GeoAI for Security

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

Assessment of UAS and Terrestrial Sensing (Integration and Field Test Report), D7.3

WP7 – Micro- & pico-satellites and UAS based data acquisition

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	C3I
Contributors	Panagiotis Pierettis (C3I), Ines Burgstaller (AIT), Matteo Marturini (AIT)
Reviewers	SATCEN (Jose Santos, Juan Francisco Romero Quesada)

Due Date	28/2/2026
Delivery Date	03/04/2026
Type	R — Document, report
Dissemination Level	PU - Public
Keywords	UAS, terrestrial sensing, data acquisition, geo-referencing, metadata specification, multi-sensor platform, optical and thermal imaging, artificial intelligence methodology, object detection, multi-modal datasets

Document History

Version	Date	Description	Comments	Edited by
0.1	12/9/2025	First draft	ToC and deliverable structure	Eirini Barri (C3I)
0.2	22/9/2025	First internal version	Chapter 1,2,3	Eirini Barri (C3I)
0.3	30/11/2025	Second Internal version	Chapter 4, all	Ines Burgstaller (AIT)
0.4	27/01/2026	Third Internal Version	All Sections	Eirini Barri (C3I)
0.5	18/02/2026	Review Version	Requesting Review (PC, Partners)	Eirini Barri (C3I)
0.6	23/03/2026	Final Draft	SATCEN review completed. Feedback incorporated.	Panagiotis Pierettis (C3I) Ines Burgstaller (AIT)
0.7	30/03/2026	Revised Final Draft	SAB security check completed. Feedback incorporated.	Panagiotis Pierettis (C3I)
1.0	02/04/2026	Final version		Panagiotis Pierettis (C3I)
1.0	03/04/2026	Final version	Submission to EC	L. Panagiotopoulou (GSH)

Legal Disclaimer

Contents

Executive Summary

Deliverable D7.3 – Assessment of UAS and Terrestrial Sensing presents the technical assessment of airborne and terrestrial sensing activities conducted within Work Package 7 (WP7) of THEIA. The deliverable addresses structured data acquisition, geo-referencing capabilities, and the definition of metadata associated with sensing platforms, alongside a desk-based review of suitable Artificial Intelligence (AI) methodologies relevant to the project's use cases.

The airborne sensing component is based on the configuration and technical assessment of a DJI Matrice 350 RTK unmanned aerial vehicle equipped with a Zenmuse H30T multi-sensor payload. The Unmanned Aerial Systems (UAS) platform is characterised in terms of its operational configuration, sensing modalities, positioning accuracy, and its capacity to generate geo-referenced optical and thermal datasets. Particular emphasis is placed on the nature of the data products produced during acquisition activities — including imagery, video, telemetry, and auxiliary measurements — and on the metadata elements required to ensure spatial traceability and reproducibility.

In parallel to the airborne component configured by C3I, terrestrial sensing activities were conducted by AIT using the AIT terrestrial multi-sensor platform, ensuring a complementary ground-based perspective within WP7.

The geo-referencing capabilities of the UAS platform, enabled through RTK GNSS positioning, timestamp synchronization, and orientation metadata, support accurate spatial documentation of airborne observations. Within the scope of this deliverable, the UAS assessment focuses exclusively on sensing characteristics and metadata availability at the acquisition stage.

In parallel, WP7 includes terrestrial data acquisition activities performed by AIT using AIT's deployable ground-based multi-sensor platform. The ground-based platform operates independently from the UAS system and is used to acquire complementary datasets in WP7. The terrestrial platform contributes additional RGB and LWIR data acquisition activities within WP7. The AI methodology review and algorithm evaluation presented in Section 4.3, performed by AIT, are based on RGB imagery from open-source datasets used to benchmark state-of-the-art detection approaches, as a basis for further developments within T10.2.

The terrestrial platform data acquisition described earlier in this chapter serves dataset enrichment and future experimentation purposes but is not used in the evaluation presented in this deliverable.

Overall, Deliverable D7.3 establishes a structured technical baseline for non-space sensing assets within WP7, focusing on airborne and terrestrial data acquisition characteristics, geo-referencing mechanisms, and AI methodology review. The document maintains a clear distinction between

sensing and processing responsibilities and does not address system-level integration or demonstration activities.

List of Tables

List of Figures

List of Acronyms / Abbreviations

Table 1. List of Acronyms/Abbreviations

Acronym / Abbreviation	Explanation
AI	Artificial Intelligence
AIT	Austrian Institute of Technology
AP	Average Precision
CNN	Convolutional Neural Network
CSS	Copernicus Security Services
DETR	DEtection TRansformer
EO	Earth Observation
EU	European Union
FPS	Frames Per Second
GA	Grant Agreement
GeoAI	Geospatial Artificial Intelligence
GDPR	General Data Protection Regulation
GNSS	Global Navigation Satellite System
JSON	JavaScript Object Notation
ML	Machine Learning
MP4	MPEG-4 Video Format
RTK	Real-Time Kinematic
ToC	Table of Contents
UAS	Unmanned Aerial System
UR	User Requirement
VHR	Very High Resolution
WP	Work Package
YOLO	You Only Look Once

1. Introduction

Addressing critical challenges such as population displacement due to conflicts, climate change impacts, extreme weather events, food insecurity, and socio-economic instability remains a priority for European security and crisis management policies. In this context, the THEIA project seeks to enhance situational awareness capabilities by supporting the structured acquisition and analysis of heterogeneous data sources relevant to displacement monitoring and crisis response scenarios.

THEIA aims to strengthen existing Copernicus Security Services by complementing space-based assets with additional sensing modalities and analytical approaches. Through the coordinated use of Earth Observation, Radio Frequency analytics, and non-space sensing platforms, the project contributes to improved information availability for decision-makers operating in demanding security and humanitarian environments.

Within this broader framework, non-space sensing assets — including UAS and terrestrial multi-sensor platforms — provide flexible and targeted data acquisition and analysis capabilities. These platforms enable the collection of high-resolution optical and thermal datasets that can complement other information sources and support geospatial analysis in dynamic operational contexts.

This document, Deliverable D7.3 – Assessment of UAS and Terrestrial Sensing, is produced within the framework of Work Package 7 (WP7). The deliverable addresses activities related to the technical characterization and data acquisition capabilities of UAS platforms equipped with multi-sensor payloads and of terrestrial multi-camera systems.

In accordance with the amended scope of work, the deliverable focuses on:

- Documenting the data acquisition processes performed using airborne and ground-based sensing platforms,
- Defining the metadata elements required to ensure accurate geo-referencing and spatial traceability,
- Providing a structured desk-research-based assessment of relevant AI) methodologies for object detection and classification in general and with regards to classes-of-interest, provided by the end-users.

This deliverable presents the assessment of UAS and terrestrial sensing within WP7.

Its detailed purpose, scope and structure are described in Sections 1.1 and 1.2

Work Package 7 – Micro-satellites, CubeSats and UAV-based data acquisition comprises the following tasks:

- Task 7.1: Micro- and pico-satellites towards advanced sensing [M2–M15] – Led by LuxSpace, supported by GSH.
- Task 7.2: Radio Frequency (RF) analytics [M2–M15] – Led by LuxSpace.
- Task 7.3: Assessment of current and planned missions against requirements, gap analysis and recommendations [M5–M15] – Led by LuxSpace, supported by GSH.
- Task 7.4: Data acquisition including relevant metadata for geo-referencing and desk research on relevant AI algorithms for UAS with different sensors and terrestrial cameras (Leader: C3I, Contributors: GSH, AIT) [M2–M15].

Deliverable D7.3 constitutes an output of WP7 and provides a structured assessment of UAS and terrestrial sensing activities carried out within the project. It consolidates the documentation of data acquisition campaigns performed using UAS and ground-based sensing platforms, the definition of associated geo-referencing metadata, and the desk-based review of relevant AI algorithms. Contributions from C3I focus on the characterisation of UAS platforms and sensing configurations, while AIT contributes to terrestrial data acquisition activities and to the evaluation of suitable AI methodologies.

1.1 Purpose and scope of the deliverable

The purpose of Deliverable D7.3 is to provide a structured technical assessment of UAS, and terrestrial sensing solutions considered within the THEIA project. The deliverable documents the data acquisition activities performed using UAS and ground-based sensing platforms, specifies the metadata elements required to ensure accurate geo-referencing, and presents desk-research based review of relevant AI methodologies.

Within the context of WP7, this deliverable characterises the sensing capabilities, operational configurations, data formats, and spatial referencing parameters of both airborne and terrestrial camera systems. The emphasis is placed on acquisition-stage characteristics and metadata availability, establishing a clear technical baseline for non-space sensing assets considered within the project. The assessment therefore encompasses both the UAS platform configured by C3I and the AIT terrestrial multi-sensor platform used for structured ground-based data acquisition and AI methodology evaluation.

The scope of this document includes the following elements:

- Sensing configuration and platform characterisation – detailing the technical characteristics of the UAS and terrestrial sensing platforms, including sensor payloads and operational parameters relevant to structured data acquisition.

- Data acquisition procedures and formats – describing the methodologies used for collecting optical and thermal data, the recording configurations adopted, and the file formats employed (e.g., imagery, video, structured metadata records).
- Geo-referencing and metadata specification – defining the spatial referencing parameters (e.g., GNSS coordinates, timestamps, altitude, orientation, sensor configuration) and the structured metadata elements required to ensure traceability, reproducibility, and spatial accuracy.
- Desk-research-based assessment of AI methodologies – reviewing state-of-the-art Artificial Intelligence approaches relevant to object detection and fine-grained classification tasks associated with THEIA use cases, as further discussed in Section 4.3
- Technical observations and considerations – summarising key findings related to sensing characteristics, data quality aspects, and practical constraints identified during acquisition activities.

The deliverable does not address system-level integration, interoperability validation, or demonstration activities. Its scope is limited to sensing characterisation, data acquisition documentation, metadata definition, and AI methodology assessment.

1.2 Structure of the deliverable

This deliverable is organised into five main chapters, each addressing a specific aspect of the assessment of UAS and terrestrial sensing activities within THEIA. The structure reflects the focus on data acquisition, geo-referencing, metadata specification, and AI methodology assessment, in accordance with the scope of WP7.

Chapter 1 – Introduction

Provides the background, purpose, scope, and structure of the deliverable, situating Deliverable D7.3 within the broader objectives of THEIA and WP7.

Chapter 2 – Technical Characterisation of the UAS Platform

Presents the technical characteristics of the UAS deployed within WP7. The chapter describes the platform configuration, sensing modalities, operational parameters relevant to structured data acquisition, and positioning capabilities enabling accurate geo-referencing.

Chapter 3 – UAS Data Acquisition and Metadata Specification

Documents the data acquisition procedures associated with the UAS platform, including recording configurations, data formats, and structured metadata elements required to ensure spatial traceability and reproducibility. The chapter defines the geo-referencing parameters and documentation practices applied at the acquisition stage.

Chapter 4 – Terrestrial Multi-Sensor Platform and AI Methodology Research and Assessment

Describes the ground-based multi-sensor platform deployed by AIT, including the data acquisition campaign conducted to enrich and diversify a pre-existing multi-modal dataset. The chapter also presents a desk-research based assessment of suitable object detection and vision-language AI methodologies for image classification.

Chapter 5 – Conclusions and Technical Observations

Summarises the key findings related to UAS and terrestrial sensing characteristics, metadata availability, and AI methodology suitability. The chapter provides consolidated technical observations derived from the assessment activities conducted within WP7.

The structure ensures a clear separation between airborne and terrestrial sensing components and maintains a focused scope on sensing characterisation, data acquisition documentation, and AI methodology review.

2. Technical Characterisation of UAS Platform

This chapter presents the technical characterisation of the UAS deployed within WP7 activities for airborne data acquisition. The platform was configured, prepared for operation, and assessed in the context of Deliverable D7.3 to support structured, geo-referenced sensing aligned with THEIA use cases and identified data needs.

The aerial system used in the assessment is the DJI Matrice 350 RTK UAS equipped with the Zenmuse H30T multi-sensor payload (source: DJI). The platform was selected due to its operational robustness, multi-sensor capability, and high-precision positioning features, which together enable reliable acquisition of optical and thermal datasets under diverse environmental and lighting conditions. The system supports real-time mission monitoring and structured post-mission data handling, facilitating consistent organisation of raw sensing outputs and associated telemetry.

The characterisation presented in this chapter is based on the practical configuration and operation of the UAS within WP7 activities. This includes mission planning considerations such as flight profiles, altitude selection, coverage parameters, and payload configuration choices, as well as the identification and handling of the data products generated during acquisition. Particular emphasis is placed on the sensing outputs produced (optical imagery, video streams, thermal data, and auxiliary measurements) and on the geo-referencing information accompanying each dataset, ensuring spatial traceability and reproducibility.

The chapter establishes a clear technical baseline for the airborne sensing component considered in this deliverable. The focus remains on acquisition-stage characteristics and metadata-relevant parameters, without addressing system-level integration or interoperability aspects. The deliverable presents representative results from the UAS and terrestrial sensing activities, illustrating the data acquisition capabilities of the deployed platforms. These results are not intended as a comprehensive operational performance evaluation, but rather as indicative examples supporting the system characterisation.

2.1 Platform Overview and Operational Context

The airborne sensing activities performed within WP7 are supported by the deployment and operational configuration of a DJI Matrice 350 RTK Unmanned Aerial System (UAS). The platform is utilised as a mobile airborne sensing asset designed to enable structured acquisition of geo-referenced optical and thermal data aligned with THEIA use cases.

The DJI Matrice 350 RTK represents an enterprise-class multi-rotor UAS platform engineered for professional data acquisition tasks in demanding operational environments. Its configuration

within WP7 was oriented toward achieving reliable sensing performance, positional accuracy, and repeatable acquisition conditions rather than exploratory or ad hoc flight recording.

The platform supports extended flight endurance, enabling meaningful coverage of areas of interest within a single mission while maintaining operational safety margins. Stable hover performance and controlled waypoint navigation are essential characteristics supporting consistent image capture and thermal data acquisition. Flight stability directly influences data quality by reducing motion artefacts, maintaining consistent viewing angles, and enabling predictable overlap between captured frames when required.

The UAS is designed to operate under varied environmental conditions, including moderate wind exposure and fluctuating ambient temperatures. Such robustness allows data acquisition activities to be conducted under realistic field conditions relevant to displacement monitoring and security-oriented use cases.

The main technical specifications of the deployed UAS platform (DJI Matrice 350 RTK) are summarised in the below table.

Parameter	Value
Dimensions (unfolded)	810 × 670 × 430 mm
Dimensions (folded)	430 × 420 × 430 mm
Diagonal Wheelbase	895 mm
Weight (with batteries)	~6.47 kg
Max Takeoff Weight	9.2 kg
Max Payload (single gimbal)	960 g
Max Flight Time	up to 55 min
Max Speed	23 m/s
Max Wind Resistance	12 m/s
Max Flight Altitude	up to 7000 m
Operating Temperature	-20°C to 50°C

Ingress Protection	IP55
Positioning Systems	GPS, GLONASS, BeiDou, Galileo
RTK Accuracy	1 cm + 1 ppm (horizontal)
Hovering Accuracy	±0.1 m (RTK)
Operating Frequency	2.4 GHz / 5.1–5.8 GHz
Supported Payloads	EO/IR cameras (e.g. Zenmuse H30T)
Gimbal Configurations	Single / dual / upward / downward
EU Class	C3

A key characteristic of the deployed UAS is its integrated Real-Time Kinematic (RTK) Global Navigation Satellite System (GNSS) capability. RTK-enhanced positioning significantly improves spatial accuracy compared to standard GNSS positioning by incorporating correction signals that reduce positional uncertainty.

Within WP7 activities, this high-precision positioning capability is particularly relevant for:

- Accurate geo-referencing of captured imagery,
- Association of sensing outputs with precise spatial coordinates,
- Reduction of cumulative positional errors across flight trajectories,
- Structured metadata generation supporting spatial traceability.

The RTK system enhances GNSS positioning by applying real-time carrier-phase corrections from a base station or network, enabling centimetre-level positioning accuracy and supporting precise geo-referencing of airborne data

T Mission planning parameters were defined based on a standard system configuration, while allowing limited adjustments depending on the operational scenario and environmental conditions. The following key parameters were consistently considered during UAS configuration:

- Flight altitude selection relative to required spatial resolution,
- Ground coverage area and trajectory design,
- Speed adjustments to balance coverage efficiency and image stability,

- Sensor orientation and operating modes,
- Recording settings for optical and thermal streams.

These configuration decisions were taken to ensure controlled and reproducible data acquisition rather than exploratory flight experimentation. The objective was to generate datasets accompanied by consistent positional and temporal information, suitable for structured documentation and subsequent analysis.

Figure 1:DJI Matrice 350 RTK UAS equipped with Zenmuse H30T multi-sensor payload

During acquisition activities, the UAS simultaneously generates multiple categories of sensing outputs, including optical and thermal imagery, video streams, telemetry data and auxiliary measurements, supported by the capabilities of the Zenmuse H30T multi-sensor payload. These data streams are synchronised with positioning information and timestamps, enabling structured documentation of sensing conditions at the moment of capture.

Parameter	Value
Optical Camera	40 MP zoom camera + 48 MP wide-angle
Thermal Camera	VOx microbolometer, 1280×1024 resolution
Thermal Spectral Range	8–14 µm
Thermal Sensitivity (NETD)	≤ 50 mK

Video Output	4K (optical), thermal video supported
Zoom Capability	up to 400× hybrid zoom
Stabilisation	3-axis gimbal
Laser Rangefinder	up to 3000 m
Night Operation	Supported (IR + thermal)
Data Outputs	Optical imagery, thermal imagery, video streams, telemetry

The platform supports both real-time monitoring during flight, through live video and telemetry transmission to the ground control station, and systematic post-flight data retrieval via onboard storage. This dual capability ensures that acquisition quality can be observed during operation while preserving full-resolution datasets for structured storage and documentation after mission completion.

Within the scope of Deliverable D7.3, the DJI Matrice 350 RTK UAS serves as the airborne sensing component responsible for generating geo-referenced datasets. Its contribution is limited to acquisition-stage sensing and metadata availability. Processing activities described later in this deliverable relate exclusively to terrestrial datasets.

The UAS platform therefore establishes the airborne sensing baseline within WP7, focusing on acquisition performance, spatial positioning capability, and metadata-relevant characteristics.

2.2 Multi-Sensor Payload Configuration – Zenmuse H30T

The DJI Matrice 350 RTK UAS deployed within WP7 is equipped with the Zenmuse H30T multi-sensor payload. The payload integrates complementary sensing modalities within a stabilised gimbal system, enabling simultaneous acquisition of optical and thermal data during a single flight mission.

The payload configuration in WP7 was designed to support structured data acquisition under varying environmental and lighting conditions, while enabling operation in real-world scenarios. Particular attention was given to sensor mode selection, recording parameters, and operational stability to ensure reproducible sensing outputs accompanied by consistent positional metadata.

The EO sensor (electro-optical), integrated in the Zenmuse H30T payload, supports high-resolution visual imagery and video acquisition. Adjustable zoom functionality enables observation of objects or areas of interest at varying distances without requiring excessive proximity to the target area. This flexibility supports both wide-area situational awareness and focused observation tasks.

Optical recording parameters such as resolution, frame rate, and encoding format were configured to balance image clarity, storage requirements, and post-mission handling efficiency. Stable gimbal control ensures consistent framing and reduces motion-induced distortion, supporting clear visual documentation of observed scenes.

The Zenmuse H30T integrates a long-wave infrared (LWIR) thermal sensor capable of detecting temperature differentials within the observed scene. Thermal sensing enables data acquisition in low-visibility conditions, including night-time operation or environments with reduced ambient light.

Thermal data acquisition is particularly relevant for:

- Detection of heat-emitting objects,
- Identification of human presence in low-light environments,

- Observation of operational activity where visual contrast is limited.

Thermal recording parameters were configured to ensure stable capture of temperature-based imagery alongside synchronised positional metadata.

The payload includes an integrated laser rangefinder enabling distance measurement to observed objects. While not used as a primary sensing modality, range information can support spatial interpretation of observations and contextual understanding of object positioning within the scene.

Distance measurements, when recorded, are associated with corresponding positional and temporal metadata generated by the UAS platform.

The integration of optical, thermal, and auxiliary sensing modalities within a single payload allows heterogeneous data capture during a single mission. This multi-modal capability supports:

- Cross-referencing between optical and thermal observations,
- Improved detection reliability under varying environmental conditions,
- Flexible adaptation to mission objectives.

Within WP7 activities, payload configuration was adapted depending on acquisition objectives and environmental conditions, ensuring that sensing outputs remained consistent with structured data documentation requirements.

All sensing outputs generated by the Zenmuse H30T payload are synchronised with telemetry and positional information provided by the UAS platform. This includes:

- Timestamp information,
- Geographic coordinates,
- Altitude data,
- Platform orientation parameters.

This synchronisation ensures that each image, video segment, or measurement can be associated with precise spatial context. The availability of consistent positional metadata forms the basis for the geo-referencing and metadata specification described in Chapter 3.

2.3 Data Products and Sensing Outputs

The UAS platform deployed within WP7 generates multiple categories of sensing outputs during airborne data acquisition missions. These outputs are produced simultaneously and are synchronised with positional and temporal metadata to ensure spatial traceability.

The data products generated during acquisition activities can be grouped into the following categories:

The optical sensor produces high-resolution RGB imagery and video streams. These outputs support visual scene documentation and enable observation of objects, infrastructure elements, and environmental features within the area of interest.

Still imagery captures discrete frames associated with specific timestamps and positional coordinates, while video streams provide continuous visual recording of flight segments. Recording configurations are selected to ensure clarity and stability while maintaining manageable storage requirements.

The thermal sensing component generates long-wave infrared imagery representing temperature differentials within the observed scene. Thermal outputs are particularly valuable in low-light or night-time conditions, as well as in scenarios where visual contrast is limited.

Thermal imagery and video are synchronised with positional data in the same manner as optical outputs, enabling spatially referenced documentation of heat-emitting objects and activity patterns.

In parallel with sensing outputs, the UAS platform continuously records telemetry information. This includes:

- Geographic coordinates (latitude and longitude),
- Altitude above ground level,
- Platform orientation parameters (roll, pitch, yaw),
- Speed and flight status indicators,
- Timestamp information.

Telemetry data are essential for associating captured imagery and video with precise spatial and temporal context. The availability of structured telemetry enables accurate geo-referencing of sensing outputs and supports reproducibility of acquisition trajectories.

Where applicable, auxiliary sensor data such as laser rangefinder measurements may be recorded during acquisition. These measurements provide distance information relative to observed objects and can support spatial interpretation of the scene.

Although auxiliary data are not the primary sensing modality, their association with positional metadata contributes to enhanced contextual understanding of observations.

A key characteristic of the deployed UAS configuration is the synchronisation of sensing outputs with telemetry and positioning information. Each image frame, video segment, or measurement

is associated with timestamp and coordinate data, ensuring that airborne observations are documented within a structured spatial reference framework.

This synchronisation forms the foundation for the metadata specification and geo-referencing procedures described in Chapter 3.

The scope of this section is limited to the identification and characterisation of sensing outputs generated at the acquisition stage. The scope of this section is limited to the identification and characterisation of sensing outputs generated at the acquisition stage. The deliverable does not address post-acquisition processing of UAS data, which is handled in subsequent WPs (e.g., WP10).

2.4 Geo-Referencing and Positioning Capabilities

Accurate geo-referencing constitutes a fundamental requirement for structured airborne data acquisition within WP7. The deployed UAS platform integrates high-precision positioning capabilities that enable spatially traceable documentation of all sensing outputs generated during flight operations.

The DJI Matrice 350 RTK UAS incorporates Real-Time Kinematic (RTK) Global Navigation Satellite System (GNSS) functionality. RTK positioning enhances conventional satellite-based navigation by incorporating correction signals that reduce positional uncertainty and improve spatial accuracy.

Under appropriate GNSS conditions, RTK capability enables centimetre-level positioning accuracy.:

- Reliable association of imagery with geographic coordinates,
- Accurate mapping of observed objects or areas of interest,
- Consistent reproduction of flight trajectories across missions,
- Reduction of cumulative positional drift over extended acquisition paths.

High-precision positioning is particularly important in scenarios requiring spatial documentation of small-scale features or when correlating airborne observations with other geospatial datasets.

In addition to spatial coordinates, each sensing output generated by the UAS is associated with timestamp information. Timestamp synchronisation ensures that imagery, video frames, telemetry data, and auxiliary measurements can be accurately aligned within a temporal sequence.

Temporal metadata supports:

- Reconstruction of acquisition timelines,
- Frame-level association between sensing outputs and telemetry records,
- Correlation of events observed during flight.

The combination of spatial coordinates and precise timestamps establishes a coherent spatial-temporal reference framework for airborne datasets.

Beyond positional coordinates, the UAS platform records orientation and attitude parameters during flight. These typically include:

- Roll,
- Pitch,
- Yaw (heading).

Orientation metadata enables accurate interpretation of camera perspective and viewing geometry. Knowledge of platform attitude at the moment of capture is essential for:

- Understanding image footprint orientation,
- Interpreting viewing angles,
- Supporting structured spatial documentation.

The inclusion of orientation parameters strengthens the completeness of metadata associated with each dataset.

Altitude information recorded during flight operations provides additional spatial context for airborne observations. Altitude data contribute to:

- Estimation of ground sampling characteristics,
- Approximation of image footprint dimensions,
- Interpretation of object scale within captured imagery.

When combined with positional and orientation data, altitude measurements contribute to a comprehensive geo-referencing framework.

All spatial and temporal parameters recorded by the UAS are associated with corresponding sensing outputs at the acquisition stage. This association may occur through:

- Embedded geospatial metadata within still imagery files,
- Synchronised telemetry logs accompanying video recordings,
- Structured metadata records generated alongside sensing data.

The availability of consistent positional and orientation metadata ensures that each dataset is spatially traceable and reproducible. This capability forms the basis for the metadata specification framework detailed in Chapter 3.

The geo-referencing capabilities described in this section relate exclusively to acquisition-stage positioning and metadata generation. The deliverable does not address geospatial post-processing workflows, orthorectification pipelines, or integration with external geospatial infrastructures. The focus remains on the availability and structure of metadata at the moment of data capture.

2.5 Data Handling at Acquisition Stage

Structured data handling at the acquisition stage is essential to ensure traceability, consistency, and reproducibility of airborne sensing outputs. Within WP7 activities, data handling procedures were defined to maintain organised storage and clear association between sensing outputs and corresponding metadata.

During flight operations, the UAS supports real-time monitoring of telemetry and video streams through the Ground Control Station (GCS). This capability enables immediate observation of sensing quality, flight parameters, and operational status.

However, full-resolution sensing outputs are preserved through onboard storage mechanisms integrated within the payload and platform. Onboard storage ensures that raw imagery, video streams, and associated telemetry records are retained without compression-related degradation that may occur during live transmission.

This dual capability — real-time monitoring combined with high-quality post-flight data retrieval — supports both operational oversight and structured documentation.

The data acquisition activities involved human presence only at a non-identifiable level (e.g. aerial perspective with no possibility of individual recognition). As such, no personal data were collected, and all data handling was conducted in accordance with applicable data protection principles and internal procedures

Following mission completion, sensing outputs are retrieved from onboard storage devices and transferred to secure local storage environments for documentation and organisation. The retrieval process preserves:

- Optical imagery files,
- Thermal imagery and video files,
- Telemetry logs,

- Auxiliary measurement records.

Care is taken to ensure that no modification of raw datasets occurs during transfer. Original acquisition files are preserved to maintain data integrity and traceability.

To ensure structured documentation, acquisition outputs are organised according to predefined naming conventions and directory structures. File organisation typically reflects:

- Mission identifier,
- Date and time of acquisition,
- Sensor modality (optical, thermal, auxiliary),
- Platform identifier.

This structured approach allows clear association between datasets and their corresponding metadata records. It also facilitates systematic documentation without implying integration into broader system architectures.

All sensing outputs are accompanied by metadata elements generated at the acquisition stage, including:

- Timestamp information,
- Geographic coordinates,
- Altitude,
- Orientation parameters,
- Sensor configuration identifiers.

Metadata may be embedded directly within imagery files or stored in associated structured records. Preservation of metadata integrity is considered essential to maintain spatial traceability and reproducibility.

Data handling procedures at the acquisition stage prioritise:

- Prevention of data loss,
- Preservation of raw dataset integrity,
- Controlled access to stored acquisition outputs.

Encrypted communication channels protect telemetry and video streams during transmission to the Ground Control Station. Access to stored data is restricted to authorised personnel involved in WP7 activities.

These measures ensure that sensing outputs remain reliable and traceable from the moment of capture through post-flight documentation.

The data handling procedures described in this section are limited to acquisition-stage storage, organisation, and metadata preservation. The deliverable does not address system-level data integration, automated processing pipelines, or interoperability mechanisms.

2.6 Security and Data Protection at Acquisition Level

The acquisition of airborne imagery and associated telemetry data may involve observation of sensitive environments, infrastructure elements, or, in certain scenarios, identifiable individuals. For this reason, security and data protection considerations are addressed at the acquisition stage of UAS deployment within WP7 activities.

During flight operations, communication between the UAS and the Ground Control Station (GCS) is conducted through encrypted transmission channels provided by the platform's communication system. This reduces the risk of unauthorised interception of telemetry or live video streams during active missions.

Access to the UAS control interface is restricted to authorised operators involved in WP7 activities. Authentication mechanisms defined at platform level ensure that only designated personnel can initiate and manage flight operations.

Following mission completion, acquired datasets are stored in secure local environments with controlled access. Access to raw sensing outputs and associated metadata is limited to authorised project personnel responsible for documentation and analysis tasks.

Data handling procedures aim to:

- Preserve dataset integrity,
- Prevent unauthorised modification,
- Avoid unintended disclosure of sensitive content.

Where acquisition activities involve environments in which individuals may be present, privacy considerations are taken into account. The scope of this deliverable is limited to technical characterisation and metadata documentation; however, awareness of potential privacy implications informs responsible data handling practices.

Security and data protection measures described in this section relate exclusively to acquisition-stage practices associated with the UAS platform. Broader cybersecurity frameworks, system-level integration safeguards, or cross-platform data exchange mechanisms are outside the scope of this deliverable.

3. UAS Data Acquisition and Metadata Specification

This chapter defines the data acquisition framework and metadata specification associated with the Unmanned Aerial System (UAS) component described in Chapter 2. The objective is to establish a structured approach to airborne sensing that ensures spatial traceability, reproducibility, and consistency of documentation.

The focus of this chapter is on acquisition-stage parameters and metadata availability rather than on operational flight campaign results. It describes the configuration principles governing data capture, the expected sensing outputs generated by the UAS platform, and the structured metadata elements required to enable accurate geo-referencing of airborne datasets.

Particular emphasis is placed on the definition of spatial, temporal, and orientation metadata associated with optical and thermal sensing outputs. These elements form the foundation for consistent documentation of acquisition conditions and provide a clear framework for associating imagery and video content with geographic coordinates.

The chapter does not address post-acquisition processing workflows, automated analytics pipelines, or system-level data integration mechanisms.

Through this structured acquisition and metadata framework, the UAS component of WP7 is assessed in terms of its capacity to produce geo-referenced sensing outputs aligned with THEIA use cases and documentation requirements.

3.1 UAS Data Acquisition Framework

The UAS data acquisition framework defined within WP7 establishes the configuration principles and documentation requirements necessary to support structured airborne sensing. The objective of this framework is to ensure that sensing outputs generated by the UAS platform are spatially traceable, reproducible, and accompanied by consistent metadata elements.

The framework focuses on acquisition-stage parameters governing how optical and thermal data are captured and how associated telemetry and positional information are recorded.

Figure 2: UAS data acquisition and documentation workflow

Figure 2 illustrates the acquisition-stage workflow associated with the UAS component. Sensing outputs generated by the UAS platform are first stored locally together with synchronised telemetry and positioning metadata. These outputs are subsequently organised and documented within a structured framework to ensure traceability and reproducibility.

To illustrate the sensing capabilities of the deployed UAS platform, example outputs from the acquisition activities are presented. The UAS system acquired both RGB (electro-optical) and thermal imagery together with associated telemetry metadata. These datasets demonstrate the capability of the platform to generate geo-referenced sensing outputs suitable for documentation and further analysis within the project.

Figure 3: Example thermal (LWIR) image acquired by the UAS payload during sensing activities.

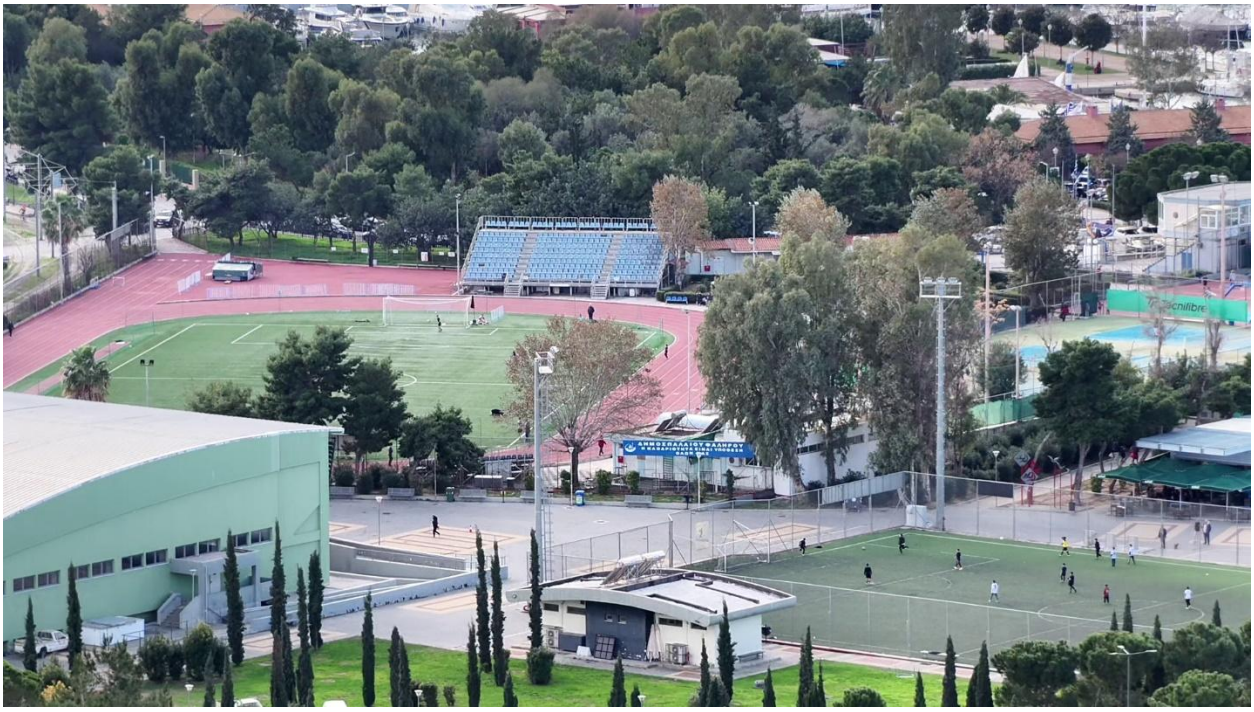

Figure 4: Example RGB frame acquired by the UAS platform during flight operations (The image includes human presence; however, no personal data can be identified, as individuals are not recognisable due to the spatial resolution and acquisition conditions.)

In addition to imagery data, the UAS platform records telemetry parameters during flight operations. These parameters include spatial position, altitude, flight speed, and platform orientation, enabling precise documentation of sensing conditions during acquisition. An example of such telemetry metadata extracted from the flight log is presented in below table:

Table 2: Example UAS Telemetry Metadata

Parameter	Example Value	Unit
Timestamp (UTC)	2026-01-11 10:27:33	ISO 8601
Altitude above sea level	19.706	feet
Height above take-off	0.010	feet
Ground speed	0.083	mph
Satellites used	8	count
Compass heading	359.42	degrees

Parameter	Example Value	Unit
Pitch	2.07	degrees
Roll	−0.47	degrees
Battery level	97	%

The workflow focuses exclusively on data capture, storage, and metadata association. It does not represent system-level integration, automated ingestion mechanisms, or downstream processing architectures.

The UAS platform allows configuration of multiple acquisition parameters that influence spatial resolution, coverage, and data quality. These include:

- Flight altitude selection in relation to desired ground detail,
- Flight trajectory planning and coverage definition,
- Speed adjustments to balance stability and area coverage,
- Sensor modality selection (optical and/or thermal),
- Recording resolution and encoding settings,
- Activation of telemetry and auxiliary measurement logging.

These parameters are defined to support structured and repeatable data capture aligned with the sensing objectives of WP7. The framework ensures that acquisition settings are documented and associated with each dataset.

The acquisition framework assumes synchronised recording of sensing outputs and telemetry information. Optical imagery, thermal imagery, and video streams are generated in parallel with positional and orientation metadata, enabling each data product to be associated with:

- Geographic coordinates,
- Timestamp,
- Platform orientation,
- Altitude and flight status indicators.

This synchronisation is fundamental to enabling geo-referenced documentation of airborne observations.

Although this deliverable does not report operational mission results, the acquisition framework has been defined with consideration of representative use-case scenarios within THEIA. These scenarios require high-resolution, geo-referenced sensing outputs that can support monitoring of dynamic environments and small-scale features.

Accordingly, the configuration principles prioritise:

- Spatial accuracy,
- Metadata completeness,
- Stable capture conditions,
- Consistent documentation of acquisition parameters.

To support traceability, acquisition parameters and recording configurations are documented alongside sensing outputs. This ensures that each dataset can be traced back to defined configuration settings and that acquisition conditions remain reproducible.

The acquisition framework therefore establishes a structured baseline for UAS-based sensing within WP7, focusing exclusively on acquisition-stage characteristics and metadata availability.

3.2 Data Formats and File Structures

The UAS platform supports standardised data representations to enable structured documentation and spatial traceability of airborne sensing outputs. The selection of data formats within the acquisition framework prioritises compatibility with widely adopted geospatial and multimedia standards, while maintaining clarity of metadata association.

Still optical imagery generated by the UAS platform may be stored in formats that support high-resolution visual capture and the embedding of geospatial metadata. When geo-referencing information is embedded directly within the image file, formats such as GeoTIFF enable spatial parameters (e.g., coordinates and altitude) to be associated with each image frame.

Where applicable, embedded metadata fields include:

- Geographic coordinates,
- Altitude,
- Timestamp,
- Camera orientation parameters.

The use of such formats ensures that still imagery retains spatial context without requiring external referencing mechanisms.

Video streams captured during airborne sensing activities are typically stored in widely adopted multimedia container formats (e.g., MP4). Video files are accompanied by synchronised telemetry logs containing spatial and temporal metadata.

Since video formats do not inherently embed full geospatial information at frame level, spatial traceability is maintained through structured association between video segments and corresponding telemetry records.

Telemetry data generated during flight operations are recorded in structured formats that may include machine-readable representations (e.g., JSON-based records). These records contain:

- Latitude and longitude coordinates,
- Altitude,
- Timestamp,
- Orientation parameters,
- Flight status indicators.

Metadata records are designed to be consistently associated with corresponding sensing outputs through shared identifiers, timestamps, or naming conventions.

To ensure structured documentation, acquisition outputs are organised according to predefined file and directory conventions. Organisational principles may include:

- Separation by mission identifier,
- Separation by sensor modality (optical / thermal),
- Chronological ordering based on acquisition time,
- Consistent naming structures incorporating timestamp and platform identifiers.

This structured approach ensures clarity of association between sensing outputs and metadata records, supporting reproducibility and traceability.

The data formats and file structures described in this section relate exclusively to acquisition-stage documentation. The deliverable does not address system-level ingestion mechanisms, data federation architectures, or interoperability frameworks. The objective is to define how sensing outputs and associated metadata are structured at the moment of capture and initial storage.

3.3 Geo-Referencing Parameters and Metadata Schema

Accurate geo-referencing is a central requirement for structured airborne sensing within WP7. The metadata schema defined for the UAS component establishes the spatial, temporal, and configuration parameters required to ensure traceability and reproducibility of sensing outputs.

The metadata framework is designed to associate each image frame, video segment, or auxiliary measurement with a well-defined spatial-temporal context. The schema includes the following categories of metadata elements.

3.3.1 Spatial Metadata

Spatial metadata elements define the geographic location of the UAS platform at the moment of data capture. These parameters enable precise positioning of sensing outputs within a geodetic reference framework.

Core spatial metadata elements include:

- **Latitude** (in decimal degrees)
- **Longitude** (in decimal degrees)
- **Altitude** (above ground level or mean sea level, as applicable)
- **Coordinate Reference System (CRS)** identifier

These parameters enable direct mapping of sensing outputs and support spatial alignment with other geospatial datasets when required.

3.3.2 Temporal Metadata

Temporal metadata ensures that each sensing output is associated with a precise timestamp, enabling reconstruction of acquisition sequences and alignment with telemetry data.

Core temporal metadata elements include:

- **Timestamp** (ISO 8601 format)
- **Acquisition sequence identifier** (where applicable)

Temporal precision supports frame-level association between imagery and telemetry records and enables chronological ordering of sensing outputs.

3.3.3 Orientation and Attitude Metadata

In addition to geographic position, the orientation of the UAS platform at the moment of capture is recorded. Orientation metadata enables interpretation of viewing geometry and supports accurate spatial contextualisation of imagery.

Core orientation parameters include:

- **Roll**
- **Pitch**
- **Yaw (heading)**

These values describe the platform's attitude relative to the Earth's reference frame and contribute to comprehensive documentation of acquisition conditions.

3.3.4 Sensor Configuration Metadata

Sensor-related metadata captures configuration settings relevant to the generation of sensing outputs. These parameters support reproducibility and clarity of acquisition conditions.

Typical sensor configuration elements include:

- Sensor modality (optical / thermal)
- Recording resolution
- Frame rate (for video)
- Zoom level (for optical sensing)
- Thermal mode selection (where applicable)

By documenting these parameters, the metadata schema ensures that datasets can be interpreted in relation to their acquisition configuration.

3.3.5 Platform Identification and Mission Context

To ensure traceability, metadata records may also include identifiers that associate datasets with specific acquisition contexts.

These may include:

- Platform identifier
- Mission or configuration identifier
- Operator reference (where applicable)
- Acquisition date

Such identifiers support structured documentation without implying operational campaign reporting.

3.3.6 Metadata Structure and Association

Metadata elements may be:

- Embedded directly within imagery files (e.g., geospatial tags), or

- Stored in associated structured records synchronised with sensing outputs.

Association between metadata and sensing outputs is maintained through shared timestamps, unique identifiers, or consistent naming conventions defined within the acquisition framework.

The metadata schema defined in this section ensures that all sensing outputs generated by the UAS platform can be spatially and temporally located within a clearly defined reference framework.

Table 2 summarises the core metadata elements associated with UAS-based sensing outputs within WP7, organised by category and acquisition-stage relevance.

Table 3: Summary of UAS Metadata Parameters

Category	Parameter	Description	Unit / Format	Source
Spatial Metadata	Latitude	Geographic latitude of UAS at time of capture	Decimal degrees	RTK GNSS
Spatial Metadata	Longitude	Geographic longitude of UAS at time of capture	Decimal degrees	RTK GNSS
Spatial Metadata	Altitude	Altitude of UAS relative to ground or sea level	Meters (m)	RTK GNSS / Altimeter
Spatial Metadata	Coordinate Reference System (CRS)	Geodetic reference system used for coordinates	EPSG code	GNSS configuration
Temporal Metadata	Timestamp	Time of data capture	ISO 8601 format (UTC)	UAS system clock
Temporal Metadata	Acquisition Sequence ID	Identifier for ordered capture sequence	Alphanumeric	Acquisition framework
Orientation Metadata	Roll	Rotation around longitudinal axis	Degrees (°)	IMU
Orientation Metadata	Pitch	Rotation around lateral axis	Degrees (°)	IMU
Orientation Metadata	Yaw (Heading)	Rotation around vertical axis	Degrees (°)	IMU / Compass
Sensor Metadata	Sensor Modality	Optical or Thermal mode	Text	Payload configuration
Sensor Metadata	Recording Resolution	Image or video resolution	Pixels (e.g., 1920x1080)	Sensor configuration
Sensor Metadata	Frame Rate (Video)	Video capture rate	Frames per second (fps)	Sensor configuration
Sensor Metadata	Zoom Level	Optical zoom configuration	Numeric level	Payload configuration
Platform Context	Platform Identifier	Identifier of deployed UAS	Alphanumeric	System configuration
Platform Context	Mission Identifier	Reference to acquisition configuration	Alphanumeric	Documentation framework

The metadata elements summarised above constitute the structured parameter set required to ensure spatial traceability, reproducibility, and clarity of documentation for UAS-derived sensing outputs.

These parameters are generated at platform level and preserved during acquisition and initial storage procedures.

The metadata schema described herein pertains exclusively to acquisition-stage documentation and geo-referencing capability. The deliverable does not address advanced geospatial processing techniques, orthorectification workflows, or integration with external geospatial infrastructures. The objective is to define the structure and availability of metadata required for accurate spatial documentation.

3.4 Metadata Association and Documentation Workflow

The metadata association and documentation workflow defined for the UAS component ensures that all sensing outputs are consistently linked with their corresponding spatial, temporal, and configuration parameters. The objective is to maintain traceability from the moment of capture through initial storage and documentation.

During acquisition, sensing outputs (optical imagery, thermal imagery, and video streams) are synchronised with telemetry data generated by the UAS platform. Synchronisation mechanisms ensure that each dataset is associated with:

- Timestamp information,
- Geographic coordinates,
- Altitude,
- Orientation parameters.

This association enables spatial-temporal traceability at frame level or segment level, depending on the sensing modality.

Following acquisition, sensing outputs and associated metadata are retrieved from onboard storage without modification of original files. Preservation of raw datasets is prioritised to ensure integrity and reproducibility.

Metadata may be:

- Embedded within imagery files through geospatial tags, or
- Stored as structured metadata records accompanying sensing outputs.

The documentation workflow ensures that metadata integrity is maintained during transfer and storage.

To support traceability and clarity, sensing outputs are organised according to predefined documentation principles. These may include:

- Grouping by mission or configuration identifier,

- Chronological ordering of acquisition outputs,
- Separation by sensor modality,
- Consistent naming conventions incorporating timestamp and platform identifiers.

Such documentation practices enable clear association between sensing outputs and corresponding metadata records without requiring external system-level integration mechanisms.

The defined metadata association framework ensures that each sensing output can be traced to:

- A specific acquisition configuration,
- A defined spatial location,
- A precise timestamp,
- Documented sensor parameters.

This structured approach supports reproducibility of acquisition conditions and establishes a transparent baseline for spatial documentation.

The documentation workflow described in this section pertains exclusively to acquisition-stage metadata handling and storage. The deliverable does not address automated ingestion pipelines, data federation layers, or integration into broader system architectures.

Processing and AI-based evaluation activities presented in Chapter 4 are independent of UAS data handling procedures.

4. Terrestrial Multi-Sensor Platform

The AIT terrestrial multi-sensor platform was deployed to conduct structured ground-based data acquisition activities within WP7. The platform operates independently from the UAS component described in previous chapters and is designed to capture complementary multi-modal datasets.

This chapter presents the terrestrial multi-sensor platform deployed within WP7 and describes its role in structured ground-based data acquisition activities. The platform was used to enrich and diversify AIT's existing multi-modal dataset through the collection of complementary optical and infrared sensor data.

In addition to data acquisition, this chapter documents research activities related to the identification and evaluation of suitable Artificial Intelligence (AI) methodologies aligned with THEIA use cases.

The activities described herein are independent of the UAS component presented in previous chapters. No processing of UAS-derived imagery is included within the scope of this chapter. The emphasis is placed on dataset composition, sensing characteristics, and the applicability of selected AI approaches within a structured evaluation context.

Through this assessment, the terrestrial sensing component contributes to the overall evaluation of non-space sensing capabilities within WP7.

4.1 Hardware Selection for the Multi-Sensor Platform

AIT's multi-sensor platform hardware consists of a modular architecture that allows adaptability of the platform with different modules and camera sensors for the desired use cases and required computational needs (see Figure 5). Additionally, apart from the modularity, a strong emphasis is put on compactness and ruggedness, to allow easy deployment of the platform in an area of interest and 24/7 outdoor operation. The platform consists of a camera housing, including different camera sensors, a GNSS module to record GPS data, a rugged housing including the computational hardware and power units and a pan-tilt unit for steering the camera head. Depending on the chosen camera sensors, the corresponding protective windows are chosen for the camera head to protect the camera sensors from dust and weather conditions without blocking the wavelengths of interest.

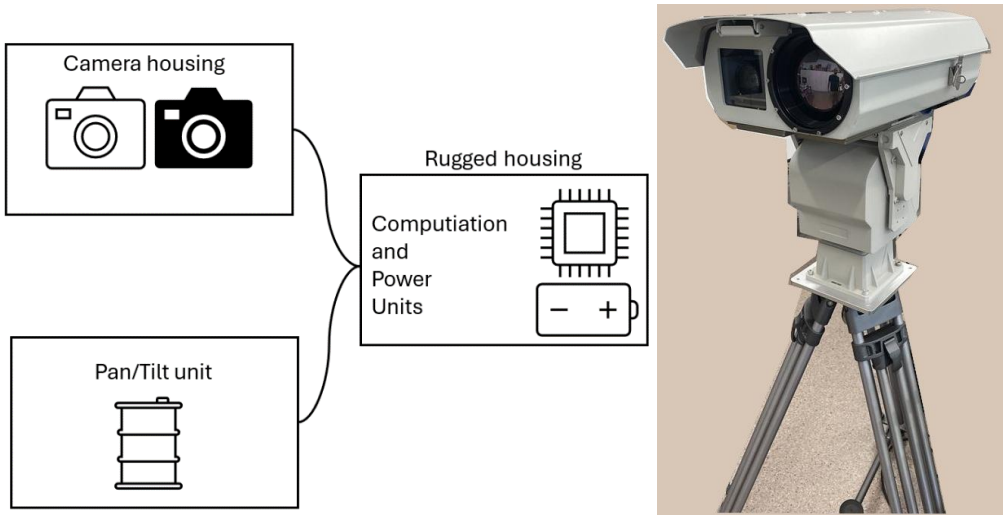

Figure 5: a) Concept of AIT's multi-sensor camera platform. b) real camera housing and pan-tilt unit

Considering the desired functionalities of the platform for THEIA aligned with Grant Agreement and User Requirements (defined in D5.1 and assigned to each project partner in D5.2) the following functionalities have been identified for the selection of camera sensors and further development of software components of the multi-sensor platform:

- Application in 24/7 outdoor operation in different weather conditions
    - Ruggedness of platform hardware, compactness for deployment (given by platform)
    - Use of complementary sensors with different wavelengths to obtain information in different environmental conditions (e.g. day, night, fog)
- Incorporation of thermal data
    - Incorporation of thermal (long-wavelength infrared (LWIR)) camera sensor into the platform
- Person, vehicle and vessel detection
    - Software requirement on object detection algorithm

Considering the identified functionalities and desired specifications regarding the sensor wavelengths, the following sensors were chosen out of a pre-existing pool of sensors of different wavelengths available from former projects (SWIR, UV, RGB, LWIR), for the camera head: A RGB camera sensor was chosen to provide data during good operational conditions during the day and generally when ambient lighting is available. A LWIR camera was chosen to fulfil the requirement of incorporating thermal data, as well as to obtain data within environmental conditions where the RGB camera sensor alone will not be sufficient (night, severe fog), due to its ability to detect heat signatures rather than relying on ambient light.

Following the identification of functionalities and choosing camera sensors based on the required spectral sensor modalities, desk research regarding suitable algorithms and publicly available

open-source datasets, as well as a data acquisition to diversify AITs custom person detection dataset were conducted.

4.2 Data Acquisition

Open-Source Object Detection Datasets in LWIR

Interest in infrared-based object detection stems from its ability to compensate for degraded RGB performance in conditions such as night-time, low light, or the presence of smoke and fog. However, progress in this domain is constrained by the limited availability of large, diverse, and consistently annotated infrared datasets. Supervised learning requires data covering a wide range of object classes and environmental conditions with varying thermal signatures (Bustos, 2023). Public infrared datasets also differ substantially in format and preprocessing, and undocumented contrast enhancement of raw thermal images complicates reproducibility and cross-dataset generalization (Danaci, 2024).

The large-scale benchmarks in the RGB domain such as the MSCOCO dataset (Lin, 2015), containing 80 classes and 330k images, the open images dataset from Google (Kuznetsova, 2020) containing roughly 600 object classes and 2 million images with bounding box annotations and the Objects365 (Shao, Objects365: A Large-scale, High-quality Dataset for Object Detection, 2019) dataset comprising 365 object classes within 2 million images, have no direct counterparts in the infrared spectrum. Among the various infrared spectral modalities, LWIR provides the largest number of public datasets, most notably FLIR-ADAS (FLIR, 2019) with 15 classes and 10k images, LLVIP (Jia, 2021) with 31k images and labelled pedestrians and CAMEL (Gebhardt, 2018) with 43k images and 5 classes. Nevertheless, these datasets are primarily collected for autonomous driving, resulting in limited diversity and a strong bias toward urban, road-centric scenes with vehicles and pedestrians. To mitigate these drawbacks a multi-modal dataset was previously created by AIT. This dataset consists of approximately 250000 images and contains three different spectral modalities: RGB, LWIR and SWIR (Short-wavelength infrared) to obtain a larger diversity than utilizing one spectral modality alone. The samples are collected and curated from publicly available datasets including COCO-person (Lin, 2015), CrowdHuman (Sharma, 2018), VisDrone (Zhu P. a., 2021), AAU-PD-T (Noor Ul Huda, 2020), CAMEL (Gebhardt, 2018), FLIR-ADAS (FLIR, 2019) and LLVIP (Xinyu Jia, 2021).

For person detection in land border environments, which is part of the THEIA use case PUC 2 “Terrestrial surveillance of population displacement and flows” the limited diversity of scenes and human poses in existing datasets, especially in the infrared domain represents a significant challenge. Unlike typical autonomous driving datasets, which predominantly depict urban settings with pedestrians in upright, unobstructed positions, land border scenarios often involve natural environments with vegetation. Therefore, a data acquisition campaign was planned and

executed in the late summer of 2025 at AIT's Seibersdorf Campus to record additional data with RGB and LWIR sensors for augmentation of the previously created multi-modal dataset.

Person Detection Data Acquisition in RGB and LWIR

To augment the custom dataset for later developments/evaluations within T10.2, the following scenarios were planned and recorded:

1. Persons with Nature Background
2. Persons (Partially) Occluded by Tall Grass
3. Persons with Varied Movements and Postures
4. Persons behind Smoke
5. Persons at Night

The hardware setup during the data acquisition included a camera head incorporating a RGB and a LWIR camera and a pan-tilt-unit to move camera head to e.g. follow persons walking through the scene. This setup was placed within the outdoor scene, at a distance and zoom such that the persons are in good view of the cameras. For the scenario recordings the format are ROS bags including the videos of the scenes, the pan-tilt-zoom metadata, as well as the GNSS position of the actors from the scenes. There were between 1 and 4 actors in each scene. These voluntary actors are employed by AIT and have given their consent to the data acquisition with consent forms, that informed the voluntary actors of their rights. Following is a description of the scenarios and acquired data:

1. Persons with Nature Background

This scenario consists of recordings of one or more persons walking concurrently in natural environments, including grassland and vegetated areas with trees in the background (see Figure 6). These recordings aim to increase the diversity of public datasets by incorporating simple scenes of persons in non-urban scenes, which are currently underrepresented especially within open-source LWIR data.

Figure 6: Single and multiple persons in nature scenes

2. Persons (Partially) Occluded by Tall Grass

This scenario consists of recordings of persons walking into and crouching behind tall grass. Due to different vegetations in nature, this scenario of partial occlusions by different types of tall grass and bushes found in nature aims to reflect realistic partial occlusions caused by diverse natural vegetation (see Figure 8).

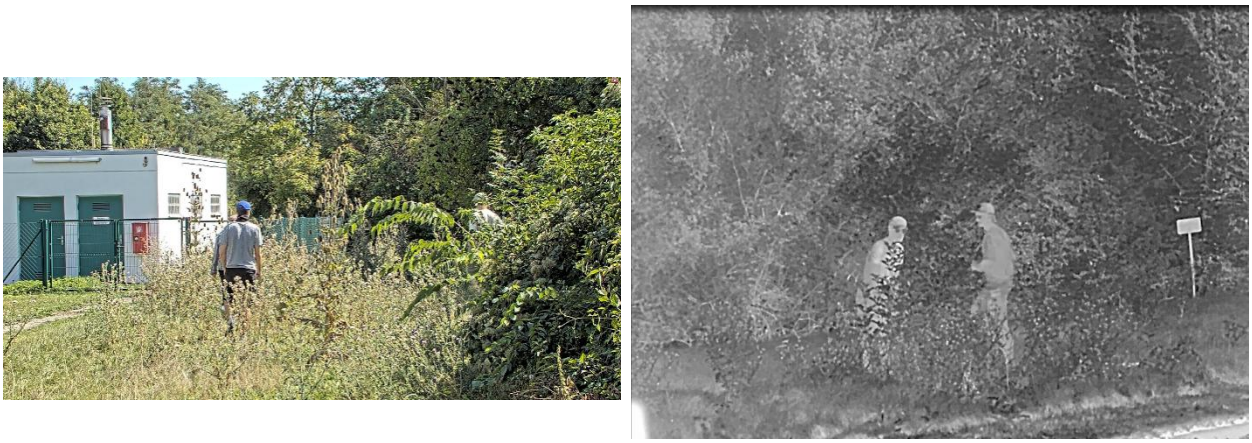

Figure 7: Persons (partially) occluded by tall grass

3. Persons with Varied Movements and Postures

Since most publicly available (LWIR) datasets contain persons walking in an upright position along different streets, persons in various positions, displaying varied movements and postures were recorded by the RGB and LWIR sensors within this scenario.

The recorded postures and movements contain crouching, sitting, jumping, lying down, standing on one leg, climbing trees and others (see Figure 9).

Figure 9:Persons in various positions in different scenes

4. Persons behind Smoke

To simulate persons within foggy natural environments, a fog machine was employed. Different scenes of persons walking and standing behind the fog were recorded with the RGB and LWIR sensors to investigate the visibility of persons behind smoke within the different wavelengths (see Figure 10). Due to the presence of wind, the density of the fog varied during the recordings.

Figure 10: Person behind Smoke

5. Persons at Night

To further explore the visibility of humans within the recordings of the RGB and LWIR sensors in diverse environments, night scenes of persons walking in nature were recorded.

Figure 11: Persons walking at night

In total 1.7GB of LWIR and 24.8GB of RGB videos of persons were acquired showing the previously introduced scenes. There are 11 individual recordings of different scenes and a total duration of raw thermal and RGB video data of 1,48 hours each. A selection of the recordings is in the annotation phase and will be utilized for algorithmic developments/evaluations in T10.2 of THEIA. The objects to be annotated are the persons shown within the scenes.

4.3 Software Desk-Research on Object Detection and Classification Algorithms

Considering the above-mentioned identified software requirement (person, vehicle, vessel detection) and chosen hardware, as well as a list of desired classes-to-be-identified provided by the end-users, SatCen, with a focus on vessel detection, a two-part desk research was conducted:

First Part Desk Research: Object Detection

Recent advances in deep learning have brought object detection to a high level of maturity and robustness (Trigka, 2025). Consequently, this research will concentrate on contemporary deep learning–based approaches and emerging methodological directions that are shaping the current and future development of object detection.

Two-stage detectors

The first approach for deep learning-based detection consisted of the following: for each image, a set of region proposals are generated. Then, representational features are extracted using Convolutional Neural Networks. Finally, a classifier is employed to assign the proposal's category labels. Key milestones of this approach have been achieved by (Girshick, Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, 2014) with the introduction of the **R-CNN** architecture, and improvements on this paradigm with **Fast-RCNN** (Girshick, Fast R-CNN, 2015) and **Faster R-CNN** (Shaoqing Ren, 2016).

One-stage detectors

A drawback of two-stage detectors is the slow processing speed for devices with limited storage and computational power. To overcome this disadvantage, another family of object detectors was introduced: one-stage detectors. This framework includes architectures that, by encapsulating all computation in a single network, directly predict class probabilities and bounding box offsets from images using a single CNN without the use of region proposal generation or post-classification refinements. Since the entire pipeline is comprised of a single network, its detection performance can be optimized end-to-end. The representative algorithm of this family is **YOLO** (Joseph Redmon, 2015). Since its introduction, multiple iterations of the YOLO family have been developed and maintained by Ultralytics (Ultralytics), each incorporating state-of-the-art techniques to improve accuracy, speed, and training stability. For example, the two more recent versions are YOLOv11 and YOLO26. **YOLOv11** was released on September 10,

2024, delivering excellent accuracy across different computer vision tasks such as object detection, image segmentation, pose estimation, tracking and classification. **YOLO26** was released at the beginning of 2026 and achieves the highest accuracy among the YOLO models on the COCO benchmark (Lin, 2015), with competitive processing speed. Another version worth mentioning which has not been developed by Ultralytics is **YOLOX** (Zheng Ge, YOLOX: Exceeding YOLO Series in 2021, 2021): this version provides competitive performance both in terms of accuracy and processing speed, while issuing the algorithm with an Apache-2.0 license which, compared to the AGPL-3.0 license, it a totally permissive license, making it usable for commercial purposes.

Transformer-based detectors

While originally introduced as a breakthrough in natural language processing, **Transformers** (Ashish Vaswani, 2017) have, in recent years, deeply affected the field of computer vision as well. The first researchers to adapt the transformer architecture to visual data were (Alexey Dosovitskiy, 2020) with the **Vision Transformer** (ViT). Instead of relying solely on convolutional layers, ViT divides input images into patches, serializes each patch into a vector, and processes them through a transformer-based encoder network as if they were token embeddings. Compared to CNNs, transformers applied to images have a better capacity to reason and exploit the relation between the object and the global image context. Building on this achievement, (Nicolas Carion, 2020) adapted the visual transformer specifically to the task of object detection and introducing **DETR** (Detection Transformer). The key innovation of DETR lies in its ability to simplify the object detection pipeline by eliminating many heuristic-based components like Non-maxima Suppression NMS and anchor boxes, which are commonly used in traditional object detection models. The main limitations with this approach were the slow training convergence and slow processing speed; these problems have been tackled by (Yian Zhao, DETRs Beat YOLOs on Real-time Object Detection, 2024) with the introduction of **RT-DETR** (Real-time DETR). This was the first transformer-based architecture that surpassed the YOLO algorithms both in terms of accuracy and speed on the COCO benchmark. An improvement on the RT-DETR architecture was achieved by **D-FINE DETR** (Yansong Peng, D-FINE: Redefine Regression Task in DETRs as Fine-grained Distribution Refinement, 2024) by reformulating bounding box regression to model localization uncertainty, reducing sensitivity to coordinate noise and accelerating convergence.

Open-Vocabulary Object Detection

One of the main limitations of the object detection paradigms discussed so far is their closed-set assumption: these networks are trained on a fixed set of object classes and can't recognize objects outside the categories seen during training. Open-Vocabulary Object Detection (Alireza Zareian, 2021) is a new paradigm that aims at extending object detection to an unbounded text vocabulary. Specifically, they first learn a visual-semantic embedding space using image-caption

data so that region features and caption words share a common representation. Then they train a two-stage detector (based on Faster R-CNN) using bounding-box annotations, while aligning the extracted region features with a visual–semantic embedding space learned from image–caption data. At inference time, the model can be given an image and an arbitrary text query, and it is expected to localize the region in the image corresponding to that text.

Follow-up work focused on improving language–vision alignment and localization quality. GLIP (Liunian Harold Li, 2022) unified object detection and phrase grounding for pre-training, showing that large-scale grounding supervision substantially improves open-vocabulary generalization. In contrast, other works such as OWL-ViT (Vision Transformer for Open-World Localization) (Matthias Minderer, 2022) adapted the ViT architecture to the open-vocabulary paradigm demonstrating that a transformer detector pretrained on image–text pairs can achieve strong zero-shot detection without explicit class-specific heads.

Recent state-of-the-art approaches, such as GroundingDINO (Shilong Liu, 2024) builds on the improved transformer architecture DINO (Mathilde Caron, 2021) and enhances it through grounded pre-training, integrating language directly into the transformer decoding process. On the other hand, YOLO-World (Tianheng Cheng, 2024) adapts the YOLO family to the paradigm of open-vocabulary recognition, enabling prompt-based detection at real-time speeds.

Evaluation

Given the vessel-detection use case, we conduct an object-detection evaluation focused on the *boat* category. The objective is to compare state-of-the-art detectors identified in the prior SOTA review and determine which methodology is best suited to this application.

To ensure a fair comparison, all closed-vocabulary models have comparable parameter counts and are pre-trained on the COCO dataset. This constraint does not apply to open-vocabulary detectors, which rely on distinct pre-training strategies involving large-scale, web-scraped datasets.

The evaluated models include:

- YOLOv11, YOLOv12, YOLO26, and YOLOX from the one-stage YOLO family.
- RT-DETR and D-Fine DETR from the transformer-based detector family.
- GroundingDINO and YOLO-World from the open-vocabulary detection paradigm.

Performance is assessed using Average Precision (AP) computed with the COCOeval API to quantify detection accuracy, alongside inference speed measured in frames per second (FPS) to capture computational efficiency.

The evaluation dataset is derived from the Pascal VOC (Williams, 2010) challenge by retaining only images annotated with the *boat* category. This yields a dataset of approximately 600 images encompassing diverse vessel types and operating contexts. Results are reported in Figure 12.

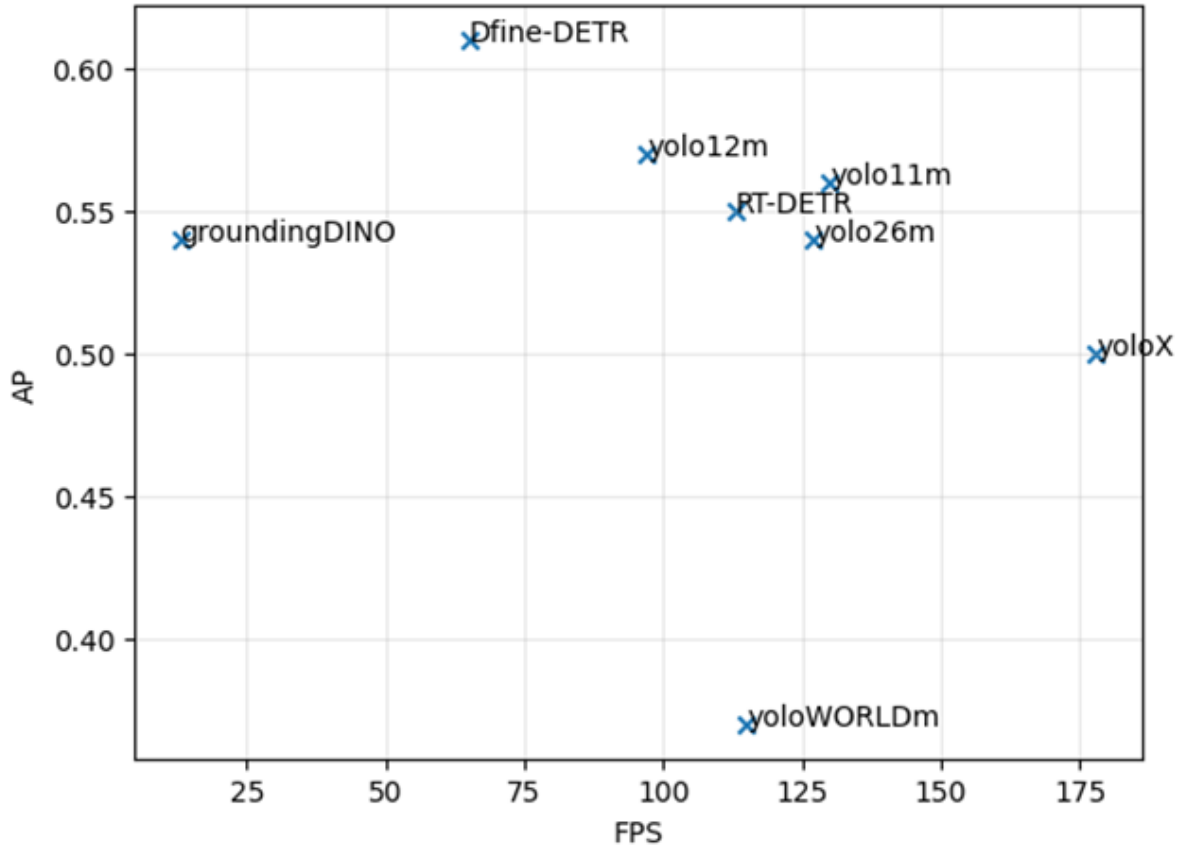

Figure 12: Evaluation of performance of different pre-trained object detectors on the boat category for the vessel detection use case

In general, the models form a Pareto-efficient curve that reflects the best achievable trade-offs between accuracy and speed. Moving along this curve, gains in frames per second (FPS) are obtained only by accepting a reduction in average precision (AP), and improvements in AP require sacrificing FPS.

The highest accuracy is achieved by Dfine-DETR, with an AP of 0.61, but this comes at the cost of a larger computational burden of 65 FPS. In contrast, YOLOX is by far the fastest model, reaching 178 FPS, while maintaining a slightly lower but still competitive AP of 0.50.

Models in the middle of the curve, including the latest YOLO variants, do not clearly outperform one another, and RT-DETR shows comparable performance within this group.

Another notable observation is that open-vocabulary object detection does not provide a clear advantage in this setting. Grounding DINO achieves a competitive AP but at an extremely low FPS, whereas YOLO-World offers competitive FPS but very low AP.

In conclusion, there is no clear out-of-the-box winner for boat detection. The choice of model should instead be guided by the target use case, such as prioritizing fast edge processing or maximizing detection accuracy.

Second Part Desk Research: Fine-Grained Classification for Objects of Interest

The object classes of interest were provided by the SatCen and consider more specific ("fine-grained") classes, than utilized within most popular benchmarks, for both land and marine vessels, with a focus on marine vessels.

METHODOLOGY

The proposed methodology for detection and fine-grained classification of specialized classes can work for both maritime vessels as well as ground vehicles. In the following the maritime use case is portrayed. The approach operates in two stages:

1. Spatial localization of vessels using an object detector
2. Semantic ("fine-grained") classification of the identified vessels using a vision-language model

RESEARCH

The following research focuses on vision–language models that can be applied to perform fine-grained classification on image crops identified by the object detection backbone. This family of architecture is designed to learn a shared semantic representation space for images and natural language. Their primary objective is to align visual content with textual concepts such that semantically related image–text pairs are mapped to nearby points in a common embedding space, while unrelated pairs are pushed apart.

The key outcome of this training paradigm is open-vocabulary recognition. Because visual features are aligned to linguistic concepts rather than fixed classes, the models can perform classification by comparing an image embedding against embeddings of arbitrary text prompts, without requiring task-specific retraining. This makes the family particularly suitable for fine-grained classification from detected image crops, where labeled data are scarce, class definitions are subtle, and category sets may evolve over time.

CLIP (Contrastive Language-Image Pre-training) (Alec Radford, 2021) was the first work to formalize and popularize the idea of using a vision–language dual encoder trained on image–text pairs specifically to enable zero-shot classification via prompt–image similarity. As training data, the authors used a large web-scraped image–text dataset with hundreds of millions of pairs filtered and structured to improve alignment quality. (Chao Jia, 2021) used the same approach but on a noisy web image–alt-text dataset of over one billion pairs with minimal filtering. The central claim is that sheer scale compensates for noise in the alt-text, enabling the model to learn strong visual and vision-language representations without expensive data cleaning or curation.

Building on the success of CLIP, **DeCLIP** (Data efficient CLIP) (Yangguang Li, 2022)**,** improves CLIP's data efficiency by augmenting the standard training paradigm with additional self-supervision, cross-modal multi-view supervision, and nearest-neighbor supervision to better exploit relationships within and across image–text pairs. **EVA-CLIP** (Quan Sun, 2023) also improves on the original CLIP by introducing more effective representation learning, optimization, and augmentation strategies, to achieve higher performance with substantially reduced training cost. Finally, **OpenCLIP** (Mehdi Cherti, 2022) is a popular open-source implementation of CLIP that, apart from providing reproducible models and training pipelines, it analyzes scaling laws for contrastive vision–language models using large public datasets.

SigLIP (Sigmoid Loss for Language Image Pre-Training) (Xiaohua Zhai, 2023) differs from standard contrastive learning with softmax normalization such as CLIP by introducing a sigmoid loss that operates solely on image-text pairs and does not require a global view of the pairwise similarities for normalization. This allows scaling up the batch size, while maintaining performance at smaller batch sizes. The second iteration of the SigLIP family, **SigLIP 2** (Michael Tschannen, 2025)**,** extends the original objective with a unified training recipe that combines captioning-based pretraining, self-supervised losses, and online data curation, resulting in an improved performance across all model scales. The SigLIP2 model is selected for experiments because it provides the most recent, strongest and most reliable image-text alignment while preserving the simple similarity-based inference used by CLIP.

5. Conclusions and Technical Observations

This deliverable has presented a structured technical assessment of UAS and terrestrial sensing components within the framework of WP7 of THEIA. In alignment with the amended Description of Action, the focus has been placed on characterising sensing capabilities, defining acquisition-stage metadata requirements for geo-referencing, and reviewing relevant Artificial Intelligence methodologies suitable for the THEIA use cases. The work reported herein establishes a technical baseline for non-space sensing assets considered within the project, without extending to system-level integration, operational validation, or demonstration activities.

The assessment covers both the UAS platform configured by C3I and the AIT terrestrial multi-sensor platform, each contributing distinct but complementary sensing capabilities within WP7.

From the perspective of airborne sensing, the assessment confirms that the selected UAS platform provides a robust technical foundation for geo-referenced multi-modal data acquisition. The combination of high-resolution optical and thermal imaging with RTK-enabled positioning allows the generation of spatially traceable sensing outputs accompanied by synchronised temporal and orientation metadata. The metadata schema defined in Chapter 3 demonstrates that all essential parameters required for accurate geo-referencing—such as latitude, longitude, altitude, timestamp, and platform attitude—are available at acquisition stage. This structured availability of metadata ensures that airborne observations can be precisely located in space and time, thereby supporting reproducibility and clarity of documentation. Within the scope of this deliverable, the UAS component has been assessed in terms of configuration and metadata readiness, without reporting operational flight campaign results or post-acquisition processing activities.

The terrestrial multi-sensor platform complements the airborne component by allowing flexible deployment in an area of interest. The data acquisition campaign conducted by AIT enriched an existing multi-modal dataset with additional RGB and LWIR imagery aligned with representative THEIA use-case scenarios. The structured dataset expansion provides valuable material for research-oriented analysis and supports exploration of object detection in diverse environmental contexts. The terrestrial platform therefore contributes not only sensing capability but also dataset diversification, strengthening the overall technical assessment of non-space sensing assets within WP7.

The desk-research-based review and evaluation of Artificial Intelligence methodologies, performed by AIT highlights important technical considerations for model selection within a THEIA use case context and the objects-of-interest. The assessment confirms that lightweight architectures can provide efficient inference performance suitable for resource-constrained environments, whereas larger transformer-based models achieve higher detection accuracy at the expense of increased computational demand. Vision-language models demonstrate potential

for enhanced semantic understanding, although their performance remains dependent on domain specificity and dataset characteristics. These observations underline that no single AI architecture is universally optimal; instead, model suitability depends on operational constraints, accuracy requirements, and computational resources. The AI-related evaluation presented in this deliverable is strictly limited to terrestrial datasets and does not include processing of UAS-derived imagery.

Taken together, the airborne and terrestrial sensing components demonstrate complementary characteristics. The UAS platform offers flexible, mobile, geo-referenced observation capability with high spatial accuracy and multi-modal sensing potential, while the terrestrial platform enables flexible deployment in an area of interest, multi-modal sensing and on-board AI-processing and geo-referencing.

It is important to emphasise that this deliverable is confined to sensing characterisation, acquisition-stage documentation, metadata specification, and AI methodology assessment. It does not address system-level integration architectures, automated data federation mechanisms, interoperability validation, or pilot deployment activities. The conclusions presented herein therefore relate to technical capability and research-oriented evaluation rather than operational performance validation.

In conclusion, Deliverable D7.3 establishes a coherent and amendment-aligned technical baseline for UAS and terrestrial sensing within WP7. By defining structured acquisition frameworks, specifying geo-referencing metadata requirements, and assessing relevant AI methodologies, the deliverable fulfils its objective of evaluating non-space sensing capabilities in support of THEIA's broader research activities.

References

- Project GA with No. 101190051
- THEIA Partners CA
- Alec Radford, J. W. (2021). Learning Transferable Visual Models From Natural Language Supervision. arXiv.
- Alexey Dosovitskiy, L. B. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv.
- Alireza Zareian, K. D.-F. (2021). Open-Vocabulary Object Detection Using Captions. CVPR 2021, (S. pp. 14393-14402).
- Ashish Vaswani, N. S. (2017). Attention Is All You Need. arXiv.
- Bustos, N. a.-Y. (2023). A Systematic Literature Review on Object Detection Using near Infrared and Thermal Images. Neurocomputing, 126804.
- Chao Jia, Y. Y.-T. (2021). Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision. arXiv.
- Danaci, K. I. (2024). A Survey on Infrared Image and Video Sets. Multimedia Tools and Applications, 16485--16523.
- FLIR, T. (2019). Teledyne FLIR.
- Gebhardt, E. a. (2018). Camel dataset for visual and thermal infrared multiple object detection and tracking. 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), 1--6.
- Girshick, R. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition.
- Girshick, R. (2015). Fast R-CNN. arXiv.
- JetPack, N. (kein Datum). https://developer.nvidia.com/embedded/jetpack.
- Jia, X. a. (2021). LLVIP: A visible-infrared paired dataset for low-light vision. Proceedings of the IEEE/CVF International Conference on Computer Vision, (S. 3496--3504).
- Joseph Redmon, S. D. (2015). You Only Look Once: Unified, Real-Time Object Detection. arXiv.
- Kuznetsova, A. a.-T. (2020). The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. In International journal of computer vision (S. 1956--1981). Springer.
- Lin, T.-Y. (2015). Microsoft COCO: Common Objects in Context. arXiv.
- Liunian Harold Li, P. Z.-N.-W. (2022). Grounded Language-Image Pre-training. arXiv.
- Mathilde Caron, H. T. (2021). Emerging Properties in Self-Supervised Vision Transformers. arXiv.
- Matthias Minderer, A. G. (2022). Simple Open-Vocabulary Object Detection with Vision Transformers. arXiv.

- Mehdi Cherti, R. B. (2022). Reproducible scaling laws for contrastive language-image learning. arXiv.
- Michael Tschannen, A. G. (2025). SigLIP 2: Multilingual Vision-Language Encoders with Improved Semantic Understanding, Localization, and Dense Features. arXiv.
- Nicolas Carion, F. M. (2020). End-to-End Object Detection with Transformers. arXiv.
- Noor Ul Huda, B. D. (2020). The Effect of a Diverse Dataset for Transfer Learning in Thermal Person Detection. Sensors 20.7.
- NVIDIA. (kein Datum). https://developer.nvidia.com/embedded/jetson-developer-kits.
- Pengfei Zhu, L. W. (2018). Vision Meets Drones: A Challenge. arxiv.
- Quan Sun, Y. F. (2023). EVA-CLIP: Improved Training Techniques for CLIP at Scale. arXiv.
- Radford, A. K. (2021). Learning transferable visual models from natural language supervision. Proceedings of the 38th International Conference on Machine Learning (ICML).
- Shao, S. (2018). CrowdHuman: A Benchmark for Detecting Human in a Crowd. arXiv.
- Shao, S. (2019). Objects365: A Large-scale, High-quality Dataset for Object Detection.
- Shaoqing Ren, K. H. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv.
- Sharma. (2018). Conceptual Captions 3M (CC3M).
- Shilong Liu, Z. Z. (2024). Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection. arXiv.
- TensorRT, N. (kein Datum). https://developer.nvidia.com/tensorrt.
- Tianheng Cheng, L. S. (2024). YOLO-World: Real-Time Open-Vocabulary Object Detection. arXiv.
- Trigka, Maria, and Elias Dritsas. "A comprehensive survey of machine learning techniques and models for object detection." *Sensors* 25.1 (2025): 214.
- Ultralytics. (kein Datum). https://github.com/ultralytics.
- Ultralytics. (2024). YOLOv11. https://docs.ultralytics.com/models/yolo11/.
- Ultralytics. (kein Datum). YOLOv8.
- Williams, M. E. (2010). The PASCAL Visual Object Classes (VOC) Challenge. International Journal of Computer Vision.
- Xiaohua Zhai, B. M. (2023). Sigmoid Loss for Language Image Pre-Training. arXiv.
- Xinyu Jia, C. Z. (2021). LLVIP: A visible-infrared paired dataset for low-light vision. Proceedings of the IEEE/CVF international conference on computer vision, 3496--3504.
- Yangguang Li, F. L. (2022). Supervision Exists Everywhere: A Data Efficient Contrastive Language-Image Pre-training Paradigm. arXiv.
- Yansong Peng, H. L. (2024). D-FINE: Redefine Regression Task in DETRs as Fine-grained Distribution Refinement. arXiv.
- Yian Zhao, W. L. (2024). DETRs Beat YOLOs on Real-time Object Detection. arXiv.

- Zareian, A. Y.-F. (2021). Open-vocabulary object detection using captions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (S. (pp. 14393–14402)).
- Zhang, W. a. (2022). Research on Camouflaged Human Target Detection Based on Deep Learning. Computational Intelligence and Neuroscience, 7703444.
- Zheng Ge, S. L. (2021). YOLOX: Exceeding YOLO Series in 2021. arXiv.
- Zhu, P. (2020). Detection and Tracking Meet Drones Challenge. arXiv.
- Zhu, P. a. (2021). Detection and Tracking Meet Drones Challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1-1.
- Zizhao Chen, Y. Q. (2025). AMFD: Distillation via adaptive multimodal fusion for multispectral pedestrian detection. IEEE Transactions on Multimedia.
- DJI (2024). Matrice 350 RTK Specifications. Available at: https://enterprise.dji.com/matrice-350-rtk/specs

END OF DOCUMENT