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

Initial Data Management Plan (DMP), D1.3

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	Vyron Antoniou (ED LUXEMBOURG)

Due Date	31/03/2025
Delivery Date	31/03/2025
Type	DMP — Data Management Plan
Dissemination Level	PU - Public
Keywords	Data management, FAIR principles, Data security, Ethical Compliance, GDPR

Document History

Version	Date	Description	Author	Description/ Action	Validation
	03/03/2025	Outline	L. Panagiotopoulou (GSH)		Deliverable leader
0.1	14/03/2025	First Draft	L. Panagiotopoulou (GSH)	Initial draft of the contents	Consortium partners comments. Feedback is collected and incorporated into the document.
0.2	21/03/2025	Draft for review	L. Panagiotopoulou (GSH)	Internal Review process	Consortium partner ED LUXEMBOURG review. Feedback is collected and incorporated into the document.
0.3	27/03/2025	Final Draft	L. Panagiotopoulou (GSH)	Finalisation and final approval	SAB security check. Feedback is collected and incorporated into the document.
1.0	31/03/2025	Final version	L. Panagiotopoulou (GSH)	Submission to EC	

Legal Disclaimer

Contents

Executive Summary

The current deliverable D1.3 “Initial Data Management Plan (DMP)” corresponds to Task T1.3 – “Initial Data Management Plan” under WP1 – “Project Management 1 – 1st period,” led by GSH, the project coordinator.

This document, which is the initial DMP of the project developed at M4, outlines how data will be collected, processed, stored, shared and preserved throughout the THEIA project lifecycle, ensuring both effective and secure data management while maintaining compliance with relevant regulations.

The document begins with an overview of THEIA’s objectives and the DMP framework, followed by a data summary and a detailed explanation of how FAIR principles will be applied. It also covers resource allocation, data security, the legal framework, IPR and ethical considerations.

As data management is an ongoing process, this Initial DMP will be updated regularly to reflect new developments and best practices.

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
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 initial version of the Data Management Plan.

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 one of the outputs of Task 1.3 and represents the third 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 DMP is to offer an outline on how data will be 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”
- 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 initial version of the DMP, with two more versions to follow: an intermediate DMP at month 15 (M15) and a final version of DMP at month 30 (M30), which will also cover the post-project period. The DMP will be regularly updated as needed to align with project progress given that project needs and datasets will change.

As the first version, Deliverable D1.3 aims to give a first report about the data that will be collected during the project, ensuring compliance with FAIR data principles (Findable, Accessible, Interoperable, Reusable). It also details how the data are stored, in which repository and how they are preserved.

Finally, although this initial version of the deliverable does not strictly follow the outline described in the GA, as the project is still in its early stages, having started only three months ago, it nevertheless addresses all the issues outlined in the GA. In the next version (M15) and the final version (M30), the template specified in T1.3 of the GA will be used.

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 utilized

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, Analyzing Data, Preserving Data, Sharing Data etc. as presented in the following Diagram (Figure 1).

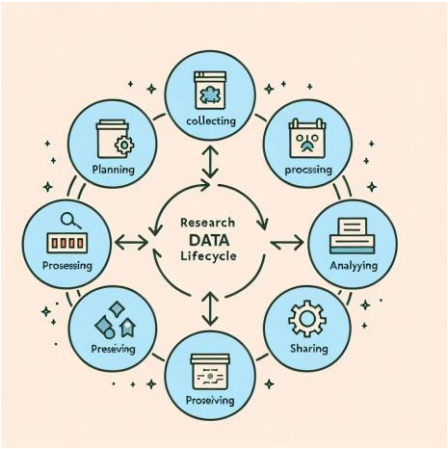

Figure 1. Data cycle

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 first version of the DMP is to:

1. Present first 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 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 will be 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, will be 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 prioritizes data protection and security policies to ensure compliance and safeguard sensitive information. Additionally, THEIA consortium partners consider important the data quality, 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 Deliverable 1.2). 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 anonymization, encryption, if necessary, etc. for mitigating risks associated with third-party access.

This data governance strategy aims to maximize the value of collected and utilized 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:

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

At the partner level, each consortium member is required to designate a Partner Data Protection Officer (ParDPO) or an individual responsible for ensuring compliance with data protection regulations.

Project THEIA relies on the ParDPOs and designated data protection contact points within partner organizations to uphold personal data protection compliance. Each of the consortium partners is responsible for the processing of personal data associated with their research activities within the project.

THEIA Partners are required to provide the contact details of their DPO if they have appointed one in accordance with their legal obligations or individual regulatory requirements. For partners not legally required to appoint a DPO, a designated data protection contact point must be assigned. This person will collaborate on legal compliance with public authorities, respond to data subject requests and address other data protection compliance matters.

The details and the names of the persons will be confirmed in the THEIA JCA.

The main responsibilities of the THEIA Data Manager include:

- 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 already mentioned above to effectively address ethical concerns within the project, an Ethical Manager has been appointed and his responsibilities include:

- Coordinating comprehensive protection against all forms of discrimination.
- Collaborating with the Project Manager on personal data protection in accordance with the relevant Regulations (GDPR).
- Defining the ethical and societal framework of the project.

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 first description of the data utilized 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 development of the initial version of the DMP, partners began defining the types of data that will be collected and generated throughout the project, along with the rationale and necessity for collecting this data.

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

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 will be regularly updated as the project progresses and will be included in following versions 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 analyze 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).

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 2. 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	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, 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, WP5, WP12, WP13, WP14
Presentations	Electronic document	Powerpoint document	All WPs

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

To support the project deliverables, data will be 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

In next Table, THEIA deliverables classified as PUBLIC.

Table 3. THEIA Public Deliverables

No	Title
D1.2	Risk identification Management & Quality assurance plan

No	Title
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)
D5.2	Enhanced Service Features
D5.3	Use Case Definition Report
D5.4	EO Data Processing Pathway Selection Guide
D5.5	THEIA Architecture Concept
D6.1	Delivery of the tailored Very-High-Resolution Earth Observation tools
D6.2	Analysis of Space-based Video and Multi-Payload Data Utilization Potential
D6.3	Selection and optimisation of the methodology of appropriate space data
D7.3	Integration and Field Test Report
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.

Table 4. THEIA Sensitive Deliverables

No	Title
D1.1	MoM of Kick-off Meeting

No	Title
D1.6	Initial Progress Report
D2.3	Progress Report
D5.1	Stakeholders engagement plan, CSS Gap Analysis Report and User Feedback Summary
D7.1	Delivery of the Micro-satellites Cubesats & UAS-based data acquisition tools
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

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.

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 persistent identifier will be adopted (please see next Section). Metadata will be used that 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).

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 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

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 centralized 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

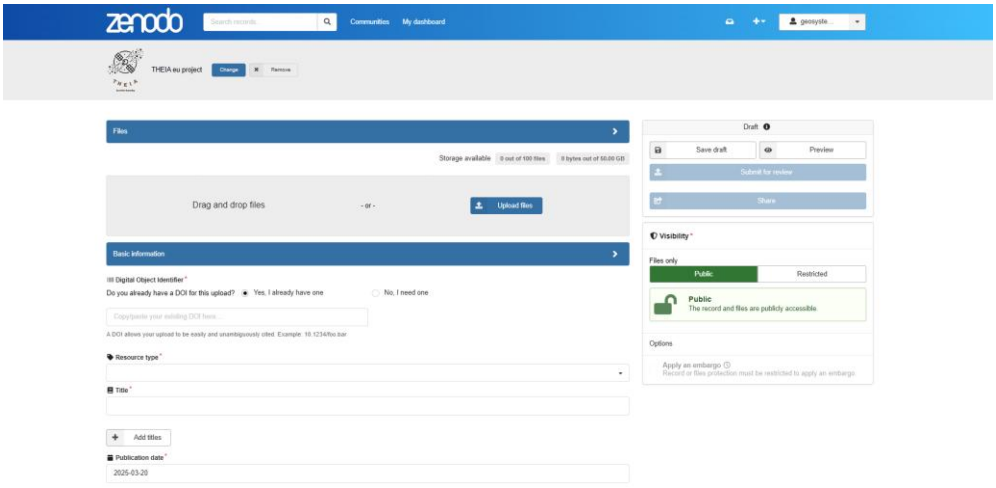

Figure 3. THEIA community in Zenodo

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.

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 standardized 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 standardized 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.

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 organization involved. Data management will be handled as part of WP1 (Project Management). Currently, GSH, as the project coordinator, along with contributions from SATCEN, AIT and C3I, 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 unauthorized access, use and distribution. Access controls and user privileges are also considered among the THEIA consortium.

All generated and collected data will be securely managed, in the duration of the THEIA project, to protect against unauthorized access and ensure there is no loss or leak of information. Additionally, an incident response plan will be developed and delivered in later version of DMP to ensure effective action in the event of a 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. Personal data will be made available for access only to authorized personnel. This responsibility concerns all partners, hence all of them will ascertain that all data is protected by complying to all security and access controls within their institution. In this context, the development and signing of a JCA is planned in the near future.

The data collected and/or generated throughout the whole period of project's execution will be held in data repositories in the servers of consortium partners.

As far as the "physical" data storage is concerned, project's documents will be kept in an office's secure environment, in computers with authentication and authorization mechanisms.

The THEIA project will implement a backup and recovery plan to ensure that the data can be recovered in the event of a disaster or system failure. Backups of the databases will be stored on the partners' premises. Data backups of devices will systematically take place every week and be stored in devices that follow the same security standards and procedures as the main server.

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).

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.

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 deliverable, specifically addressing: (a) the types of personal data and special categories of data utilized, (b) the processing of personal data and its lawful basis and (c) compliance with the data minimization 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.

The legal basis for processing is defined in Article 6(1) of the GDPR.

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 5. 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

9 IPR

The consortium recognizes 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 maximize the successful implementation of project results.

Knowledge and IPR management are embedded within the framework of the CA as issues related mainly to Results and Access Rights (Chapter 8 and 9 respectively), which aligns with the policies and guidelines of EC funded projects under Horizon Europe. The CA defines the general terms and conditions under which partners can access existing knowledge or new knowledge generated by other consortium members.

The IPR aspects will be further elaborated throughout the project. During the first project period (M1-M15), Task T13.3, "Identification of IPR Issues and Patentable Content," will focus 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.

Finally, it is important to note that the appropriate IPR protections will be in place to safeguard the designs of both the components and the platform itself, ensuring the preservation of the commercialization strategy. For all other aspects, the project's policies will be aligned with Open Science principles.

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 Ethics Advisor, 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 represents the initial THEIA DMP at Month 4 of the project. It mainly serves as a guideline for data management, outlining the key principles the project follows, including Open Science and FAIR principles. Additionally, it provides an overview of the security and ethical considerations related to data management.

These principles will be followed through the use of state-of-the-art tools and standards, such as the Open Science initiative and the Zenodo depository for research data, to guarantee that the THEIA results (open data, open science publications etc.) will be properly preserved and remain accessible and available for use even after the end of the project's lifecycle.

Besides, this initial version of the DMP also includes the first results of consolidated feedback from all partners regarding an initial data survey within the consortium. This was conducted using an Excel file titled "*Inventory of Existing and New Data*", where all partners began defining the types of data to be collected and generated throughout the project, along with the rationale and necessity for collecting this data.

Next, the critical issue of personal data is addressed in alignment with other project's deliverables and activities.

Finally, the THEIA DMP, is a living document that will continue to be updated regularly, as data management remains an ongoing task relevant to almost all project's 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.

END OF DOCUMENT