phone number, role within the organisation. Country of organisation in which the end user is employed, their organisation, and organisational unit. Analysed Test data (Name, surname, e-mail, organisation, images, videos)	Consortium Partners & Personnel employed by Consortium Partners	Partner information of those participating in the testing site simulating the operational environment from targeted real border locations.	Art. 6(1)(b): Necessary for system testing and validation activities under the GA.	No
T12.4	Data available from T12.3: Partner information (i.e. Name, surname, email address, phone number, role within the organization, images, videos) Analysed Test data (Name, surname,	Consortium Partners & Personnel employed by Consortium Partners	Video and image data from dedicated recording sessions (e.g., pilot demonstrations) will be processed for presentation to end-users during the validation and evaluation sessions.	Art. 6(1)(b): Necessary for validation tasks under the GA.	No

	e-mail, organisation, images, videos				
T13.1	Names, contact details, images	Consortium Partners Target audiences, stakeholders and end users	Personal of project partners will be included where sharing the activities and outcomes, as well those included in content created on demonstrations. Personal data of target audiences, stakeholders and end users to promote the project's outcomes will be involved, specifically in relation to a mailing list.	Art. 6(1)(b): Performance of a contract Necessary for dissemination, communication, and stakeholder engagement, with opt-out options	No
T13.2	Names and email addresses,	Consortium Partners	To establish commercial partnerships.	Art. 6(1)(b): Performance of a contract Necessary for responsible communication, and dissemination.	No
T13.3	Names and email addresses,	Consortium Partners	To establish commercial partnerships.	Art. 6(1)(b): Performance of a contract Necessary for responsible communication, and dissemination.	No
T14.1	Names, contact details, images	Consortium Partners Target audiences, stakeholders and end users	Personal of Consortium partners will be included where sharing the activities and outcomes, as well those included in content created on demonstrations. Personal data of target audiences, stakeholders and end users to promote the project's outcomes will be involved, specifically in relation to a mailing list.	Art. 6(1)(b): Performance of a contract Necessary for dissemination, communication, and stakeholder engagement, with opt-out options	No

T14.2	Names and email addresses,	Consortium Partners	To establish commercial partnerships.	Art. 6(1)(b): Performance of a contract Necessary for responsible communication, and dissemination.	No

9. IPR

9.1 General Principles

The consortium recognises that effective knowledge and IPR management is essential for fostering smooth collaboration among members and ensuring the successful exploitation and sustainability of THEIA outcomes both during and after the project. By safeguarding partners' interests through structured knowledge management and protection measures, potential information bottlenecks related to confidentiality will be mitigated. This, in turn, will enhance market visibility and maximise the successful implementation of project results.

9.2 IPR Framework and Governance

The management of intellectual property rights in THEIA is governed by Article 16 of the GA and the detailed provisions in Annex 5, supplemented by the Consortium Agreement-CA (Chapters 8 and 9). The GA defines the ownership of results (foreground IP), access rights to background and results, obligations for protection and exploitation, and conditions for transfer or licensing of IP, including specific rules on strategic autonomy, security-sensitive information and protection obligations.

9.3 Ownership and Access Rights

Results generated by the project are owned by the beneficiary(ies) that generate them, in accordance with GA Article 16.2. Joint ownership applies where results are jointly generated and cannot be separated. Partners grant each other royalty-free access to background and results as needed for implementing project tasks, and fair and reasonable access for exploitation purposes, subject to the terms of the CA and GA Annex 5. Access rights shall be requested in writing and granted in accordance with the timelines and conditions set out in the CA. Prior to any dissemination, beneficiaries will assess whether protection of results is appropriate and feasible, in order to avoid jeopardising potential intellectual property rights.

9.4 Protection and Exploitation of Results

The consortium commits to adequately protecting results where commercial or industrial exploitation is feasible, in line with GA obligations. Protection mechanisms include copyright, patents, trade secrets, and open-source licensing, as appropriate to each result. The detailed exploitation and IPR management strategy, including identification of Key Exploitable Results

(KERs), ownership mapping, and protection routes, is documented in the project's Exploitation Plan (D13.2, D13.5, and D14.2).

9.5 IPR and Open Science Compliance

For results not subject to commercial exploitation or confidentiality requirements, the project will comply with Open Science principles as defined in GA Annex 5, including open access to publications (CC BY license) and research data management aligned with FAIR principles (see Sections 2.4.1 and 4 of this DMP).

9.6 IPR Identification and Management Activities

During the first project period (M1-M15), Task T13.3 - "Identification of IPR Issues and Patentable Content," focused on identifying key IPR considerations. In the second period, Task T14.4 - "IPR Management and Patenting Activities", will address IPR protection and management. Both tasks are led by THEIA's ethical partner, MPL. Preliminary IPR assessments are reflected in D13.2 and D13.5.

9.7 Granting Authority Rights and Strategic Autonomy

The consortium acknowledges the Granting Authority's rights under GA Article 16.3 to use non-sensitive project outputs for policy and communication purposes. The consortium will ensure that results remain free from control by non-eligible countries, in compliance with strategic autonomy requirements under GA Annex 5.

10. Ethics

WP3 – "Ethics 1 - 1st Period" and WP4 – "Ethics 1 – 2nd Period", as well as the associated deliverables D3.1 "THEIA Ethics Version 1 (1st period)", D3.2 "THEIA Ethics Version 2 (1st period)", and D4.1 "THEIA Ethics (2nd period)", address the ethics requirements that the THEIA project must comply with in relation to its objectives, methods, processes, tasks and results. The ethics partner, MPL, is responsible for overseeing these aspects. These ethics requirements primarily focus on the processing of personal data and, therefore, data protection.

The work on Ethics within THEIA focuses on ensuring compliance with relevant laws and regulations, particularly concerning ethics, data protection and privacy. This includes addressing ethical issues, adhering to the GDPR and other relevant regulations and preparing the necessary legal documentation. Privacy and ethics approvals for each pilot will be obtained from the relevant authorities, with particular attention to gender, inclusion and social norms. Finally, the project aims to establish clear guidelines within the space sector and provide constructive evaluations to support comprehensive space segment policies.

Project THEIA aims to leverage geospatial data and AI to analyze population displacement caused by conflicts and other crises. It will develop digital systems to process large-scale personal data from geospatial sources and social media, integrating AI models to study migration flows. Given the ethical concerns related to privacy and potential biases, safeguards have been identified during the project's ethics self-assessment. According to the evaluation process, additional measures are necessary to mitigate biases in AI algorithms due to the extensive use of GeoAI and machine learning. For this reason, Deliverable D15.1 "OEI - Requirement No. 1" was added and submitted in the first month of the project's implementation which expands on those safeguards, detailing when and how ethical issues will be addressed, following the 2017 European Code of Conduct for Research Integrity.

11. Conclusions

This deliverable D1.5 constitutes the intermediate DMP of THEIA at M15 of the project. It serves as an updated reference for data management practices, describing the core principles applied within the project, including Open Science and FAIR principles. It also presents an updated overview of security and ethical aspects related to data handling.

These principles are implemented through the use of established tools and standards, such as Open Science practices and the Zenodo repository, ensuring that THEIA outputs (e.g. open data and open-access publications) are properly preserved and remain accessible beyond the project's lifetime.

Furthermore, this intermediate DMP incorporates consolidated feedback from all partners following a consortium-wide data survey. The survey was conducted using the Excel file *"Inventory of Existing and New Data"*, through which partners identified the types of data to be collected and generated, together with the justification for their collection.

The document also addresses the management of personal data, in alignment with other project deliverables and activities.

Finally, the THEIA DMP is a living document that will be continuously updated throughout the project, as data management is an ongoing activity relevant to most project WPs.

ANNEX I

Template for Horizon Europe DMP

EU Grants: Data management plan (HE):V1.1 – 01.04.2022

DATA MANAGEMENT PLAN

(To be filled in and uploaded as deliverable in the Portal Grant Management System, at the due date foreseen in the system (and regularly updated).

⚠ *The template is recommended but not mandatory. If you do not use it, please make however sure that you comply with the research data management requirements under Article 17 of the Grant Agreement.)*

PROJECT	
Project number:	[project number]
Project acronym:	[acronym]
Project name:	[project title]

DATA MANAGEMENT PLAN	
Date:	[dd/mm/yyyy]
Version:	[DMP version]

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

What types and formats of data will the project generate or re-use?

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

What is the expected size of the data that you intend to generate or re-use?

What is the origin/provenance of the data, either generated or re-used?

To whom might your data be useful ('data utility'), outside your project?

2. FAIR data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Will metadata be offered in such a way that it can be harvested and indexed?

2.2. Making data accessible

Repository:

Will the data be deposited in a trusted repository?

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Will the data be accessible through a free and standardized access protocol?

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

How will the identity of the person accessing the data be ascertained?

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

2.3. Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Will your data include qualified references1 to other data (e.g. other data from your project, or datasets from previous research)?

2.4. Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

---

[1] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)*

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Will the provenance of the data be thoroughly documented using the appropriate standards?

Describe all relevant data quality assurance processes.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

3. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

Who will be responsible for data management in your project?

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

5. Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Will the data be safely stored in trusted repositories for long term preservation and curation?

6. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

7. Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.0	05.05.2021	Initial version (new MFF).
1.1	01.04.2022	Reformatted to align with other deliverables templates.

ANNEX II

Inventory of Existing and New Datasets

Existing Datasets per WP and per Task

Task/ Partner	Existing data acquired from public/open sources or elsewhere that will be used	Type of data and data sources for these data	Kind of data	Format of data (e.g. filename extension) & Expected size of the data (estimated or known)	How will data be made available in the project internally?	Will these data be made available to external actors?	Will any personal data be used for the purposes of the project?
Clarifications/ examples/ possible answers:	*Public databases, Wikipedia, media website, social media, responses to polls and surveys, EO and non-EO data, Copernicus open datasets etc.*	*Interviews, surveys, measurements, numerical models, administrative records, etc.* *Google search, official companies' websites, national portals, Copernicus Data Space Ecosystem, end-users and stakeholders*	*Numeric (databases, spreadsheets), text (documents), image, audio, video, software, and/or mixed media*	*Usually how the data is encoded for storage, often reflected by the filename extension (.pdf, .xls, .doc, .txt, or .rdf).*	*Internal repository*	*Public deliverables, with consent forms*	*Yes/ No*
T1.1/ GSH	EC produced material for Horizon Europe projects. Legal and financial data of consortium partners.	Reports, Administrative material. EC official website	Text, Images	.pdf, .doc	Internal repository	Some of them	Yes. Personal data of individual partners e.g. Contact details.
T1.2/ GSH	EC produced material for Horizon Europe projects.	Reports, Administrative material. EC official website	Text, Images	.pdf, .doc	Internal repository	Some of them	Yes. Personal data of individual partners e.g. Contact details.
T1.3/ GSH	EC produced material for Horizon Europe projects.	Reports, Administrative material. EC official website	Text, Images	.pdf, .doc	Internal repository	Some of them	No
T2.1/ GSH	EC produced material for Horizon Europe projects. Legal and financial data of consortium partners.	Reports, Administrative material. EC official website	Text, Images	.pdf, .doc	Internal repository	Some of them	Yes. Personal data of individual partners e.g. Contact details.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under GA 101190051. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

T2.2/ GSH	EC produced material for Horizon Europe projects.	Reports, Administrative material. EC official website	Text, Images	.pdf, .doc	Internal repository	Some of them	Yes. Personal data of individual partners e.g. Contact details.
T3.1/ MPL	Benchmark legal analysis of the applicable legislation to the research in the project.The report is based on publicly available legal informationwas already submitted.	Text, public legal information.	Text	.doc	Internal repository	Public deliverables	The content is a written report without any personal data in it, except from the names of the authors and reviewers.
T4.1/ MPL	Benchmark legal analysis of the applicable legislation to the research in the project.The report is based on publicly available legal informationwas already submitted.	Text, public legal information.	Text	.doc	Internal repository	Public deliverables	The content is a written report without any personal data in it, except from the names of the authors and reviewers.
T5.1/ GSH	1.Documents about Copernicus Security Services, 2. List of contacts with authorites cooperating with CMS (EMSA), CBS (FRONTEX), 3. Images and video about CSS, 4. Contact information of agencies that cooperating with SESA (SATCEN)	Online documents, websites. Official companies' websites, EU websites, Copernicus Services website, Google search	1. Text (documents), 2. image, 3. video, 4. Contact details (Emails)	.pdf, .txt	Internal repository	Public deliverables	No
T5.2/ AIT	X	X	X	X	X	X	X

T5.3/GSH	Data No data were used during the design of the architecture. Sources website examples https://martinfowler.com/ Elsevier excerpts book examples FOUNDATIONS OF SOFTWARE TESTING: Graham, Veenendaal et al SOFTWARE ENGINEERING, Ninth Edition: Ian Sommerville Guide to the Software Engineering Body of Knowledge, version 3.0 Elements of Reusable Object-Oriented Software: Gamma, Helm, et al	websites, books	documents	.pdf, html	X	X	X
T5.4/ C3I	X	X	X	X	X	X	X
T6.1/ OHB DS	Public sample data from Very High Resolution (VHR) Earth Observation resources, including imagery from Copernicus Contributing Missions and commercial satellite constellations.	Data include optical and SAR imagery. Sources are primarily from ESA's Earth Observation portal and the Copernicus Data Space Ecosystem. E.g. sample data provided by ESA: https://earth.esa.int/eogateway/news/sample-data-products-from-esas-third-party-missions	Image, geospatial	mostly .tiff. Approximately 100GB.	Internal repository	Already publically available.	No
	Scientific Papers and Research Articles: Gathering insights and findings which could provide valuable context .	Scientific journals	Text-based documents	.pdf	Internal repository	Already publically available.	No

T6.2/ SPACE-SI	Own satellite video and images	Own satellite video and images	Video, images	.mp4, .tif	Internal repository	No	No
T6.3/ OHB DS	Public sample data from Very High Resolution (VHR) Earth Observation resources, including imagery from Copernicus Contributing Missions and commercial satellite constellations.	Data include optical and SAR imagery. Sources are primarily from ESA's Earth Observation portal and the Copernicus Data Space Ecosystem. E.g. sample data provided by ESA: https://earth.esa.int/eogateway/news/sample-data-products-from-esas-third-party-missions	Image, geospatial	mostly .tiff. Approximately 100GB.	Internal repository	Already publically available.	No
	Scientific Papers and Research Articles: Gathering insights and findings which could provide valuable context .	Scientific journals	Text-based documents	.pdf	Internal repository	Already publically available.	No
T7.1/ LXS	We will collect data from varios EO databases, Space Agencies, satellite manufacturers, and scientific papers. Key sources include: - **Websites**: info gathered from websites related to EO and satellite missions.	Measurements (like technical specifications of different satellite missions), textual descriptions (missions documentation), etc.	The project will primarily use text-based and numeric data, including: - Spreadsheets with satellite missions information - Documents, or web documents, describing	The data will primarily consist on: - **Text Documents**: Word Documents (.doc, .docx): Expected size per document: 1-10 MB. PDF Documents (.pdf): Expected size per	Internal repository (in THEIA's Dropbox)	In principle is only to be shared with the Commission, but not expected to make it available with external actors, unless it would be explicited	No, all the info contained in these documents is supposed to be public and not having any personal data.

	- **Databases**: databases retrieved from EO databases and repositories - **Scientific papers:** relevant research articles accessed through different platforms like Springer Nature and ESA EO Science for Society. Additionally, we will leverage internal knowledge and previous project data. All collected data, will be properly documented, listed and stored, to ensure integrity and accessibility.	Google search, official companies' websites, national portals, Copernicus Data Space Ecosystem, end-users and stakeholders	satellite missions characteristics.	document: 1-5 MB. - **Spreadsheets**: Excel Files (.xls, .xlsx): Expected size per file: 1-100 MB (in case of extensive catalogs or datasets). - **Compiled Links:** Word or PDF Documents containing links: Expected size per document: 1-2 MB. Most of the data will be website-based, compiled links or relevant info into Word or PDF documents, or included in Excel spreadsheets.		required - not considered.	
T7.2/ LXS	We will collect data from varios EO databases, Space Agencies, satellite manufacturers, and scientific papers. Key sources include: - **Websites**: infor gathered from websites related to EO and satellite missions.	Measurements (like technical specifications of different satellite missions), textual descriptions (missions documentation), etc.	The project will primarily use text-based and numeric data, including: - Spreadsheets with satellite missions information - Documents, or web documents, describing	The data will primarily consist on: - **Text Documents**: Word Documents (.doc, .docx): Expected size per document: 1-10 MB. PDF Documents (.pdf): Expected size per	Internal repository (in THEIA's Dropbox)	In principle is only to be shared with the Commission, but not expected to make it available with external actors, unless it would be explicited	No, all the info contained in these documents is supposed to be public and not having any personal data.

	- **Databases**: databases retrieved from EO databases and repositories - **Scientific papers:** relevant research articles accessed through different platforms like Springer Nature and ESA EO Science for Society. Additionally, we will leverage internal knowledge and previous project data. All collected data, will be properly documented, listed and stored, to ensure integrity and accessibility.	National Space Agencies, EO databases (e.g. EO portal), satellite manufacturarers (e.g. ICEYE), Copernicus Data Space Ecosystem, Google searchs for supplementary information, etc.	satellite missions characteristics.	document: 1-5 MB. - **Spreadsheets**: Excel Files (.xls, .xlsx): Expected size per file: 1-100 MB (in case of extensive catalogs or datasets). - **Compiled Links:** Word or PDF Documents containing links: Expected size per document: 1-2 MB. Most of the data will be website-based, compiled links or relevant info into Word or PDF documents, or included in Excel spreadsheets.		required - not considered.	
T7.3/ LXS	Relevant data will provide from Task 7.1 and 7.2 and from WP5 (Task 5.1). It will be data generated during the project (therefore included in the New Datasets tab) However, if further research is required, additional data may be sourced from sources such as EO satellite missions, EO databases, etc.	Measurements (like technical specifications of different satellite missions), textual descriptions (missions documentation, user requirements provided by WP05), etc.	The project will primarily use text-based data, including: - **Websites:** infor gathered from websites related to EO and satellite missions. - **Databases**: databases retrieved from EO databases and repositories	The data will primarily consist on: - **Text Documents (probably for the user requirements deliverable)**: Word Documents (.doc, .docx): Expected size per document: 1-10 MB. PDF Documents (.pdf): Expected size per document: 1-5 MB. - **Spreadsheets (from new research, from previois Task (7.1/7.2/5.1)**: Excel Files (.xls, .xlsx): Expected size per file: 1-10 MB.	Internal repository (in THEIA's Dropbox)	In principle is only to be shared with the Commission, but not expected to make it available with external actors, unless it would be explicited required - not considered.	No, all the info contained in these documents is supposed to be public and not having any personal data.

		Outcomes from Task 7.1, Task 7.2 and Task 5.1 from the THEIA project, National Space Agencies, EO databases (e.g. EO portal), satellite manufacturarers (e.g. ICEYE), Copernicus Data Space Ecosystem, Google searchs for supplementary information, etc.		- **Compiled Links:** Word or PDF Documents containing links: Expected size per document: 1-2 MB.			
T7.4/ C3I	Public databases, data from previous projects, sensor log files, Publicly available traffic data from government agencies, AI model training datasets from open-source repositories.	Images, videos, sensor data, JSON/CSV files with measurements, Traffic camera footage, vehicle movement logs, Sensor logs from robotic platforms	Images, videos, numerical sensor data, Video feeds, tabular traffic statistics, Numerical data, diagnostic reports	.tiff, .mp4, .csv, .json, .bin, jpeg, .pdf 1GB–100GB - Depends on the mission and the file. For example: One flight of 5-10 minutes has 15-30MB. If the mission is longer, the data can reach up to 10-20GB	Shared via secure project cloud, Accessible through shared repository, Shared through internal data server,	Possibly, under strict agreements	It depends. For example if the UAV camera captures faces, license plates or any other personal data, then must follow GDPR processing rules
T8.1 / WTOC	Google portal, X	Types of data: text, dates, social media metrics Data sources: X API, Google search, Wikipedia	text, numerical, datetime	.csv, .json Expected size: < 1GB	Internal repository	Data will be available to partners for processing.	No
T8.2 / ED	Public databases publicly available on the Web	Official and institutional data	Geospatial data extracted from spatial databases or data types that can be georeferenced e.g. csv with coordinates	Vector data (shapefiles, geopackages etc.), Raster data (several image formats e.g. geotif) and spreadsheet-like data (e.g. csv, xlsx). Size is unknown as the final Areas of Interest (AoIs)	Through spatial databse (dockerized). On top of this there will be APIs providing JSON data	Yes. This is publicly available information	No

				have not been defined yet.			
T8.3 / ED	Public databases publicly available on the Web. Social media APIs.	Companies Websites and Open Sources openly available on the Web	Geospatial data extracted from spatial databases or data types that can be georeferenced e.g. csv with coordinates	Vector data (shapefiles, geopackages etc.), Spreadsheet-like data (e.g. csv, xlsx). Size is unknown as the final Areas of Interest (AoIs) have not been defined yet.	Through spatial databse (dockerized). On top of this there will be APIs providing JSON data	Yes. This is publicly available information	No
T9.1/ ICCS	Copernicus Data Space Ecosystem, ESA Open Access Hub (Sentinel-2 L2A), USGS Earth Explorer (Landsat 8)	Multispectral satellite imagery from Sentinel-2 L2A and Landsat 8	Optical multispectral imagery	Sentinel-2 L2A: JPEG2000 (.jp2), GeoTIFF; Landsat 8: GeoTIFF. Several GBs per scene.	Internal repository, cloud storage, or project-specific data platforms	Yes, publicly available through open data portals	No
T9.1/ SPACE-SI	Own satellite video and images	Own satellite video and images	Video, images	.mp4, .tif	Internal repository	No	No
T9.2/ C3I	Copernicus/CREODIAS EO products + CREODIAS catalogue/metadata (for testing/access/exchange)	EO satellite data accessed through CREODIAS (Copernicus Data Space Ecosystem) + platform APIs/metadata	Geospatial raster EO products + metadata	EO product formats (e.g., Sentinel products/GeoTIFF/NetCDF as provided) + JSON metadata via API; size TB-scale depending on AOI/time	Internal repository set up on the Creodias Virtual Machine	Only to authorised users (not public) via secure distribution/access control	No
T9.3/ CREO	CREODIAS platform (part of Copernicus Data Space Ecosystem), Copernicus Program datasets and services, DIAS Front Office cloud infrastructure	Satellite imagery, geospatial datasets, Copernicus services data, cloud resource usage metrics	Numeric (geospatial data, satellite imagery), metadata	.tiff, .nc, .csv, .json, .xml. There is resource of 82PB satellite data mapped to the machine.	Internal repository set up on the Creodias Virtual Machine	Data will be available for partners to perform processing.	No
T9.4/ CREO		Satellite imagery, national geospatial datasets					No

	The data which might be applicable for further use and which might be easily added to the repository may include open datasets		Numeric (geospatial data, satellite imagery), metadata	.tiff, .nc, .csv, .json, .xml. Potentially access to some OGC services. Volume of gathered data will vary depending on partner's needs. From dozens of GB to dozens of TB.	FTP server set up un the Creodias Virtual Machine with possibility of connection to the repository set up on the machine.	Data will be available for partners to perform processing.	

T10.1 / GSH	First dataset/group Satellite data, ground sensors data, open-source intelligence data. Input from WP6, WP7 & WP8 Second dataset/group UNHCR GIS DATA: Refugee camps and other people of concern's location OCHA Services: AL Zaatari Refugee camp, Mafraq Governorate, Jordan UN geospatial layers: Refugee Camps	Input from from WP6, WP7 & WP8	Images, text	Input from WP6, WP7 & WP8	Internal repository	Two datasets are identified. The one contains data requested by the tools of the THEIA system. The other one contains data provided by SATCEN, the refugee camps data. Concerning the first dataset, excluding ground-sensors which possess the capability of human identification, as GSH, and referring to satellite data, . As the rest of the data of the first dataset are concerned, maybe the relevant partners must answer. WTOC, ED, AIT. Concerning the second dataset, namely the data provided by SATCEN, they are open. Consequently, they are already available for use	No

						by external actors.	

T10.1/ LXS	The project will use existing data from public and open sources, including satellite imagery from the Copernicus Programme (Sentinel-1 and Sentinel-2), or any other Programme from those missions collected on WP6 and WP7 (e.g. CCM). Also may benefit from some open libraries and frameworks that can be highly useful for machine learning algorithms applied to image processing, data fusion, and the interpretation of satellite imagery for automated object detection. And probably also from pre-trained ML models from different repositores (e.g. European AI resource platform - AI4EU).	Satellite imagery data Metadata associated with the sat. Imagery (acquisition date / sensor type / resolution / etc.) Existing pre-trained machine learning models. Copernicus Open Access Hub or Copernicus Data Space Ecosystem - for Sentinel products. Supplementary geospatial datasetsor processed satellite products, from National data portals (e.g. French or Luxembourgish IGN) Open Libraries like GDAL OpenStreetMap for the land mask.	Image: Satellite imagery Numeric: ML model parameters Text: Metadata Software: ML model code	Imagery: GeoTIFF (mostly) or Sentinel.safe products (jp2) - size variable depending on the AOI and resolution - MB to GB Models (YOLO - onnx; programming langua: C#) Metadata: JSON, XML or similar.	Results are sent to MQTT broker - then CDO will receive the results.	Open data products can be shared. Results of detections, identifications, route predictions is up to the project to share them or not. If data comes from commercial providers (in case some images are acquired from them) may have license restrictions	No personal data is expected unless VHR imagery could make identifible any sensitive location Implementation of preprocessing for filtering out any piece of sensitive data or sensitive areas.
T10.2/AIT	Public (open-source) Datasets from Publications / Kaggle /Github, and possibly previously recorded datasets from AIT.	Type: Image/Video sensor measurements from different sensors with different wavelengths, including ground-truth annotations Source: Previously recorded datasets from AIT or publicly available, open-source datasets from Kaggle, Github, other online available sources	Images/Videos with ground-truth annotations and optionally corresponding meta-data (timestamps, geo-location)	.jpg or .png for images, .csv or .json for ground truth. For model training around 20GB of data, however a storage or retention of the data on the platform is not foreseen.	Not planned, open-source data is already publicly available	No.	TBD, RGB datasets can include images of persons.

T10.2/ C3I	UAS-acquired imagery and video datasets (C3I), combined with publicly available EO datasets and metadata where applicable	UAS platforms (C3I), EO datasets (Copernicus, open repositories), metadata from sensing systems	mage and video data, telemetry metadata, geospatial datasets	Common formats such as .mp4, .jpg/.png, and .json for metadata; dataset size varies depending on acquisition campaigns (typically several GB per dataset)	Internal project repositories and data exchange mechanisms within THEIA platform	No (or limited to selected datasets via project deliverables if applicable)	No (data collected at non-identifiable level)
T10.3 / ED	Official websites, media websites, social media	Georeferenced publicly available information	RSS Feeds, spreadsheets, geospatial/georeferenced data	XML-based, csv, xlsx, geospatial formats	Through spatial databse (dockerized). On top of this there will be APIs providing JSON data	Sensitive deliverables	No
T10.4/ SPACE-SI	Own satellite video and images	Own satellite video and images	Video, images	.mp4, .tif	Internal repository	No	No
T10.4/ LXS	The project will use existing data from public and open sources, including satellite imagery from the Copernicus Programme (Sentinel-1 and Sentinel-2), or any other Programme from those missions collected on WP6 and WP7 (e.g. CCM).	Satellite imagery data Metadata associated with the sat. Imagery (acquisition date / sensor type / resolution / etc.) Existing pre-trained machine learning models.	Image: Satellite imagery Numeric: ML model parameters Text: Metadata Software: ML model code	Imagery: GeoTIFF (mostly) or Sentinel.safe products (jp2) - size variable depending on the AOI and resolution - MB to GB Models (YOLO - onnx; programming langua: C#)	Results are sent to MQTT broker - then CDO will receive the results.	Open data products can be shared. Results of detections, identifications, route predictions is up to the project to share	No personal data is expected unless VHR imagery could make identifible any sensitive location

	Also may benefit from some open libraries and frameworks that can be highly useful for machine learning algorithms applied to image processing, data fusion, and the interpretation of satellite imagery for automated object detection. And probably also from pre-trained ML models from different repositores (e.g. European AI resource platform - AI4EU).	Copernicus Open Access Hub or Copernicus Data Space Ecosystem - for Sentinel products. Supplementary geospatial datasetsor processed satellite products, from National data portals (e.g. French or Luxembourgish IGN) Open Libraries like GDAL OpenStreetMap for the land mask.		Metadata: JSON, XML or similar.		them or not. If data comes from commercial providers (in case some images are acquired from them) may have license restrictions	Implementation of preprocessing for filtering out any piece of sensitive data or sensitive areas.
T10.4/ ICCS	Copernicus Data Space Ecosystem, ESA Open Access Hub (Sentinel-2 L2A), USGS Earth Explorer (Landsat 8)	Multispectral satellite imagery from Sentinel-2 L2A and Landsat 8	Optical multispectral imagery	Sentinel-2 L2A: JPEG2000 (.jp2), GeoTIFF; Landsat 8: GeoTIFF. Several GBs per scene.	Internal repository, cloud storage, or project-specific data platforms	Yes, publicly available through open data portals	No
T11.1/ GSH	satellite, UAS and IoT, OSINT	Type: Images, text, video Sources: CREODIAS, missions specified in WPs 6 and 7, Google Trends, X	raster, Video, text	*GeoTIFF, COGS, JSON*	central database, directly from components peripheral to CDO (Central data storage and orchestration)	Depends on the legal set of principles which are enforced upon the project conduct. To the extent that the legal set of principles allows it, partners who are responsible for the mechanisms of retrieval may decide on this.	*no*

T11.2/ GSH	No significant data or sources can be included at the moment. The reason is that the task follows task 11.1, or at least a significant part of 11.1.	X	X	X	X	X	X
T12.1/ ED	n/a	n/a	n/a	n/a	n/a	n/a	n/a
T12.2/ SATCEN	Not applicable						
T12.3/ GSH	Satellite data, ground sensors data, open-source intelligence data from WP6, WP7 & WP8	from WP6, WP7 & WP8	images, text				
T12.3/ SPACE-SI	Own satellite video and images	Own satellite video and images	Video, images	.mp4, .tif	Internal repository	No	No
T12.4/ SATCEN	Internal reports from system evaluation such as satisfaction ratings, task completion times, error rates, etc.	Interviews and evaluation forms	text	.doc	Internal repository	TBC	No
T13.1/ ICCS	Public communication data, aggregated analytics	Reports, companies websites, google search	Text, numeric	.pdf, .doc, .xls	Internal repository	Public deliverables	No
T13.2/ GSH	Data on industry trends, target market demographics, market needs, competitors.	Surveys, Public databases, reports. Companies websites, Google search	Text, numeric	.pdf, .doc, .xls	Internal repository	Public deliverables	No

T13.3/ MPL	Intelectual Property explotation strategy for the research in the project, and patenting strategy. Data used by other partners in their reports about the technological solutions researched in the project may reflected also here. The potential patenting information and steps for the registration are reflected in this content.	Text, mostly public legal information, some data from the technological solutions necessary for the legal analysis.	Text	.doc	Internal repository	Public deliverables	The content is foreseen to be a written report without any personal data in it, except from the names of the authors and reviewers. This may vary depending on the processing that other partners do in their reports to describe the technological solutions. Personal data will be anonymised.

T14.1/ ICCS	Public stakeholder information	Reports, companies websites, google search	Text, numeric	.pdf, .doc, .xls	Internal repository	Public deliverables	No
T14.2/ GSH	Data on industry trends, target market demographics, market needs, competitors.	Surveys, Public databases, reports. Companies websites, Google search	Text, numeric	.pdf, .doc, .xls	Internal repository	Public deliverables	No

T14.3/ MPL	Intelectual Property explotation strategy for the research in the project, and patenting strategy. Data used by other partners in their reports about the technological solutions researched in the project may reflected also here. The potential patenting information and steps for the registration are reflected in this content.	Text, mostly public legal information, some data from the technological solutions necessary for the legal analysis.	Text	.doc	Internal repository	Public deliverables	The content is foreseen to be a written report without any personal data in it, except from the names of the authors and reviewers. This may vary depending on the processing that other partners do in their reports to describe the technological solutions. Personal data will be anonymised.

New Datasets per WP and per Task

Task/ Partner	List the data that will be generated in the context of the project	Kind of data	Format of data (e.g. filename extension) & Expected size of the data	Will data or products/services generated in the project be made available?
T1.1/ GSH	CA among partners, MoM	text	.doc, .pdf	Public deliverables
T1.2/ GSH	Risk Registry, QAP	numeric, text	.doc, .pdf, .xls	Public deliverables
T1.3/ GSH	List of existing and new datasets	numeric, text	.doc, .pdf, .xls	Public deliverables
T2.1/ GSH	CA among partners, MoM	text	.doc, .pdf	Public deliverables
T2.2/ GSH	List of existing and new datasets	numeric, text	.doc, .pdf, .xls	Public deliverables
T3.1/ MPL	Benchmark legal analysis of the applicable legislation to the research in the project. The content is a written report without any personal data in it, except from the names of the authors and reviewers. The report is based on publicly available legal information, and was already submitted, without any new data produced. Questionnaires from partners. Information about the research in the project is collected from the partners in the project, to evaluate ethical aspects with a foreseen impact in the project.	text (documents), numeric outputs from the questionnaire results	.doc , .txt	Public deliverables

T4.1/ MPL	Benchmark legal analysis of the applicable legislation to the research in the project. The content is a written report without any personal data in it, except from the names of the authors and reviewers. The report is based on publicly available legal information, and was already submitted, without any new data produced. Questionnaires from partners. Information about the research in the project is collected from the partners in the project, to evaluate ethical aspects with a foreseen impact in the project.	text (documents), numeric outputs from the questionnaire results	.doc , .txt	Public deliverables
T5.1/ GSH	1.Questionnaire for CSS users, 2. Feedback from end users on CSS gaps and needs, 3. List of THEIA stakeholders 4. List of user requirements identified by end users, 5. THEIA results to be provided by all technical partners, 6. List of gaps in the CSS that will be filled by the THEIA project. 7. Vector files delineating specific AOIs of interest to the project, provided by SATCEN, and available at Dropbox.	text (documents), numeric outputs from the questionnaire results	.pdf, .xls, .doc, .txt	The gap analysis results will be shared with project partners and stakeholders, and some of the results will be made publicly available by public deliverables
T5.2/ AIT	X	X	X	X
T5.3/GSH	X	X	X	X
T5.4/ C3I	X	X	X	X
T6.1/ OHB-DS	Very-High-Resolution Earth Observation (VHR EO) sample data from Copernicus Contributing Missions and European national missions.	Very-High-Resolution optical and radar imagery, and related metadata	mostly .tif (GeoTiff). Approximately 65 GB.	The data is publically available. The data is primarily used internally to produce the report D6.1. Possibly the data can be reused for development or testing purposes in WP10.
T6.2/ SPACE-SI	Satellite video and images	Video, images	.mp4, .tif	Datasets will be available to the partners of the project.
T6.3/ OHB-DS	A filtered list of VHR missions (.csv) from T6.1.	text, numeric	.csv. Few kB	The data is already available to partners through D6.1.

T7.1/ LXS	- EO missions of interest for the WP7 - optical and SAR EO missions - list with screened missions and relevant parameters about them (whenever info was found)	Mostly text data - Microsoft Excel files.	The generated data will be provided in .xls format - the expected size won't be bigger than 5MB (1-5MB expected)	Data will be available for the rest of the partners, and for the Commission, but it was considered as **Sensitive** therefore it won't be publicly available.
T7.2/ LXS	- missions of interest for the WP7 - Radio Frequency missions - list with screened missions and relevant parameters about them (whenever info was found)	Mostly text data - Microsoft Excel files.	The generated data will be provided in .xls format - the expected size won't be bigger than 5MB (1-5MB expected)	Data will be available for the rest of the partners, and for the Commission, but it was considered as **Sensitive** therefore it won't be publicly available.
T7.3/ LXS	- broader evaluation of identified EO and RF missions, considering their capabilities and perfomance and how do they align with the stakeholders requirements identified in Task 5. - List of identified gaps between EO and RF mission's current and near future capabilities and stakeholders WP5 requirements. - List of recommendations for addressing each identified gap, with proposed enhancements. - These will be included in D7.1 and D7.2	Mostly text data - Microsoft Word and Excel files.	The generated data will be provided in .doc and .xls format - the expected size will be around 5-20MB	Data will be available for the rest of the partners, and for the Commission, but it was considered as **Sensitive** therefore it won't be publicly available.

T7.4/ C3I	1. Video Data Kind of Data: Raw video footage, edited video content, animations, recordings Format of Data: .mp4, .mov, .avi, .mkv, .webm Size: Varies by resolution, frame rate, and length. Could range from a few hundred MBs to several GBs. 2. Audio Data Kind of Data: Recorded interviews, voiceovers, background music, sound effects Format of Data: .mp3, .wav, .aac, .flac Size: Depending on the quality and length of the audio. For example, 1 hour of uncompressed .wav could be around 600 MB. 3. Textual Data Kind of Data: Scripts, project documentation, reports, metadata (e.g., for video labeling or transcripts) Format of Data: .txt, .docx, .pdf, .csv Size: Generally small in size, ranging from a few KBs to several MBs depending on content. 4. Image Data Kind of Data: Thumbnails, project-related images, infographics, storyboard frames Format of Data: .jpg, .png, .bmp, .gif Size: Could vary from a few KBs to a couple of MBs per image, depending on resolution.	Video Content, Project-Related Videos, Recorded Materials, Raw Footage	Common video file formats like: .mp4 (Most common format for sharing videos) .mov (Apple format, often used for higher quality) .avi (Older format but still used for specific purposes) .mkv (Often used for high-quality video storage) .webm (Web-friendly, used for online sharing) .bin (Usual format of log files) // The size of the data will depend on the project scope, content quality, and resolution. As mentioned earlier, video data might range from 500 MB for SD footage to 20 GB or more for high-definition or 4K video.	This would depend on the specifics of the project. For instance, if the project involves publishing videos publicly or sharing internally.

	5. 3D Models / Graphics Kind of Data: 3D models, animations, renders (if part of the project) Format of Data: .obj, .fbx, .stl, .dae, .gltf Size: Size will vary based on the complexity of the models (can be hundreds of MBs to a few GBs). 6. Data Logs or Analytics Kind of Data: Logs tracking project progress, performance metrics, or analytics data (e.g., if it's a web-based or interactive project) Format of Data: .log, .csv, .json Size: Generally smaller files (a few KBs to MBs depending on the granularity of the data). 7. Software Code or Scripts Kind of Data: Source code, scripts for automation, or application development (if relevant) Format of Data: .py, .js, .html, .css, .java Size: Typically small, a few KB to a couple of MBs depending on the complexity of the codebase. 8. Project Management Data Kind of Data: Task tracking, project timelines, budget documents, project schedules Format of Data: .xlsx, .csv, .docx, .pdf Size: Generally small to medium (could range from a few KB to MB).			

	9. Backup or Archived Data Kind of Data: Compressed backups of video files, project-related assets for archiving or safekeeping Format of Data: .zip, .tar, .7z Size: Varies widely depending on the number and size of files. 10. Data Products/Deliverables (Post-Project Outputs) Kind of Data: Final product such as a completed video, interactive media, website, report, etc. Format of Data: .mp4, .pdf, .html, .zip, .exe (for interactive products) Size: Varies widely based on project output, from a few MBs to several GBs. 11. Log files Files directly from autopilot of the UAV - Format of Data: .bin, etc.. Size: 50MB - 10GB and more. It depends.			

T8.1 / WTOC	- Anonymized and curated crowdsourced data will be available to partners for processing but not stored - Topic modeling files with metadata containing analytical indicators	text, numeric, interactive figures	JSON, html Expected size: <1GB	Data and analytical outputs will be available to partners internally and to end-users through the THEIA platform.
T8.2 / ED	Spatial analysis datesets based on the openly available existing data.	Numeric (databases, spreadsheets), Rasters data	Spatial database with indicators and measurements. GeoJSON and JSON data through dedicated APIs Expected size: <1GB	Sensitive Deliverable. Datasets will be shared with project partners and stakeholders.

T8.3 / ED	Spatial analysis datesets based on the openly available existing data.	Numeric (databases, spreadsheets), Rasters data	Spatial database with indicators and measurements. GeoJSON and JSON data through dedicated APIs Expected size: <1GB	Sensitive Deliverable. Datasets will be shared with project partners and stakeholders.
T9.1/ ICCS	Satellite Imagegery Data from Copernicus services (mainly) and other services (i.e., USGS) for assisting the purpose of model development in other tasks.	Satellite Images	NetCDF, Expected Size: ~4.5GB per DC product (i.e., in case we need to store data cubes in yearly basis, the storage is estimated ~3TB)	Sensitive Deliverable. Datasets will be shared with project partners and stakeholders.
T9.1/ SPACE-SI	Satellite video and images	Video, images	.mp4, .tif	Datasets will be available to the partners of the project.
T9.2/ C3I	Secure CREODIAS exchange service (connector/API integration), security configuration & access policies, deployment scripts, technical documentation, test/audit logs	Software + configuration + documentation + (system/audit) logs	.py/.ts, .yaml/.json, .md/.pdf, .log (+ container images); size: MB–few GB	Internal/authorised access only
T9.3/ CREO	There are no new datasets in this task. CREO is delivering input datasets from Creodias repository.	N/A	N/A	Available for partners to perform processing.
T9.4/ CREO	In this task CREO is gathering needed data from external sources. Thus, no new datasets are generated.	Numeric (geospatial data, satellite imagery), metadata	Volume of gathered data will vary depending on partner's needs. From dozens of GB to dozens of TB.	Available for partners to perform processing.
T10.1/ GSH	Tents, Vehicles (OHB), Vessels (LXS), ship-to-ship. Corresponding satellite images	bounding boxes on images, images	Format: JSON, GeoTIFF.	Datasets will be available only to the partners of the project for the duration of the project
T10.1 & T10.4/ LXS	Processed imagery - results from the data fusion techniques (including Json's with detections, identification, route predictionl, etc). ML model code - result from the Task 10.1 (e.g ML model for automating detection / recognition of Objects of Interest) Different DLV reports DLV 10.1 and 10.4	Image: results from the different processes. Json files with results (detection, identification, etc.) Text: Metadata / Reports Software: ML model code	Imagery: GeoTIFF (mostly) or Sentinel.safe products (jp2) - size variable depending on the AOI and resolution - MB to GB Models (YOLO - onnx; programming langua: C#) Metadata: JSON, XML or similar.	These deliverables are considered sensitive, therefore won't be available (only internaly)

T10.2 / AIT	Video data + metadata (e.g. geo-location of camera) for a dataset specific to our use case (24/7 outdoor utilization, different sensors) is recorded within T7.4, for later use in T10.2	Images/ videos plus corresponding metadata	Videos and metadata, size of new datasets (raw video data): 1.7GB of LWIR and 24.8GB of RGB videos.	Currently not planned, project outcome will be the deliverables (sensitive!).
T10.2/ C3I	Processed datasets derived from UAS-acquired imagery and video, including fused data outputs and associated metadata generated within Task 10.2.	Processed optical and thermal imagery, annotated video data, fused multi-sensor datasets, geospatial data layers (including positioning and temporal information), and structured metadata describing acquisition parameters, sensor characteristics and processing outputs.	Standard formats such as .mp4, .jpg/.png and .json for metadata; size depends on processing workflows, typically ranging from several GBs per dataset	Limited availability within the consortium; selected results may be included in project deliverables
T10.3 / ED	Spatial and trend analysis based on the georeferencing and analysis of existing data	Gespatial datasets (vector and/or raster)	Spatial database with indicators and measurements. GeoJSON and JSON data through dedicated APIs Expected size: <1GB	Datasets will be shared with project partners and stakeholders.
T10.4/ LXS	Processed imagery - results from the data fusion techniques (including Json's with detections, identification, route predictionl, etc). ML model code - result from the Task 10.1 (e.g ML model for automating detection / recognition of Objects of Interest) Different DLV reports DLV 10.1 and 10.4	Image: results from the different processes. Json files with results (detection, identification, etc.) Text: Metadata / Reports Software: ML model code	Imagery: GeoTIFF (mostly) or Sentinel.safe products (jp2) - size variable depending on the AOI and resolution - MB to GB Models (YOLO - onnx; programming langua: C#) Metadata: JSON, XML or similar.	These deliverables are considered sensitive, therefore won't be available (only internaly)
T10.4/ SPACE-SI	Satellite video and images	Video, images	.mp4, .tif	Datasets will be available to the partners of the project.

T11.1/ GSH	no data generation is currently foreseen in CDO. CDO is presently oriented towards storage and orchestration mainly. In other words, not a highly-algorithmic nature is foreseen.	NA	Size: Concerning the size of the data an estimate does not exist. However, it must be noted that images, at least, and possibly the corresponding detections as well will be deleted periodically.	NA
T11.2/ GSH	all data referenced in 11.1. The reason is that the intention is to provide all data and functionalities to CSS, in order to enhance it operationally.	same as 11.1	same as 11.1	same as 11.1
T12.1/ ED				
T12.2/ SATCEN	Not applicable			
T12.3/ GSH	Object detections and identification. Images patches or whole images, which correspond the detections. Suggestions/actionable data	12.3 is the demonstration phase, so all data generated by every module will be generated.	all data generated by other modules	Datasets will be available to the partners of the project.
T12.3/ SPACE-SI	Satellite video and images	Video, images	.mp4, .tif	Datasets will be available to the partners of the project.
T12.4/ SATCEN	Textual documents regarding the evaluation of the performance and the impact of THEIA platform. These reports will reflect activities, operational processes, practices and outcomes from the final project evaluation	text	.docx	TBC
T13.1/ ICCS	Printed dissemination materials such as newsletters, leaflet, brochures and posters etc. Online dissemination materials such as social media posts, eposters, press releases etc.	Text, numeric	.doc, .pdf, .ppt	Public deliverables and website
T13.2/ GSH	Detailed information about the business, the products and the services. Explotation and marketing plans. Financial data and projections.	Text, numeric	.doc, .pdf, .xls	Public deliverables
T13.3/ MPL	Intelectual Property explotation strategy for the research in the project, and patenting strategy. Data used by other partners in their reports about the technological solutions researched in the project may also be reflected here. The potential patenting information and	Text, numeric	.doc	Public deliverables

	steps for the registration are reflected in this content. There is not any new data processed or identified under this work.			
T14.1/ ICCS	Printed dissemination materials such as newsletters, leaflet, brochures and posters etc. Online dissemination materials such as social media posts, eposters, press releases etc.	Text, numeric	.doc, .pdf, .ppt	Public deliverables and website
T14.2/ GSH	Detailed information about the business, the products and the services. Explotation and marketing plans. Financial data and projections.	Text, numeric	.doc, .pdf, .xls	Public deliverables
T14.3/ MPL	Intelectual Property explotation strategy for the research in the project, and patenting strategy. Data used by other partners in their reports about the technological solutions researched in the project may also be reflected here. The potential patenting information and steps for the registration are reflected in this content. There is not any new data processed or identified under this work.	Text, numeric	.doc	Public deliverables

END OF DOCUMENT

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under GA 101190051. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.