D1.2 Self-assessment & data management plan v1

Work Package:	WP1
Lead partner:	ENG
Author(s):	Gabriele Giunta (ENG), Emilia Gugliandolo (ENG), Irene Bicchierai (RESIL), Francesco Brancati (RESIL), Gerasimos Antzoulatos (CERTH), Ilias Gialampoukidis (CERTH), Vinod Ahuja (DFSL), Ilias Gkotsis (KEMEA), Galatea Kapellakou (KEMEA), Pantelis Velanas (ACCELI), Luigi Coppolino (CeRICT), Thomas Andrejak (CSNov), Gilles Lehmann (CSNov), Paulo Chaves (INOV), Helen Gibson (CENTRIC), Leonidas Perlepes (STWS), Dimitris Vamvatsikos (RG), Souzana Touloumtzi (NOA), Xavier Pothrat (CS), Mathieu Schmitt (SPACEAPPS), Leslie Gale (SPACEAPPS)
Due date:	M6
Version number:	1.0 **Status:** Final
Dissemination level:	Public

Project Number:	883284 **Project Acronym:** 7SHIELD
Project Title:	Safety and Security Standards of Space Systems, ground Segments and Satellite data assets, via prevention, detection, response and mitigation of physical and cyber threats
Start date:	September 1st, 2020
Duration:	24 months
Call identifier:	H2020-SU-INFRA-2019
Topic:	SU-INFRA01-2018-2019-2020 Prevention, detection, response and mitigation of combined physical and cyber threats to critical infrastructure in Europe
Instrument:	IA

Horizon 2020 European Union Funding for Research & Innovation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883284

Revision History

Revision	Date	Who	Description
0.1	29/01/2021	ENG	First release of the template
0.2	10/02/2021	ENG	First contribution
0.3	19/02/2021	ENG, RESIL, CERTH, NOA, CS	Assessing plan of the project objectives
0.4	23/02/2021	ENG, KEMEA, NOA, ACCELI, CERTH, INOV, DFSL, CSNov, CENTRIC, STWS	Integration of contributions; Dataset collection; Executive summary, Introduction and Conclusion.
0.5	24/02/2021	ENG, RG	Integration of contributions; Ready for peer review
1.0	26/02/2021	ENG	Final version after internal peer review

Quality Control

Role	Date	Who	Approved/Comment
Internal review	26/02/2021	CENTRIC	Accepted with minor changes
Internal review	26/02/2021	SPACEAPPS	Accepted with minor changes

Disclaimer

This document has been produced in the context of the 7SHIELD Project. The 7SHIELD project is part of the European Community's Horizon 2020 Program for research and development and is as such funded by the European Commission. All information in this document is provided 'as is' and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability with respect to this document, which is merely representing the authors' view.

Executive Summary

The project addresses the challenges associated with the security and resilience of EU Ground Segments of Space Systems. Today, the ground segments of space systems receive massive amounts of satellite data. A physical/cyber-attack to their installations or communication networks, respectively, would cause debilitating impact on the distribution of satellite data and in its data storage, access and exchange affects not only the reliability of space data, but also their FAIR standards: findability, accessibility, interoperability and reusability. To reach this goal, the project makes use of advanced technologies for data integration, processing, analytics and visualisation as well as data security and cyberthreat protection to assess the prevention, detection and mitigation of threats, both physical and cyber. Moreover, running of the project requires the collection and generation of data to achieve the main project objectives and to directly support the technical development of 7SHIELD tools or the development of operational processes. This data is required to manage the project, disseminate the information about it, analyse and exploit its results. Moreover, the technology-oriented WPs will process both open and closed source data. Collecting and generating of the mentioned data requires compliance with data management strategies suggested by the EC. In accordance with the guidelines on data management in Horizon 2020, a data management plan is required to monitor the collected/generated data with respect to their privacy and confidentiality, ensure that the legal and potential ethical standards for data generation, use, storage and share are applied throughout the project.

This deliverable is the first version of the Data Management Plan (DMP) of the 7SHIELD project, project, i.e. a short, general outline of the project policy for data management. The described policy reflects the current state of consortium agreements regarding data management and is consistent with those referring to the exploitation and protection of results. The 7SHIELD Initial Data Management Plan (DMP) has the objective to detail specifics of data which have already been collected/generated (or are foreseen to be collected/generated) during the lifespan of the project. In particular, it includes a summary of the data and how they will be FAIR (i.e., Findable, Accessible, Interoperable, and Re-usable) based on the "Guidelines on FAIR Data Management in Horizon 2020". The overall purpose of the DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project.

The DMP is a living document which will be kept updated during the whole lifetime of the project, since data generation and collection, and therefore data management, will be active in 7SHIELD for a considerable time after the submission of the initial version of the Data Management Plan. The datasets may also be altered due to converging factors, such as project maturity, shifts in consumer usage, legislative changes, etc. This document can evolve along the project and it will be updated in D1.4 Self-assessment & data management

plan v2 (M18). The update will include new sets of data and changes in consortium policies and datasets management.

Table of Contents

List of figures

List of Tables

Definitions and acronyms

AUC	Area Under Curve
C2	Command and Control
CA	Consortium Agreement
CAP	Common Alerting Protocol
CAPEC	Common Attack Pattern Enumeration and Classification
CI	Critical Infrastructure
CIP	Critical Infrastructure Protection
COCO	Common Object in Context
COTS	Common-off-the-shelf
C/P	Cyber/Physical
CSV	Comma-Separated Values
CVE	Common Vulnerabilities and Exposures
CVS	Concurrent Versions System
CWE	Common Weakness Enumeration
DB	Database
DMP	Data Management Plan
DoA	Description of Action
EC	European Commission
ECSCI	European Cluster for Securing Critical Infrastructures
EDXL	Emergency Data Exchange Language
EU	European Union
EUCI	EU-classified information
FAIR	Findable, Accessible, Interoperable and Re-usable
FAQ	Frequently Asked Questions
FLIR	Forward Looking InfraRed
FOAF	Friend of a friend
FPR	False Positive Rate
FPS	Frames Per Seconds
GA	Grant Agreement
GDPR	General Data Protection Regulation
HTML	HyperText Markup Language
IA	Innovation Activity
IMA	Impact-making Activity
IMO	Impact-making Objective
IO	Innovation Objective
IP	Intellectual Property
IPR	intellectual property rights
IVUL	Image and Video Understanding Lab
JSON	JavaScript Object Notation
KAUST	King Abdullah University of Science and Technology
NIR	Near-Infrared

NVD	National Vulnerability Database
OGC	Open Geospatial Consortium
ORD	Open Research Data Pilot
OWL	Web Ontology Language
PC	Project Coordinator
PCAP	Packet Capture
PDF	Portable Document Format
SC	Scientific Coordinator
SGS	Satellite Ground Station
SOSA	Sensor, Observation, Sample, and Actuator
SPGU	Situational Picture Generation and Update
SSN	Semantic Sensor Network
TM	Technical Manager
TPR	True Positive Rate
UA	User-oriented activity
UAF	Unified Alert Format
UAV	Unmanned Aerial Vehicle
UCF	University of Central Florida
UO	User-oriented Objective
VOC	Visual Object Classes
VPN	Virtual Private Network
WP	Work Package
XML	eXtensible Markup Language

1. Introduction

This deliverable is the first version of the Data Management Plan (DMP) for the 7SHIELD project. Data of different nature will be collected, processed, and generated during the lifetime of the 7SHIELD project. Some of these data might contain personal information and thus require a clear data management plan on how they are to be handled, i.e., stored, processed, accessed, and protected against unauthorised or improper use, etc. The first version of this deliverable is submitted on the sixth month of the project follows the template provided by the European Commission[1]. In particular, this report provides an analysis of the main elements of the data management policy that will be used by the Consortium with regard to all the datasets that will be generated by the project, describing rules, best practices and standards that will be used with regard to the datasets preparation, cleansing and processing, including data analysis and analytics. The deliverable includes information related to accessibility, intelligibility, usability and interoperability of the data gathered and takes into account privacy and security aspects.

This document can evolve along the project and it will be updated in D1.4 Self-assessment & data management plan v2 (M18). The update will include new sets of data and changes in consortium policies and datasets management.

1.1. Purpose of the document

The overall purpose of the DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project. It will contribute to the management of data in the project through the following steps:

- Outline the types of data collected and generated (or foreseen for collection and generation) during the course of the 7SHIELD project;
- Describe the methodology and standards required, but also identify whether and how data will be collected, shared, exploited, re-used or made accessible for verification, and how they will be curated and preserved;
- Specify the degree of privacy and confidentiality of the collected/generated data;
- Outline the considerations and measures that are foreseen for the adequate management of the data from the legal, ethical, and security points of view;
- Outline the main elements of the data management policy that will be followed by the 7SHIELD consortium to handle collected/generated data with respect to their sensitiveness during and after the project;
- Ensure project research data and records are accurate, complete, authentic, interoperable and reliable;

[1] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

- Enhance data security and thereby minimize the risk of data loss;
- Ensure research integrity and reproducibility by others,

The described policy reflects the current state of consortium agreements regarding data management and is consistent with those referring to exploitation and protection of results.

1.2. Structure of the document

This deliverable is structured as follows:

- Section 2 defines the assessing plan of the project objectives. The specific objectives for the project as described in section 1.1 of the DoA are listed and the work carried out towards the achievement of each listed objective has been assessed, considering also the expected KPIs and target value.
- Section 3 defines the data management structure to be followed during the 7SHIELD project according to the Guidelines on FAIR Data Management in Horizon 2020.
- Section 4 describes the information about 7SHIELD datasets: at this early stage of the project, 18 datasets have been identified.
- Section 5 presents IPR plan that focuses on the careful handling of IPR issues in 7SHIELD project.
- Section 6 presents the main conclusions and the references to the documentation sources used in this deliverable are included in the reference section.

2. Assessing Plan of the project objectives

7SHIELD addresses the security and resilience of EU Ground Segments of Space Systems, meeting the crosscutting and the sectoral criteria of the EU critical infrastructures (EC Directive 2008/114). The project aims at providing a holistic framework enabling the deployment of innovative services for cyber-physical protection of ground segments. The framework will enhance the infrastructures' protection capabilities, while integrating or interoperating with existing protection solutions already deployed. 7SHIELD resolves some Innovation, User-oriented and Impact-Making Objectives, as shown in the Figure 2-1.

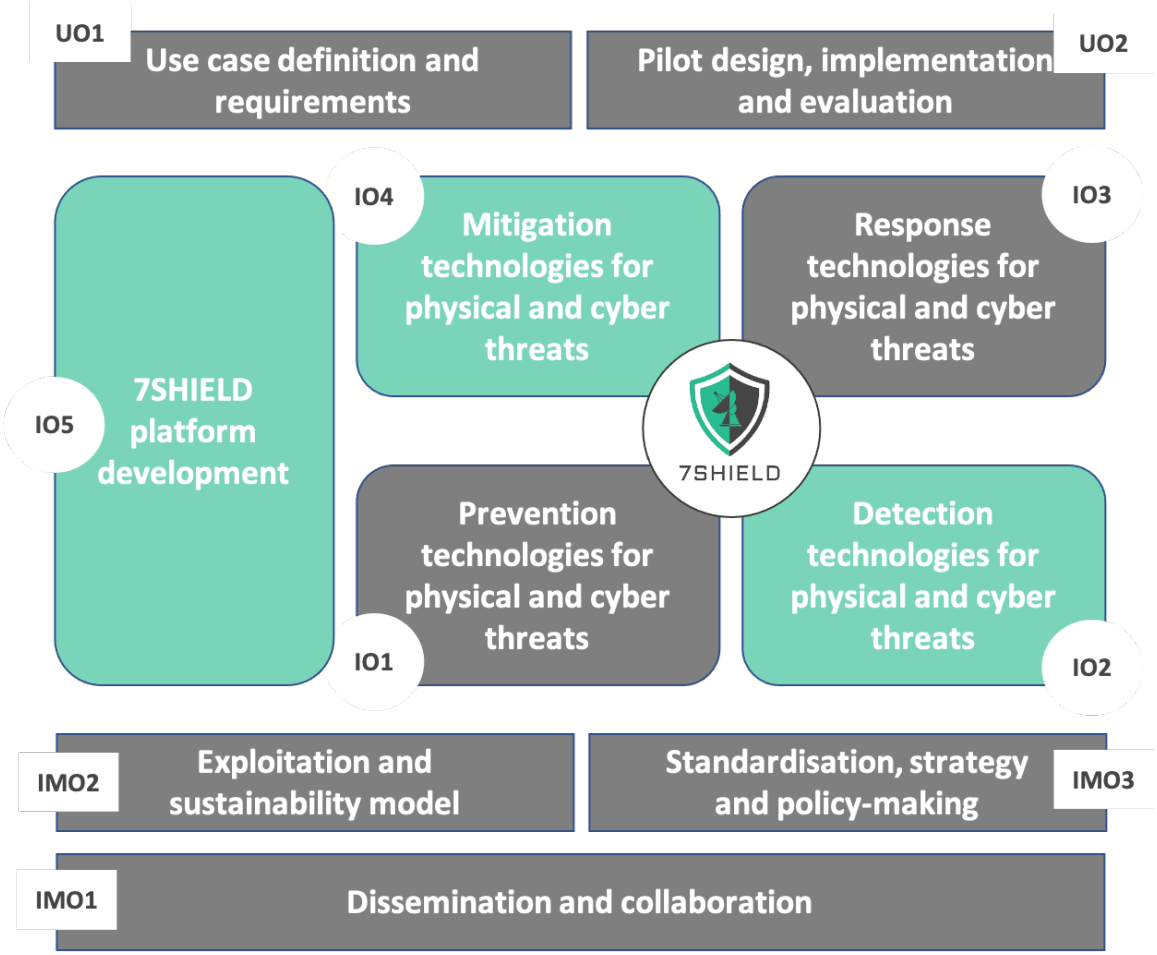

Figure 2-1: 7SHIELD objectives

Each objective includes some specific activities. To follow the progress of the project activities on a regular basis and better understand potential difficulties, the deliverable reports the achievements and the activities carried out for each objective and tries to quantify, and thereby measure, a number of KPIs related to each activity and objective. KPIs ensure project management effectively and efficiently monitor the project evolution and progress towards such objectives. 7SHIELD has identified a number of KPIs related to each objective and activity to ensure the highest impact, as well as the quality and success of the expected outputs. Some activities have reached more than one achievement in the first six

months of the project while others have just started. This document is constantly evolving so that progress can be taken into account in relation to each objective.

2.1. Innovation objectives (IOs) and innovation activities (IAs)

IO1. Prevention technologies for physical and cyber threats: 7SHIELD will assess the vulnerability of each asset of the space ground segment, whether it is some mechanic component or data asset. The foreseen technologies which contribute to the pre-crisis phase involve also analytical prediction models for future threats, secure authentication mechanism for data access and cascade effects from combined cyber-physical attacks.

In this period, the main achievements are related to the **vulnerability estimation for risk assessment** activities and in particular to the requirement definition for the CIRP platform and to the definition of a concreate roadmap for its integration with the technology for assessing cascading effect from physical and cyber-attacks. Moreover, the SSO functionality and components to be developed for a **secure authentication mechanism** have been established. Regarding the **cascading effects**, research on threat modelling and cascading risk assessment methods has been conducted and an analysis of the existing threats and attack paths modelling methods and tools to define a first conceptualization of the cascading effect. The achievements related to the **Cyber and Physical Threat Intelligence** have produced an analysis of the different data source to identify cyber and physical threats along with an understanding of the main capabilities of the current solutions.

IO2. Detection technologies for physical and cyber threats: 7SHIELD will encompass state-of-the-art detection technologies to seamlessly and accurately identify potential physical or/and cyber threats to Space Systems, ground Segments and Satellite data assets.

Regarding the **data acquisition and pre-processing methodologies**, different operational scenarios have been analysed and further details on the ground segment infrastructures have been gathered. Moreover, the best materials and protection of UAV has been conducted and a first parametric configuration and the identification of most appropriate cameras and possible other sensor for the algorithms have been provided.

For **video surveillance technologies**, the functional requirements for the Face Detection and Face Recognition module as well as for the Video-based Object Detection and Activity Recognition module have been specified along with the first version of the architecture for face recognition. Moreover, the dataset to be used for object detection have been searched, gathered and prepared.

Cyber-attack detection main achievements are related to the selection of relevant sensors and data flows useful for the validation of the cyber-attack detection methods and a first list of sensors and probes has been prepared.

For **thermal and near-infrared image processing**, a study on IR camera and a definition of preliminary data formats and the API interface for the camera have been conducted.

Regarding the **Innovative Laser-based technologies**, for PLS, LFS and 3D MND sensors, JSONs samples have been completed, upgrades on hardware, software and algorithms have started. In general, main improvements have been done on these sensors.

For **Physical and Cyber Threat Detection and Early Warning**, the Detection and correlation architecture has been designed and the correlation process has been shared.

IO3. Response technologies for physical and cyber threats: 7SHIELD will tailor innovative technologies to monitor the evolvement of a physical or/and cyber-attacks, to strengthen the responsiveness and social awareness.

Most of the activities related to this objective have just started. The main achievements reached are related to the **semantic representation and linking for reasoning and decision-making**. In particular, related ontologies and standards have been evaluated to be adapted to the 7SHIELD needs. An example ontology (OWL, Protégé) has been launched.

IO4. Mitigation technologies for physical and cyber threats (including novel installation designs): 7SHIELD will consolidate the appropriate actions to mitigate the consequences of physical and cyber-attacks, focusing on the services continuity.

All the activities related to this objective have just started so, the main achievements will be reported in the next deliverable but annotated during the progress of the project.

IO5. 7SHIELD platform development: This objective will deal with the design of 7SHIELD platform's architecture and integration of all subsystems.

7SHIELD platform integration activities will start at M7. In the meantime, 7SHIELD architecture has been defined.

The **Data Models** considered in the main 7SHIELD components and a first set of known ontologies have been analysed. The design of a Situational Information Model has started and a format to be to be adopted for the exchange of information related to alerts, threats, and combined threat scenarios has been established.

Activities related to the **user interfaces implementation** will start at M7.

2.2. User-oriented objectives (UO) and user-oriented activities (UA)

UO1. Use case definition and requirements: In 7SHIELD the systematic approach/study of the specific user requirements will bring out the general rules and procedures to consolidate them into a standardized framework which ensures the safety and security of Space Systems, ground Segment and Satellite data assets.

Use case design, stakeholder engagement and user requirements activities resulted in the definition of five use cases and 19 user scenarios of cyber, physical and combined cyber-physical attacks. 10 focus groups have been implemented by the Pilot Use Case Leaders for the design of the use cases and scenarios, with the participation of end-users from the Ground Segment operations teams, asset security management professionals, and first responders. Diverse end-users and stakeholders were engaged and participated in the use cases design and collection of user requirements. 16 questionnaires were answered by Ground Segment professionals, critical infrastructure protection experts, and first responders. At the end, 250 functional and non-functional user requirements were defined by the end-users and stakeholders.

The general **security requirements** and principles that will guide the development of the 7SHIELD system and its modules have been defined and established. More than 40 security requirements have been defined, including measures for access control, secure user authentication, traffic monitoring and encryption, data integrity, minimization of vulnerabilities, and secure backups.

Activities to ensure compliance with **legal and ethical requirements** have been take place

UO2. Pilot design, implementation and evaluation: The 7SHIELD needs to be evaluated and testing within the operational environment to probe its performance to cope with real hazardous situations in Critical Infrastructures.

All the activities related to this objective will start at M8, the main achievements will be reported in the next deliverable but annotated during the progress of the project.

2.3. Impact-making objectives (IMO) and impact-making activities (IMA)

IMO1. Dissemination and collaboration: In the context of 7SHIELD we aim at disseminating the project results with an emphasis in wide number of actors in the whole security management cycle, governmental authorities and academic institutes, as well as to the security agencies. The project aims at establishing close collaborations with existing projects working in similar research domains and external bodies.

For the **dissemination and communication of the project results**, the National/Regional Space Agencies community, the Ground segment operators' community and the Security experts community have been approached through the participation to several events.

Regarding the **collaboration and clustering**, 7SHIELD has integrated the Critical Infrastructure Protection community as it is now a member of the European Cluster for Securing Critical Infrastructures (ECSCI).

IMO2. Exploitation and sustainability model: This specific objective is linked to the achievement of an exploitable and sustainable model of 7SHIELD results and solutions at different levels as key to the success of the project vision.

The demonstrations are planned later in the project. The main achievements will be reported in the next deliverable but annotated during the progress of the project.

IMO3. Standardisation, strategy and policy-making: This objective deals with the 7SHIELD's aims to standardize and demonstrate strategies and policies to prevent, early detect, response and mitigate of amalgamated attacks in physical and cyber manner. The IMO3 is decomposed into the following activities.

For the **policy framework, standardisation, strategy and policy-planning**, the European Ground segment security policy and standards community has been approached.

2.4. Project Objectives

In this section are listed the specific objectives and activities for the project as described in the DoA. The work carried out, during the first six months, towards the achievement of each objective and the assessment of the progress in relation to the specific activities, considering also the expected KPIs, is described in detail.

2.4.1. Innovation objectives (IOs) and innovation activities (IAs)

Table 2-1: 7SHIELD Innovation objectives and innovation activities achievements

Innovation Objectives	Innovation Activities	KPIs
IO1. Prevention technologies for physical and cyber threats	**IA1.1** Vulnerability estimation and classification per asset for risk assessment (KR01)	**KPI A1.1.1** Integrated Scientific Models **KPI A1.1.2** Ingested datasets size
	IA1.2 Secure authentication mechanism for data access (KR02)	**KPI A1.2.1** Success in authentication/authorisation attempts according to the

		different user identity profiles.
	IA1.3 Cascading effects from physical and cyber-attacks due to their interdependencies (KR03)	**KPI A1.3.1** Number of identified threats due to cascading effects identified in pilot sites;
	IA1.4 Cyber and Physical Threat Intelligence (KR04)	**KPI A1.4.1** Accuracy, Error rate.
Achievements		
IA1.1 *In this period, the activities have been focused to the definition of the requirement for the porting CIRP platform provided by STWS. Details about this platform and the methodology has been presented to the WP3 Team, furthermore a roadmap for the integration of this tool with the technology for assessing cascading effect from physical and cyber-attacks have been drafted.*		
IA1.2 *In this time frame, we achieved the following results: i) Pilots specific needs and requirements have been collected by means of questionnaires, and then reviewed by technical partners. This paved the way for the kick-off of the definition of SSO functionality and components to be then developed in the project, and a first draft of a technical document has been shared with project partners. Moreover, the deployment infrastructure (e.g. Cloud Provider, Container orchestration system) has been chosen.*		
IA1.3 *In this period, the technical activities aiming at achieving this KPI have delivered a research of state of the art and requirements on threat modelling and cascading risk assessment methods and tool and an analysis of the existing threats and attack paths modelling methods and tools, focusing on the identification of issues and the definition of requirements for modelling complex dependencies. A preliminary conceptualization of the cascading effect issue and known modelling solutions to address such concepts have also been devised.*		
IA1.4 *The research belonging this innovation activity have produced the following achievements: i) Analysis of the different typology of data source could be relevant to identify cyber and physical threats in the context of the project domains, ii) Study of the main capabilities integrated in the current available solutions, both open source and commercial.*		
Innovation Objectives	**Innovation Activities**	**KPIs**
IO2. Detection technologies for physical and cyber threats	**IA2.1** Data Acquisition and pre-processing methodologies at the edge (KR05)	**KPI:** Duration of continuous autonomous operation of each type of agent in one battery charge; **KPI:** Amount of time needed to perform surveillance coverage

		mission, examining cooperative navigation and control scenarios; **KPI:** Size of monitored area per agent (for multi-agent mission) during 24h.
	IA2.2 Video surveillance technologies for physical attacks (KR06-KR07)	KPI: Accuracy and detection latency. For detection accuracy: False Positive Rate (FPR), True Positive Rate (TPR) and Area Under Curve (AUC) For detection latency: Frames per seconds (FPS).
	IA2.3 Cyber-attack detection mechanism (KR08)	KPI1: # of cyber attacks with high impact (based on technical/scientific literature) detected; KPI2: # of misuse cases with high impact (based on technical/scientific literature) detected; KPI3: Performance penalty of TE technology.
	IA2.4 Thermal and near-infrared image processing for man-made threats detection (KR09)	KPIs: Classical detection measures (Recall, Precision, F1-Measure) and tracking measures and real-time performance measures.
	IA2.5 Innovative Laser-based technologies for the detection of ground-based and aerial threats detection (KR10)	KPIs: Taking pictures of intruders (human, vehicle and drone), using slaved PTZ camera, and following up throughout the track.
	IA2.6 Combined Physical and Cyber Threat Detection and Early Warning (KR11)	KPIs: Detection of the artificially added threat data in the "normal" logs.
Achievements		
IA2.1 *In this period the technical activities for the IA2.1 have been focused on the following actions:* • *ACCELI organised an internal specification workshop of adapting ACCELI UAVs on 7SHIELD use case.* • *Analysis of different operational scenarios have been realised.*		

- *A questionnaire to the 7SHIELD end users has been prepared and shared in order to obtain details related to the PUCs and ground segment infrastructures.*
- *Concerning the UAV Payload, ACCELI has conducted a literature survey on frame materials and protection of UAV sensitive parts from harsh environments (low temperatures, rain, snow).*
- *A preliminary parametric configuration of NVIDIA Jetson GPU (standalone operation) as well as systemic configuration of ACCELI UAV PIXHAWK 4 chipset with Arduino chipset (successful collaboration of both subsystems) and a 1st operational validation with embedded intelligence have been carried out.*
- *Market survey for the identification of most appropriate cameras and possible other sensors, considering the operational needs of CERTH algorithms.*
- *Specification of autonomous flight functionalities considering use case scenarios as they have been described in DoA, has been realised.*

IA2.2

In this period the technical activities for the IA2.2 have been focused on the following actions:

- *Specified the functional requirements for the Face Detection and Face Recognition module as well as for the Video-based Object Detection and Activity Recognition module*
- *Completed data collection from surface web (public datasets) for face detection and recognition.*
- *Produced the first draft of face recognition output in JSON format through collaboration with other WP4 partners.*
- *Designed the first version of face recognition module architecture to fit user requirements and use case scenarios.*
- *Research over the dataset to be used for object detection. Study of the object or interest taking into account the user requirements.*
- *Initial filtering of the activities to be monitored.*
- *Gathering and preparing of datasets to be used for training.*

IA2.3

During this period the following tasks have been carried out as far as IA2.3 is concerned:

- *Questionnaires on use cases and D2.1 are being analysed for selecting relevant sensors and representative data flows useful for the validation of the cyber-attack detection methods*
- *A list of sensors and probes has been prepared and features of these components have been specified*
- *An adaptation of the SIEM system has been done in order to translate its output into the IDMEF format*
- *Additional operators have been implemented in order to allow SIEM to manage actuators and sensors availability*

IA2.4

For the IA2.4, the technical activities in the reported period included the:

- *Study, definition and acquisition of the IR camera*
- *Definition of preliminary data formats to interface with reaming 7Shield architecture*
- *Implementation API interface for the camera*

- *Launch of the definition of the MMAS architecture and identification of datasets for training*

IA2.5

For IA2.5, the task started at M3. For all 3 sensors - PLS, LFS and 3D MND - following tasks undertaken during the referenced period (M3-M6):

1. *JSONs samples for 7SHIELD control station / platform requirements have been completed and sent for study at WP level for T4.7. Further detector survey requirement for T4.7 submitted.*
2. *Upgradation actions on hardware, software and algorithms are in progress.*
3. *Housings, lasers and lenses are being upgraded with fresh design and innovation to meet 7SHIELD technical requirements, and to withstand weather condition requirements.*

IA2.6

For the IA2.6, the technical activities in the reported period focused on the:

- *Design of the Detection and correlation architecture*
- *Delivering the Detector questionnaire filled by WP4 partners*
- *Sharing the detection and correlation process between partners*
- *Complete of the first draft of unified format alert definition*

Innovation Objectives	Innovation Activities	KPIs
IO3. Response technologies for physical and cyber threats	IA3.1 Semantic representation and linking for reasoning and decision-making (KR12)	KPIs: Quality (e.g. Content Quality Metric, Structural Quality Metric [2] and completeness metrics will be applied in the ontology. Response time will be computed in the population tool. Accuracy and precision will be calculated in the reasoning process.
	IA3.2 Crisis level classification from multimodal data fusion (KR13)	KPIs: Precision and accuracy in the crisis level estimation.
	IA3.3 Decision Support mechanism (KR14)	KPIs: Quickness and quality of information provided and calculated in the reasoning process.
	IA3.4 Social awareness and interaction with the citizens (KR15)	KPIs: user acceptance rating during pilot testing and debriefing. Increase engagement with messages (likes, shares, comments, replies, link follows, etc.)
	IA3.5 Intruding UAV neutralisation (KR16)	KPIs: Flying Hunter flies to the intruding drone on the command of the operator,

		homing on to the drone, catching the drone and bringing it back to designated ground area
Achievements		
IA3.1 *Our progress the past few months concerning the IA3.1 can be summarized in the following action points:* • *Related ontologies and standards in7SHIELD domain of interest (e.g. SSN, SOSA, beAWARE, WADM) have been reviewed and evaluated, so they can be reused with the necessary modifications to satisfy the 7SHIELD KB needs.* • *Model requirements elicitation through interaction with technical partners and end-users (questionnaire) and clarification of input data format (JSON).* • *Example ontology (OWL, Protégé) has been launched to be further developed, based on so far user requirements, specifications and discussions (output samples, PUCs).*		
IA3.2 *Activity started in M6*		
IA3.3 *Activity started in M6*		
IA3.4 *Activity started in M6*		
IA3.5 *Activity started in M6*		
Innovation Objectives	**Innovation Activities**	**KPIs**
IO4. Mitigation technologies for physical and cyber threats (including novel installation designs)	**IA4.1** Development of service continuity scenarios for cyber-attacks (KR17)	**KPI**: Downtime of critical services.
	IA4.2 Development of service continuity scenarios for physical attacks (KR17)	**KPIs**: 7SHIELD service continuity planning will focus on ensuring that the critical services, as will be defined by the Ground Space Segment Operators (WP5, T5.4), will be delivered throughout the physical crisis under discussion (WP5, WP7), and that the minimum Acceptable Downtime of critical services is achieved.
Achievements		
IA4.1 *Activity started in M6*		
IA4.2 *Activity started in M6*		

Innovation Objectives	Innovation Activities	KPIs
IO5. 7SHIELD platform development	**IA5.1** 7SHIELD platform integration (KR18)	KPI: 7SHIELD modules integrated and deployed in the Framework
	IA5.2 Data Models for Combined Detection (KR19)	KPI: Semantic concept defined
	IA5.3 User interfaces/Command and Control (C2) (KR20)	**KPI**: Common Operational Picture refresh updates
Achievements		
IA5.1 *The 7SHIELD platform integration activities will start at M7.* *The 7SHIELD architecture has been defined and it is the starting point for the 7SHIELD platform integration activities.*		
IA5.2 *The data models considered in the main 7SHIELD components were analysed to find potential points of interaction between them.* *A first set of known ontologies was analysed with the aim to identify those that can be adopted in the context of 7SHIELD.* *A first draft version of an ontology to be used in the context of WP4 was developed.* *The Unified Alert Format (UAF) was defined to be adopted in 7SHIELD for the exchange of information related to alerts, threats, and combined threat scenarios.* *The design of a Situational Information Model has started.*		
IA5.3 *Activities related to the user interfaces implementation will start at M7. Therefore, no updates are available for this objective.*		

2.4.2. User-oriented objectives (UO) and user-oriented activities (UA)

Table 2-2: 7SHIELD User-oriented objectives and user-oriented activities achievements

User-oriented Objectives	User-oriented Activities	KPIs
UO1. Use case definition and requirements	**UA1.1** Use case design, stakeholder engagement and user requirements	KPIs: User-defined requirements that are clear and broad enough in order to ensure that all stakeholders' needs are met. At least 15 questionnaires are answered by ground stations professionals from at least 5 independent organizations. At least 3 focus groups are implemented and at least 15 user scenarios are proposed.

	UA1.2 Security requirements	KPIs: Secure access to the system, secure communications.
	UA1.3 Ethics and legal framework	KPIs: Demonstrate that research activities and expected results respect and promote the European Convention on Human Rights and the EU's Charter of Fundamental Rights and enhance European and local values, in accordance with the public sense of fairness.

Achievements

UA1.1

The user-oriented activities implemented during the first 6 months resulted in the following achievements:

- *Five use cases of five different EU countries (Finland, Spain, Greece, Belgium and Italy) have been thoroughly designed.*
- *A total of 19 user scenarios of cyber, physical and combined cyber-physical attacks have been proposed.*
- *10 focus groups were implemented by the Pilot Use Case Leaders for the design of the use cases and scenarios, with the participation of end-users from the Ground Segment operations teams, asset security management professionals, and first responders.*
- *Diverse end-users and stakeholders were engaged and participated in the use cases design and collection of user requirements, including Ground Segment facility managers, IT managers and data center technical managers, Ground Segment engineers, operators, cybersecurity engineers, telecommunications systems engineers, system administrators, DevOps engineers, cybercrime experts, critical infrastructure protection experts, ONDA DIAS infrastructure operations managers, ICE Cubes operators, and Ground Segment first responders.*
- *16 questionnaires were answered by Ground Segment professionals, critical infrastructure protection experts, and first responders.*
- *250 functional and non-functional user requirements were defined by the end-users and stakeholders that participated in the user requirements survey.*

UA1.2

During the first 6 months, the general security requirements and principles that will guide the development of the 7SHIELD system and its modules were defined and established, with the guidance of KEMEA's critical infrastructure protection experts. The security requirements have been communicated to the technical partners and aim to ensure the development of a cyber-secure system, limiting the risks of data breach and ensuring secure data exchange and storage. More than 40 security requirements have been defined, including measures for access control, secure user authentication, traffic monitoring and encryption, data integrity, minimization of vulnerabilities, and secure backups.

Secure access to the system and secure communications will be tested during the pilot implementation activities.
UA1.3 *The user-oriented activities implemented during the first 6 months have been closely monitored by CENTRIC to ensure compliance with legal and ethical requirements. User and stakeholder engagement and participation in the research activities has taken place in accordance with GDPR and human rights.*

User-oriented Objectives	User-oriented Activities	KPIs
UO2. Pilot design, implementation and evaluation:	**UA2.1** Development of the validation scenario and evaluation methodology	KPIs: Evaluation metrics, User satisfaction metrics, user feedback, system usability metrics.
	UA2.2 Field demonstrations, testing and training	KPIs: User satisfaction metrics, user feedback, system usability metrics.
Achievements		
UA2.1 *The reporting period does not cover these activities, which will start in M8.*		
UA2.2 *The reporting period does not cover these activities, which will start in M8.*		

2.4.3. Impact-making objectives (IMO) and impact-making activities (IMA)

Table 2-3: 7SHIELD Impact-making objectives and impact-making activities achievements

Impact-making Objectives	Impact-making Activities	KPIs
IMO1. Dissemination and collaboration	**IMA1.1** Dissemination and communication of the project results	KPIs: At least two domain-specific communities for dissemination and clustering.
	IMA1.2 Collaboration and clustering with other SU-INFRA-01 projects	KPIs: At least two domain-specific communities for dissemination and clustering.
Achievements		
IAM1.1		

- *The **National/Regional Space Agencies community** has been approached with the integration of 2 people in the Advisory Board coming from National Space Agencies (Maria de Fatima Mattiello Francisco from INPE and Julien Airaud from CNES) and the participation to several ESA events (Phi Week and ESA Ground Segment security policy and standards).*
- *The **Ground segment operator community** has been approached through the participation to several events (ESA Ground Segment security policy and standards, GROUND SYSTEM ARCHITECTURE WORKSHOP)*
- *The **Security expert community** has been approached through the participation to one event (Nicosia Risk Forum 2020)*

IMA1.2

*7SHIELD has integrated the **Critical Infrastructure Protection community** as it is now a member of the European Cluster for Securing Critical Infrastructures (ECSCI).*

Impact-making Objectives	Impact-making Activities	KPIs
IMO2. Exploitation and sustainability model	**IMA2.1** Market analysis and existing business models	KPIs: Demonstrations to at least two other external installations and comparison.
	IMA2.2 Exploitation plan and Intellectual Property (IP) protection for the proposed tools	Demonstrations to at least two other external installations and comparison.

Achievements

IMA2.1

The demonstrations are planned later in the project

IMA2.2

The demonstrations are planned later in the project

Impact-making Objectives	Impact-making Activities	KPIs
IMO3. Standardisation, strategy and policy-making	**IMA3.1** Policy framework	KPIs: At least two domain-specific communities for dissemination and clustering.
	IMA3.2 Standardisation, strategy (investment measures) and policy-planning	KPIs: At least two domain-specific communities for dissemination and clustering

Achievements

IMA3.1

*The **European Ground segment security policy and standards community** has been approached through the participation to the ESA Ground segment security policy and standards event organised by ESA/ESOC in February 2021*
IMA3.2 *The **European Ground segment security policy and standards community** has been approached through the participation to the ESA Ground segment security policy and standards event organised by ESA/ESOC in February 2021*

3. Data Management Structure

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- The handling of research data during & after the end of the project
- What data will be collected, processed and/or generated
- Which methodology & standards will be applied
- Whether data will be shared/made open access and
- How data will be curated & preserved (including after the end of the project).

A DMP is required for all projects participating in the extended ORD pilot, unless they opt out of the ORD pilot. Due to the Restricted and classified nature of the project, the Consortium has decided to Opt-out of the Commission's Open Research Data Pilot (ORD pilot) before signing the Grant Agreement.

"7SHIELD will cross-cut domains in the protection of critical infrastructure. Data related to each of these domains may be security sensitive and should not be exposed publicly to prevent misuse of the data by potential bad actors trying to subvert the very security systems the project is aiming to put in place. Furthermore, data collected by sensors and other means may expose the range of security systems be commercially sensitive and proprietary to specific installations and not be for general public consumption. That said, 7SHIELD, despite opting out of the open data pilot, will aim to make any non-sensitive data available where appropriate. These opportunities will be monitored within the data management plan."

However, 7SHIELD Consortium has decided to submit a DMP on a voluntary basis[2]. A Data Management Plan (DMP) describes the data management life cycle for all data sets that will be collected, processed or generated by the research project. It is a document that outlines how research data will be handled during a research project, and even after the project is completed, describing which data will be collected, processed or generated and following specific methodologies and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved. Each dataset that will be generated by the project has to be described in compliance with the five dimensions provided by the EU Commission:

[2] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

- **Dataset identification and description (reference and name)**: a unique persistent identifier for the data set as well as a description, which specifies the origin, scope, scale, partners and link(s) to the corresponding publications (if any),
- **Standards and metadata**: a reference to relevant standards and a description of the metadata schema adopted to describe the data.
- **Data access and sharing**: all the information concerning access and reuse of the dataset including the nature of access (open or restricted), the tools or software needed, the reference and type of the repository where data are stored.
- **Archiving and preservation**: long-term preservation procedures, costs and volume of preserved data.
- **Ethics, privacy and legal requirements:** Any requirements to respect ethical and privacy regulations as well as legal compliance.
- **Storage and backup**: the standards and procedures for storing and backing-up the data, ensuring integrity, access control and security

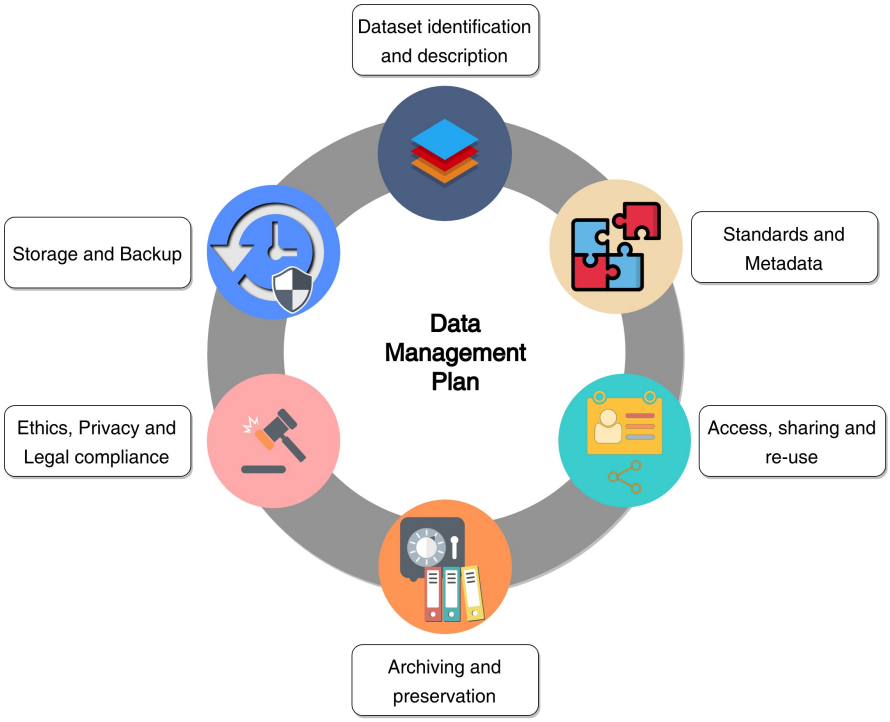

Figure 3-2: Data Management Plan Key dimensions

The 7SHIELD DMP has been developed by taking into account the template of the Guidelines on Data Management in Horizon 2020[3]. This document aims to help applicants and beneficiaries of projects to meet their responsibilities with regards to research data quality, sharing and security. In addition to the guidelines provided by the European

[3] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Commission, this document also refers to the plan to address the ethical issues related to data that will be collected during the project timeframe.

Therefore, the document follows the above structure, with each dimension presented in a dedicated section. Additionally, template for Dataset Management and for the Consent Form are reported in Annex. In Section 4, 7SHIELD Datasets reports each 7SHIELD dataset in detail. Moreover, this DMP is oriented to:

- The consortium partners;
- All stakeholders involved in the project;
- The European Commission.

This document can evolve along the project and it will be updated in D1.4 Self-assessment & data management plan v2 (M18). The update will include new sets of data and changes in consortium policies and datasets management.

3.1. 7SHIELD Purposes

The project addresses the security and resilience of EU Ground Segments of Space Systems, meeting the crosscutting and the sectoral criteria of the EU critical infrastructures (2008/14). Due to the growing reliance on satellite communications for today's economy and governments, the ground segments of space systems receive massive amounts of satellite data. A physical/cyber-attack to their installations or communication networks, respectively, would cause debilitating impact on public safety and security of EU citizens and public authorities. A physical attack on a space ground segment makes the distribution of satellite data problematic and, on the other hand, a cyber-attack in its data storage, access and exchange affects not only the reliability of space data, but also their FAIR standards: findability, accessibility, interoperability and reusability.

7SHIELD provides a holistic framework able to confront complex threats by covering all the macro-stages of crisis management, namely pre-crisis, crisis and post-crises phases. Specifically, 7SHIELD framework enables the deployment of innovative services for the cyber–physical protection of ground segments, such as e-fences, passive radars and laser technologies, multimedia AI technologies, that enhance their protection capabilities, while integrating or interoperating with existing protection solutions already deployed at their installations. The project makes use of advanced technologies for data integration, processing, analytics and visualisation as well as data security and cyberthreat protection to assess the prevention, detection and mitigation of threats, both physical and cyber. The project will be evaluated and demonstrated in five installations of ground segments of space systems.

3.2. Data Summary

3.2.1. Purposes of data

The purpose of collection and generation of data throughout the 7SHIELD project is to achieve the main project and to directly support the technical development of 7SHIELD tools or the development of operational processes. 7SHIELD includes a number of technology-oriented WPs (WP3, WP4, WP5 and WP6) that will process both open and closed source data. Moreover, collection and generation of data are necessary to manage the project, disseminate the information about it, analyze and exploit its results.

3.2.2. Data sources

The following data sources in the 7SHIELD project can be identified at this early stage:

- *Questionnaires, surveys and interviews with end-users*: a set of interviews with end-users will be conducted as the part of activities within the tasks T2.1, T2.2, T2.3 and T2.4 in WP2 to investigate specific requirements and needs. Moreover, the project partners were interviewed about the data they generated and collected within the project to prepare this deliverable.
- *Data collected from the field by physical and cyber sensors:* 7SHIELD project will apply different technologies to collect data:
    - UAV equipped with a set of different types of cameras deployed on the field, used both for control and monitoring purposes, as well as allowing the live collection of raw and pre-processed data from all platforms. Moreover, different sensor payloads can be installed on-board the UVs depending on the specific operation needs.
    - Thermal and near-infrared (NIR) cameras are used to detect the presence of intruders, such as moving objects and people, within the boundaries of an area under surveillance. This system is based on a network of near-infrared (NIR) and thermal sensors that support the automated monitoring server.
    - Innovative laser-based detection system is used for detection of ground based and aerial intrusion. Laser- based technology is LIDAR technology (2D and 3D). When an intruder (human or drone) cuts across the laser screen, the intruder is detected and its location information is sent to the control station (a PC), a slaved camera is made to turn onto the intruder, follows its movements and records its locations continuously.
    - Cyber and physical threat intelligence data based on open-source feeds and commercial providers is collected and analysed.
    - A face detection and recognition framework that will process both single shots and video streams from the surveillance cameras is also considered.

Face detection and face recognition will be deployed and linked either with a local host database.

- Wearable sensors using IoT: wearables for the team members using common-off-the-shelf (COTS) components as much as possible and a dedicated terminal to connect the sensors with local IoT communications. Each team member will be a sensor and a same time receptor of the decisions taken. Engagement rules and other hierarchical constraints and pre-operation relevant data should be acquired to improve effectiveness and better support for the operation. Moreover, a connection to the main system of 7SHIELD to acquire data and provide feedback, is capable with à priori information loaded (adding to any that is collected locally) to operate and provide valid outputs.
- Social media posts and other correspondence focusing on the output of credible voices such as ground space segment operators, civil contingency organisations and other stakeholders.

These technologies will be developed within the WP3, WP4, WP5, WP6 and used in pilots as part of WP7.

Other data sources in the 7SHIELD project will identified in subsequent stages:

- Workshops, demonstrations and piloting in WP7 (T7.3, T7.4 and T7.5),
- Meetings and stakeholders' engagement in WP8 (T.8.2, T8.3 and T8.4).

A multimodal data flow will be collected from these data sources. Leveraging innovative techniques for extracting and filtering features and information, only the "relevant" data for the specific 7SHIELD domain will be identified, selected, and processed.

Different datasets will be taken into account:

- Dataset 1: User Requirements
- Dataset 2: Security Requirements
- Dataset 3: Risk Assessment
- Dataset 4: Common Weakness Enumeration
- Dataset 5: Common Vulnerabilities and Exposures
- Dataset 6: Common Attack Pattern Enumeration and Classification
- Dataset 7: Data collection from UAVs and processing at the edge
- Dataset 8: Face detection and face recognition from video surveillance
- Dataset 9: Object detection and activity recognition from video content
- Dataset 10: Cyber-attack detection methods

- Dataset 11: Infrared and thermal image processing for the detection of man-made disasters
- Dataset 12: Laser-based technologies for the detection of ground-based and aerial threat detection
- Dataset 13: Combined Physical and Cyber Threat Detection and Early Warning
- Dataset 14: Semantic Representation
- Dataset 15: Data Severity Level
- Dataset 16: Emergency Response Plan
- Dataset 17: Social Awareness
- Dataset 18: Pilot Critical Operation

Each of these has specific type, attributes and dimensions and will be treated in different manner. The general strategy for data management will be based on the identification and classification of data generated and collected, standards and metadata to be used, exploitation and availability of data as well as how the data will be shared and archive to preserve the information.

3.2.3. Types of data

Several types of data are acquired and generated in 7SHIELD. A first classification can be done but, during the project, other types of data not envisaged at this time could be considered and included in the next version of this deliverable:

(i) **Personal data**, which will be used to provide personalized guidelines and 7SHIELD support including profile data and data from end-user/user group activity. Under article 4 of the GDPR[4] the personal data is defined as any information relating to an identified or identifiable natural person (data subject). In turn, the "identifiable natural person" is "anyone who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier (e.g. IP addresses) or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person"[5]. The personal data will be collected and processed within the project under the current EU regulations.

(ii) **Evaluation data and other-not personal-data**, including data used for assuring the functionality of the 7SHIELD solution, data collected during the scenarios and pilot use cases to assess the evolution of the users, and data from demonstration and

[4] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
[5] GDPR, art. 4.

evaluation tasks. Different measures will be applied depending on the utilisation of the gathered data.

(iii) **Data collected from the field by physical and cyber sensors**, including data coming from different technologies and sensors (cyber and physical) applied in the project to collect data.

3.3. FAIR data

3.3.1. Making data findable, including provisions for metadata

Data standards are the rules by which data are described and recorded. In order to share, exchange, and understand data, the format as well as the meaning has to be standardized. Standardization can be achieved by subdividing information into two main categories: data values, that are responsible of the result that can be obtained by the analysis of the data, and metadata, that allow users to understand, analyze, synthesize and research the datasets, as well as following and monitoring the progress of the research project. The use of data standards allows agencies to move from "project-based" data files to "enterprise" data files - and vice versa. In other words, the data become usable to more than just the project or person that created the data, because whoever uses the data knows the data will be in an expected format and what it represents. Since datasets will come from distributed and heterogeneous sources, an ontology-based data mapping approach will be designed to characterize these datasets. Each data source will provide metadata that expresses the rules/conditions that each schema realizes.

The metadata for the different identified datasets will be generated either automatically or through manual content annotation. A metadata framework will be used to identify the data types, owners and allowable use. This will be combined with a controlled access mechanism and with efficient encoding and encryption mechanisms. For the considered datasets, no definitive standards have been identified yet, but a set of metadata will be defined in relationship with the own data source. For this reason, more mature description of metadata used will be provided during the project development, with respect to the needs that could arise and in particular, as part of the work in various WPs related to the development of technical solutions (WP3, WP4, WP5 and WP6), conducting pilots and evaluating their results (WP7) and will be duly reported in the second version of the present deliverable.

3.3.2. Making data openly accessible

In the context of Research and Development, Open Access typically focuses on access to "scientific information", which refers to two main categories:

- Peer-reviewed scientific research articles (published in academic journals);
- Scientific research data (data underlying publications and/or raw data) [8].

The EU does not impose to researchers the obligation to publish their Results. It is always up to them to decide whether they want to publish some results or not. If researchers decide to proceed with such publication, art. 29 of the GA should be respected. More specifically, Art.29 of the GA foresees that "beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results." Therefore, open access becomes an issue only if publication is chosen as a means of dissemination.

At the same time, Open Access does not aim to affect IPRs. The beneficiary has to decide if, when and which Results to disseminate, in a way that does not affect the IP generated by research results and more specifically in a way that will not disrupt the decision to exploit research results commercially, e.g. through patenting. The decision on whether to publish through open access must come after the more general decision on whether to publish directly or to first seek protection for IPRs. This means that the beneficiary will decide the dissemination of the scientific information, having taken into consideration if and how to protect IPRs generated in the lifetime of the Project.

Following the general aim of the EU Open Science policy, Open Access objective is to enable the replicability and/or the uptake of research results by others. The end goal is to enable these research results to be used in an Open Innovation context, thereby speeding up the uptake of innovation with high impact for society. [EARTO Paper: Towards a Balanced Approach Between IPRs and Open Science Policy 31 July 2020, 2.1 Complementarity Between Open Science and IPRs, p.7]

Open Access does not aim to demise or diminish IPRs preventing industry from securing the element of shared "value capture" essential to Open Innovation. IPR remains a key to offer balanced rights to both the users and creators of Open Science content.

7SHIELD recognises the importance of making the research output of the project accessible as widely as possible. To this end, where permitted by the sensitivity of the data, the consortium takes active approach to the open access policy in Horizon 2020 in order to promote diffusion of knowledge and dissemination. More specifically, Open Access i.e. free on-line access, such as the "green[6]" or "gold[7]" model will be provided for the peer-reviewed scientific publications that relate to the project scientific results. Moreover, presentations by project participants about the project and public deliverables will be made publicly available through the project's website and will be posted in a public service like

[6] Self-archiving / 'green' open access – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. H2020 "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020", https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

[7] Open access publishing / 'gold' open access - an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers. H2020 "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020", https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

SlideShare. These will all be licensed via a Creative Commons license, like the project documents.

For what concerns data management, some of the collected data, in particular that concerning business-relevant data, organisational data, or personal data, could be sensitive and in this case will not be made available. Therefore, the consortium opts-out from the Pilot on Open Research Data in H2020. The data collected will only be exploited and/or shared/made accessible to project partners with a direct requirement to support their respective piece of work within 7SHIELD. Some data from end-users may not be made available if it is felt that it contains information of sensitive nature which would be unnecessarily disclosed. All data will be anonymised and held with the secure systems of the appropriate partner requiring access to the data. A data minimization policy will be also adopted. The latter implies that no data which is not strictly necessary for running the activity will be collected from individual participants and therefore processed.

7SHIELD framework will provide security policies in order to maintain the integrity of data and to make sure that the data will not be accessible by unauthorized parties or susceptible to corruption of data. In 7SHIELD, a Secure authentication mechanism for data access (T3.2) is designed and developed to support:

- Secure personal data storage in the system backend;
- Secure encrypted personal data search;
- Expressive and advanced access control over encrypted data;
- Secure data integrity verification.

A hybrid encryption mode that merges symmetric encryption, attribute-based encryption, proxy re-encryption and searchable encryption is used in the project. Light-weight hash-based message authentication is considered to achieve the integrity of personal data. Access control mechanisms will be developed for the authentication processes to allow secure data access in a controlled manner. This security level is needed to ensure the data access control to authorized users and entities.

3.3.3. Making data interoperable

Data interoperability is an important aspect of 7SHIELD's data management strategy while it enables to foster collaboration and increase the efficiency of data's use between the partners. Whenever possible, existing, well-defined data-exchange standards will be used.

Moreover, a 7SHIELD ontology will be created, based on the user requirements and standard representations. The metadata will be effectively represented via semantic models, populating the ontological structures, building on top of existing ontologies in the context of H2020 security projects. These structures will build upon already existing standards for semantically representing geospatial, multimedia and user information. The reasoning mechanism will create a cross-modal linking of sources and historical data to

increase the completeness and intelligence of the models. Many ontological frameworks have been proposed which are related with the 7SHIELD ontologies: The Semantic Sensor Network (SSN) ontology [3] models sensor devices, systems and procedures; the Time ontology [4] models temporal aspects such as temporal entities, time positions, durations etc. and relations such as before, finishes, during etc.; OGC GeoSPARQL standard [5] models geospatial objects and their topological and geometrical properties. Additionally, more abstract ontologies may be deployed, like e.g. FOAF [6], which connects people and information using the Web, or DOLCE+DnS Ultralite [7] which establishes interoperability among domain-specific ontologies.

3.3.4. Increase data re-use

The datasets, in general, will be stored during the periods necessary for achieving the purposes of its collection and later when it is necessary for the use within the project or later to employ and disseminate the project's results. For sensitive/restricted data access restrictions will be enforced (e.g. by requiring specific credentials, anonymization) while providing the specific data to authorized users only.

3.4. Allocation of resources (Responsibility and Resources)

All research data collected as part of this project and all the results are owned by the beneficiary that generates them. The whole Consortium will take responsibility for the collection, management, storage, security, sharing and quality assurance of the research data. ENG as the leader of T1.5 and of the WP1 that includes the preparation of DMP, have a particular responsibility in creating and updating the DMP. Each 7SHIELD partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation.

3.5. Data security

The 7SHIELD project will involve activities or generate results raising security concerns (in contrast it will contribute to resolve serious security-related challenges), as well as, due to the nature of the topic, i.e. Critical Infrastructures, will generate and handle classified information 'EU-classified information' (EUCI) as foreground.

The classification of the information of this project, according to Security Scrutiny after the proposal evaluation phase and the 7SHIELD GA, is EU-RES[8] (the lowest classification level of 4 levels).

The assessment and the monitoring of these EUCI will be managed from the Security Advisory Board (3 members from three different partners) and the Project Security Officer

[8] information and material the unauthorised disclosure of which could be disadvantageous to the interests of the European Union or of one or more of the Member States.

(chairman of the SAB). In this direction and based on the EU Legislations[9], the SAB & PSO have produced the Project Security Management Plan (an internal document) in order to help the members of the consortium with EUCI.

The PSMP provides a simple but efficient guide for the management of classified information among the partners of the 7SHIELD Project, in order to avoid security breach and support the secure implementation of the project, provides instructions on the protection of Classified Information that are provided or generated on behalf of the 7SHIELD project.

Based on this document, all Consortium Partners are informed on their obligations to deter breaches of security and compromise of Classified Information, by following instructions on the classification of the information, security procedures, including the handling and transfer of Classified Information, and visit procedures for 7SHIELD project.

There are specific rules for producing, accessing, opening, sending, reading, handling, discussing, exchanging, keeping safe/protecting etc the classified information.

For example:

EUCI will only be produced on a "safe" computer (not connected to any network)

The members of the consortium will not discuss EUCI with project partners over the phone or mail or on an unsecure area

EUCI will be stored in a suitable locked office

EUCI will be exchanged only after following a specific procedure (encryption of the information)

More specifically, below the foreseen process on how to exchange EUCI (e.g. in the framework of a deliverable) is presented:

1. Deliverable Leader using a safe computer (without any network) in order to encrypt the document using the approved cryptographic software ZED!
2. Using an electronic storage (USB) transfer the encrypted document to a computer with network connection
3. Sending through mail the encrypted document to the partners
4. Sending the password (key) to the partners through SMS
5. Partners receive the encrypted document, transfer the document using an electronic storage to a "safe" computer

[9] Commission's provisions on security and the rules on security as laid down in Commission Decision 2015/444 of 13 March 2015 on the security rules for protecting EU classified information & Commission Decision (Eu, Eratom) 2019/1962 "Implementing Rules for Handling Restreint UE/EU RESTRICTED information"17-10-2019

6. Partners decrypt the document using the password that have received to their mobile phones via SMS
7. Partners contribute to the specific deliverable and follows the same process to send back their contribution to the deliverable leader, until the final version is ready
8. The final version of the deliverable is reviewed by the SAB and when the security check is completed, the SAB will communicate with the Deliverable Leader in order to send it to the Coordinator.
9. The Coordinator then proceeds to the submission of the deliverable to the EC, following the agreed and appropriate procedures.

Respective procedures are followed also for the production and release of any dissemination material or scientific publications to deter potential data breach.

In order to ensure the efficient EUCI handling by all partners who are being involved with classified information, a training has been provided by the SAB, combined with the PSMP and with the necessary tools, such as the cryptographic software in order to be able to exchange EUCI. Through this process, all the data (EUCI) will be kept safe and not being used for any purpose other than that of carrying out the Grant Agreement.

Last but not least, sensitive information that may not be EUCI but still need to be secured and handled with caution, are foreseen within 7SHIELD activities. These kinds of data will be stored in a secure environment. More specifically, sensitive information needs to be stored in the appropriate infrastructure and format, corresponding to the related requirements and specifications of each pilot. Accessibility to the information needs to be maintained controlled and the networking configuration should not allow data duplication and circulation.

The following security measures will be applied:

- Data generated in the pilots will be stored in each pilot site based on the security measures specified in national legislations and the European General Data Protection Regulation.
- End-users will be requested to fill an informed consent form, which indicates the possible usage of their data.
- Files with personal information, including personal data and data collected from the 7SHIELD tools, will be protected by means of robust encryption schemes.
- Evaluation results of the pilots will be anonymised, applying the appropriate security measures.
- Ethical Committee approval from different pilot site will be sought if necessary.

3.6. Ethics and legal compliance

Given the complexities of data protection law across Europe, a data protection policy in line with relevant EU, national and local policies for the 7SHIELD project will be agreed within the first months of the project, in line with the agreed informed consent and other legal requirements. Finally, all data protection documentation will be centrally held by the project and will therefore be available for audit. Moreover, WP1 will make sure that data collection, management and access throughout and after the project is properly addressed.

All the personal data collected in the project will be processed under the EU's Data Protection laws, where the main legislation is the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, also known as the 'General Data Protection Regulation' (GDPR), which entered into force on 25 May 2018.

Further information on how personal data collection and handling should be approached in the 7SHIELD project, as well as other legal and ethical requirements are provided in WP9 and in particular in D9.1.

As explained, the DMP is a living document which will be kept updated during the whole lifetime of the project, since data generation and collection, and therefore data management, will be active in 7SHIELD for a considerable time after the submission of the initial version of the Data Management Plan. Whenever necessary this part of the deliverable will be updated accordingly.

4. 7SHIELD Project datasets

4.1. DATASET 1: User Requirements

Table 4-4 Dataset 1 – User Requirements

Template Field	Description
Overview	
Dataset Name	User requirements
Dataset Category	Primary data collected by partner (NOA) in 7SHIELD
Partner	NOA
Provider (if different from partner)	FMI, DEIMOS, DES, SPACEAPPS, SERCO
Work Package	2
Task/Deliverable	T2.2/D2.2
Details	
Short Description	Functional and non-functional user requirements of the 7SHIELD system collected by end-users and stakeholders (first responders) of the 7SHIELD Consortium via offline questionnaires.
Existing already before the 7SHIELD project?	NO
Use in 7SHIELD	The user requirements will be shared with the technical partners of 7SHIELD, who will translate them into technical specifications of the 7SHIELD system architecture in deliverable D6.1. The user requirements constitute the core of deliverable *D2.2 Consolidation of Stakeholder Requirements*.
Use beyond 7SHIELD	Not foreseen.
Storage and access details	
Is the data open?	No (Classified Information: EU-RESTRICTED)
Available to 7SHIELD partners?	YES (through deliverable D2.2)
Access control	Following handling EUCI procedures (e.g. use of encryption software to exchange and read the report).
What kind of processing is involved?	Processes on EUCI handling, based on Commission Decision 2015/444/EC and the developed PSMP by the SAB in order to support the consortium (e.g.

	encryption/decryption process through approved software, in order to exchange the classified deliverable between the partners involved).
What kind of derivative data is produced?	No derivative data.
How it becomes accessible to stakeholders outside the consortium?	Only after need-to-know basis and approval by the SAB.
Data flows and views	The user requirements will be reviewed by the 7SHIELD technical partners for the definition of technical specifications of the 7SHIELD system corresponding to each user requirement.
How can be managed after the project?	Only generic data that are not linked to CIs and are not contradictory with the Commission Decision 2015/444/EC and the developed PSMP.
License	Not applicable.
Type and format	.doc and .pdf
Data Size	Less than 5 MB.
Storage location	Partners' PCs
Storing responsible	NOA
Secure storage procedures	Classified information in electronic format will be stored encrypted through ZED! or following any other procedure indicated by the PSMP (e.g. secure room/administrative area for hardcopies).
Metadata	Not applicable.
Ethics and Data Protection	
Dataset contains personal data?	Data: name, surname, job title, expertise, employer. Informed consent was given for user/reuse. Personal data is anonymised.
DPIA required	NO
Dataset ethics and legal requirements	D9.1, Recruitment of participants, Informed Consent. Requirement respected.

4.2. DATASET 2: Security Requirements

Table 4-5 Dataset 2 – Security Requirements

Template Field	Description
Overview	
Dataset Name	Security Requirements
Dataset Category	Primary data collected by partner (KEMEA) in 7SHIELD Synthetic / generated data (Data generated from bibliographic research; analysis deriving from processed public data such as for example papers, documents, standards etc. processed)
Partner	KEMEA
Provider (if different from partner)	FMI, SPACEAPPS, SERCO, NOA, DEIMOS
Work Package	2
Task/Deliverable	T2.3/D2.5 (input also to D2.3)
Details	
Short Description	Literature and background review, and answers retrieved from questionnaires (to be) distributed to and interviews taken by members of the consortium but potentially also from outside the Consortium
Existing already before the 7SHIELD project?	NO
Use in 7SHIELD	• Elicitation of security requirements through the use case scenarios as well as through the technological solutions • To draft Security Requirement Report (D2.5 – M21) but also contribute to other deliverables (D2.2 – M6, D2.4 – M14)
Use beyond 7SHIELD	NONE
Storage and access details	
Is the data open?	No (Classified Information: RESTREINT UE)
Available to 7SHIELD partners?	YES (through D2.4 report)
Access control	Following handling EUCI procedures (e.g. use encryption software to exchange and read the report)

What kind of processing is involved?	Processes on EUCI handling, based on Commission Decision 2015/444/EC and the developed PSMP by the SAB in order to support the consortium (e.g. encryption/decryption process through approved software, in order to exchange the classified deliverable between the partners involved)
What kind of derivative data is produced?	No derivative data.
How it becomes accessible to stakeholders outside the consortium?	Only after need-to-know basis and approval by the SAB
Data flows and views	Security requirements are used in order to define system specification and architecture and are further used to validate/evaluate the final outcome.
How can be managed after the project?	Only generic data that are not linked to CIs and are not contradictory with the Commission Decision 2015/444/EC and the developed PSMP
License	N.A.
Type and format	.xls, .doc
Data Size	~5mb
Storage location	Partners' PCs and project MS TEAMS
Storing responsible	KEMEA
Secure storage procedures	Classified information in electronic format will be stored encrypted through ZED! or following any other procedure indicated by the PSMP (e.g. secure room/administrative area for hardcopies)
Metadata	N.A.
Ethics and Data Protection	
Dataset contains personal data?	• Deliverable to be submitted in May 2022. Data expected to be collected: name/surname & email. • Informed consent will be given for use/reuse. • Personal data will be anonymised.
DPIA required	Data Privacy Impact Assessment Required - NO
Dataset ethics and legal requirements	D9.1, Recruitment of participants, Informed Consent. Requirement respected.

4.3. DATASET 3: Risk Assessment

Table 4-6: Dataset 3 - Risk Assessment Dataset Management

Template Field	Description
Overview	
Dataset Name	Risk Assessment Data
Dataset Category	Derived data
Partner	RESIL
Provider (if different from partner)	n.a.
Work Package	WP3
Task/Deliverable	T3.1, T3.3
Details	
Short Description	The dataset will contain threats and vulnerabilities associated with the pilot infrastructures and the level of risk (likelihood, impact) connected with them.
Existing already before the 7SHIELD project?	NO
Use in 7SHIELD	The dataset will be used to perform analysis, run "what if" scenarios and better plan response and mitigation.
Use beyond 7SHIELD	n.a.
Storage and access details	
Is the data open?	NO
Available to 7SHIELD partners?	YES with restricted access
Access control	The access is regulated by EU legislation
What kind of processing is involved?	n.a
What kind of derivative data is produced?	n.a.
How it becomes accessible to stakeholders outside the consortium?	Not accessible.

Data flows and views	The data will be entered in the system by operators, then will be elaborated by tools. They will be available for planning countermeasures and mitigation.
How can be managed after the project?	n.a.
License	n.a.
Type and format	Ecore XML, CSV
Data Size	n.a
Storage location	RESIL server
Storing responsible	RESIL
Secure storage procedures	VPN, firewall, two factors authentication, secure communication
Metadata	n.a.
Ethics and Data Protection	
Dataset contains personal data?	No
DPIA required	n.a.
Dataset ethics and legal requirements	n.a.

4.4. DATASET 4: Common Weakness Enumeration

Table 4-7: Dataset 4 - CWE Dataset Management

Template Field	Description
Overview	
Dataset Name	CWE (Common Weakness Enumeration)
Dataset Category	Publicly available dataset
Partner	RESIL
Provider (if different from partner)	MITRE Corporation
Work Package	3
Task/Deliverable	T3.3
Details	

Short Description	A Community-Developed List of Software & Hardware Weakness Types. It can be downloaded from: https://cwe.mitre.org/data/downloads.html
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	It will be integrated in the MBDA module for associating weaknesses to assets and components of ground segment of space systems
Use beyond 7SHIELD	It serves as a common language, a measuring stick for security tools, and as a baseline for weakness identification, mitigation, and prevention efforts
Storage and access details	
Is the data open?	Yes – Public;
Available to 7SHIELD partners?	YES
Access control	The dataset can be downloaded in different alternative formats without any need of authentication
What kind of processing is involved?	The dataset is stored in a database and then retrieved for associating weaknesses to blocks and interfaces of a MBDA module's profiles/models
What kind of derivative data is produced?	The list of weaknesses affecting all the components and interfaces of a use case system
How it becomes accessible to stakeholders outside the consortium?	The dataset is already available. The derived data can be exported in a standard format e.g., Ecore XML.
Data flows and views	The dataset is manually downloaded from MITRE website and imported into a database. The database is then accessed by the module to retrieve weaknesses and associate them to system components.
How can be managed after the project?	The dataset is already available. The derived data, i.e., the list of weaknesses affecting all the components and interfaces of a use case system could be made publicly available for dissemination purposes and provided on a website in the form of downloadable Ecore XML.
License	CWE™ is free to use by any organization or individual for any research, development, and/or commercial purposes, per the CWE Terms of Use available at https://cwe.mitre.org/about/termsofuse.html
Type and format	XML, CSV, HTML, PDF

Data Size	10MB
Storage location	Dataset is available at https://cwe.mitre.org/ The downloaded version will be stored on RESIL's servers
Storing responsible	RESIL
Secure storage procedures	VPN, Firewall, two factors authentication, secure communication
Metadata	xsd
Ethics and Data Protection	
Dataset contains personal data?	no
DPIA required	no
Dataset ethics and legal requirements	none

4.5. DATASET 5: Common Vulnerabilities and Exposures

Table 4-8: Dataset 5 - CVE Dataset Management

Template Field	Description
Overview	
Dataset Name	CVE (Common Vulnerabilities and Exposures)
Dataset Category	Publicly available dataset
Partner	RESIL
Provider (if different from partner)	MITRE Corporation
Work Package	3
Task/Deliverable	T3.3
Details	
Short Description	A list of records for publicly known cybersecurity vulnerabilites. It can be downloaded from: https://cve.mitre.org/data/downloads/index.html
Existing already before the 7SHIELD project?	YES

Use in 7SHIELD	It will be integrated in the MBDA module for associating vulnerabilities to assets and components of ground segment of space systems
Use beyond 7SHIELD	Is used in numerous cybersecurity products and services from around the world, including the U.S. National Vulnerability Database (NVD)
Storage and access details	
Is the data open?	Yes – Public;
Available to 7SHIELD partners?	YES
Access control	The dataset can be downloaded in different alternative formats without any need of authentication
What kind of processing is involved?	The dataset is stored in a database and then retrieved for associating vulnerabilities to blocks and interfaces of a MBDA module's profiles/models
What kind of derivative data is produced?	The list of vulnerabilities affecting all the components and interfaces of a use case system
How it becomes accessible to stakeholders outside the consortium?	The dataset is already available. The derived data can be exported in a standard format e.g., Ecore XML.
Data flows and views	The dataset is manually downloaded from MITRE website and imported into a database. The database is then accessed by the module to retrieve vulnerabilities and associate them to system components.
How can be managed after the project?	The dataset is already available. The derived data, i.e., the list of vulnerabilities affecting all the components and interfaces of a use case system could be made publicly available for dissemination purposes and provided on a website in the form of downloadable Ecore XML.
License	As stated in https://cve.mitre.org/about/termsofuse.html : MITRE grants a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare derivative works of, publicly display, publicly perform, sublicense, and distribute the CVE®
Type and format	XML, CSV, HTML, Text
Data Size	35MB
Storage location	Dataset is available at: https://cve.mitre.org/data/downloads/index.html The downloaded version will be stored on RESIL's servers

Storing responsible	RESIL
Secure storage procedures	VPN, Firewall, two factors authentication, secure communication
Metadata	xsd
Ethics and Data Protection	
Dataset contains personal data?	no
DPIA required	no
Dataset ethics and legal requirements	none

4.6. DATASET 6: Common Attack Pattern Enumeration and Classification and Exposures

Table 4-9: Dataset 6 - CAPEC Dataset Management

Template Field	Description
Overview	
Dataset Name	CAPEC (Common Attack Pattern Enumeration and Classification)
Dataset Category	Publicly available dataset
Partner	RESIL
Provider (if different from partner)	MITRE Corporation
Work Package	3
Task/Deliverable	T3.3
Details	
Short Description	A publicly available catalogue of common attack patterns https://capec.mitre.org/index.html
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	It will be integrated in the MBDA module for determining possible attack patterns undermining assets and components of ground segment of space systems

Use beyond 7SHIELD	It can be used by analysts, developers, testers, and educators to advance community understanding on how adversaries exploit weaknesses in applications and other cyber-enabled capabilities and enhance defenses.
Storage and access details	
Is the data open?	Yes – Public;
Available to 7SHIELD partners?	YES
Access control	The dataset can be downloaded in different alternative formats without any need of authentication
What kind of processing is involved?	The dataset is stored in a database and then retrieved for determining possible attack patterns undermining components and interfaces of a MBDA module's profiles/models
What kind of derivative data is produced?	The list of attack patterns undermining a use case system
How it becomes accessible to stakeholders outside the consortium?	The dataset is already available. The derived data can be exported in a standard format e.g., Ecore XML.
Data flows and views	The dataset is manually downloaded from MITRE website and imported into a database. The database is then accessed by the module to retrieve attack patterns and associate them to system components.
How can be managed after the project?	The dataset is already available. The derived data, i.e., the list of attack patterns undermining a use case system could be made publicly available for dissemination purposes and provided on a website in the form of downloadable Ecore XML.
License	As stated in https://capec.mitre.org/about/termsofuse.html The MITRE Corporation (MITRE) hereby grants you a non-exclusive, royalty-free license to use Common Attack Pattern Enumeration and Classification (CAPEC™) for research, development, and commercial purposes. Any copy you make for such purposes is authorized provided that you reproduce MITRE's copyright designation and this license in any such copy.
Type and format	XML, CSV, HTML
Data Size	3MB
Storage location	Dataset is available at: https://capec.mitre.org/data/downloads.html

	The downloaded version will be stored on RESIL's servers
Storing responsible	RESIL
Secure storage procedures	VPN, Firewall, two factors authentication, secure communication
Metadata	xsd
Ethics and Data Protection	
Dataset contains personal data?	no
DPIA required	no
Dataset ethics and legal requirements	none

4.7. DATASET 7: Data collection from UAVs and processing at the edge

Table 4-10 Dataset 7 - Data collection from UAVs and processing at the edge

Template Field	Description
Overview	
Dataset Name	UAV Data
Dataset Category	Raw data by sensors or embedded camera
Partner	ACCELI
Provider (if different from partner)	N/A
Work Package	WP4
Task/Deliverable	T4.1 Data collection from UAVs and processing at the edge D4.3 Data collection from UAVs and processing at the edge techniques
Details	
Short Description	The data fusion and object detection/identification algorithms collect data from sensors (e.g. LiDAR, Camera and UAV), fuses them and send the output results to the 7SHIELD C2.
Existing already before the 7SHIELD project?	NO

Use in 7SHIELD	Will provide alerts in case of a human or other object type intrusion in specific areas. Also, will send videos of inspected areas.
Use beyond 7SHIELD	- To train new object detection/identification algorithms - To train new UAV architectures
Storage and access details	
Is the data open?	TBD
Available to 7SHIELD partners?	YES
Access control	Through 7SHIELD control room or directly through 7SHIELD repository
What kind of processing is involved?	Image processing embedded algorithms provided by CERTH in the framework of the T4.3 task.
What kind of derivative data is produced?	N/A at this stage
How it becomes accessible to stakeholders outside the consortium?	N/A at this stage
Data flows and views	7SHIELD UAV will collect video data through its embedded camera and raw data from the embedded sensors. It will receive notifications and commands from 7SHIELD C2 and it will transmit the outcomes from the edge processing to C2 (JSON Format)
How can be managed after the project?	N/A at this stage
License	N/A at this stage
Type and format	RSTP, XML, JSON, etc.
Data Size	Up to 10 GB.
Storage location	End user sites
Storing responsible	End Users
Secure storage procedures	N/A at this stage
Metadata	N/A at this stage
Ethics and Data Protection	
Dataset contains personal data?	Does the dataset contain personal data, and if YES, which?

	NO Was informed consent given for use/reuse? YES Is personal data anonymised? YES
DPIA required	Data Privacy Impact Assessment Required / NO
Dataset ethics and legal requirements	N/A at this stage

4.8. DATASET 8: Face detection and face recognition from video surveillance

Table 4-11: Dataset 8 – Labelled Faces in the Wild

Template Field	Description
Overview	
Dataset Name	Labelled Faces in the Wild
Dataset Category	Publicly available dataset
Partner	CERTH
Provider (if different from partner)	University of Massachusetts – Computer Vision Lab
Work Package	WP4
Task/Deliverable	T4.2 Face detection and face recognition from video surveillance D4.1 Video surveillance techniques: Initial release
Details	
Short Description	A database of face photographs designed for studying the problem of unconstrained face recognition. The data set contains more than 13,000 images of faces collected from the web.
Existing already before the 7SHIELD project?	Yes
Use in 7SHIELD	Will be used to train and test the face recognition module
Use beyond 7SHIELD	This is a well-known public benchmark for face verification, widely referenced in the related literature. It can be used for research purposes outside the 7SHIELD scope.

Storage and access details	
Is the data open?	Yes
Available to 7SHIELD partners?	Yes, from the original provider.
Access control	The dataset is available to download from the website: http://vis-www.cs.umass.edu/lfw/
What kind of processing is involved?	Facial feature extraction will be performed on the images.
What kind of derivative data is produced?	Facial features will be derived from the dataset images, which will be then used to train face recognition models. The features and models will exist in the 7SHIELD storage.
How it becomes accessible to stakeholders outside the consortium?	From the original provider.
Data flows and views	The raw data may enter the system as an input to the face recognition module. It will be used to train and test face recognition models. The raw data will not be forwarded to other modules or system components.
How can be managed after the project?	The dataset will not continue to be stored or managed outside the lifetime of the project.
License	Public Domain
Type and format	JPG images
Data Size	326.7 MB
Storage location	CERTH local server
Storing responsible	CERTH
Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	There are no additional metadata.
Ethics and Data Protection	
Dataset contains personal data?	The dataset contains personal data of famous celebrities (i.e., photos of their faces). The data will be used for research purposes within the scope of the 7SHIELD project. The dataset was collected from another institution, other than CERTH, and was made publicly available to the web. No formal consent has been

	requested. The distributor has provided direct access to retrieve all the data through their website using a published download link, without requiring to fill a consent request. The dataset, as distributed originally, includes non-anonymised personal data.
DPIA required	No
Dataset ethics and legal requirements	Legal req.: the photos in the dataset are in the public domain and are free of any copyright restrictions. Proper citation will be added in the project's deliverables in order to credit the creators of the dataset. Ethics req.: • Privacy – the photos found in this dataset contain personal identifiers (i.e., faces) of famous people that have consented to the public use of their photograph. If withdrawal of this consent is communicated to the dataset provider, CERTH will also ensure the deletion of the indicated photograph from its training model. • Bias - as mentioned in the dataset's main page, some groups (such as women) are not equally represented in the dataset. CERTH will amend this issue with the use of additional publicly available datasets that will enable the creation of a non-biased face recognition module. • Data minimisation principle - the data to be processed from this database will be strictly necessary for the training of the face recognition module and will not be used for any other purpose. The dataset's processing will be done in accordance with deliverables produced within the context of WP9 – Ethics requirements.

Table 4-12: Dataset 8 - WIDER Face

Template Field	Description
Overview	
Dataset Name	WIDER Face
Dataset Category	Publicly available dataset
Partner	CERTH
Provider (if different from partner)	Multimedia Laboratory, Department of Information Engineering, The Chinese University of Hong Kong
Work Package	WP4

Task/Deliverable	T4.2 Face detection and face recognition from video surveillance D4.1 Video surveillance techniques: Initial release
Details	
Short Description	This is a database that contains over 30000 images which mostly show people participating in various activities of everyday life. The human faces appear with a high degree of variability in scale, pose and occlusion.
Existing already before the 7SHIELD project?	Yes
Use in 7SHIELD	Will be used to train and test the face detection module
Use beyond 7SHIELD	This is a well-known public benchmark for face detection, widely referenced in the related literature. It can be used for research purposes outside the 7SHIELD scope.
Storage and access details	
Is the data open?	Yes
Available to 7SHIELD partners?	Yes, from the original provider.
Access control	The dataset is available to download from the website: http://shuoyang1213.me/WIDERFACE/
What kind of processing is involved?	Facial feature extraction will be performed on the images.
What kind of derivative data is produced?	Facial features will be derived from the dataset images, which will be then used to train face detection models. The features and models will exist in the 7SHIELD storage.
How it becomes accessible to stakeholders outside the consortium?	From the original provider.
Data flows and views	The raw data may enter the system as an input to the face detection module. It will be used to train and test face detection models. The raw data will not be forwarded to other modules or system components.
How can be managed after the project?	The dataset will not continue to be stored or managed outside the lifetime of the project.
License	Public Domain
Type and format	JPG images, TXT files

Data Size	3.7 GB
Storage location	CERTH local server
Storing responsible	CERTH
Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	There are additional metadata in the form of bounding box annotations of human faces (pixel coordinates) for every image.
Ethics and Data Protection	
Dataset contains personal data?	The dataset contains personal data (i.e., photos of human faces). The data will be used for research purposes within the scope of the 7SHIELD project. The dataset was collected from another institution, other than CERTH, and was made publicly available to the web. No formal consent has been requested. The distributor has provided direct access to retrieve all the data through their website using a published download link, without requiring to fill a consent request. The dataset, as distributed originally, includes non-anonymised personal data.
DPIA required	No
Dataset ethics and legal requirements	Legal req.: the photos in the dataset are in the public domain and are free of any copyright restrictions. Proper citation will be added in the project's deliverables in order to credit the creators of the dataset. Ethics req.: • Privacy – the photos found in this dataset contain personal identifiers (i.e., faces) of people that have consented to the public use of their photograph. If withdrawal of this consent is communicated to the dataset provider, CERTH will also ensure the deletion of the indicated photograph from its training model. • Data minimisation principle - the data to be processed from this database will be strictly necessary for the training of the face recognition module and will not be used for any other purpose. The dataset's processing will be done in accordance with deliverables produced within the context of WP9 – Ethics requirements.

Table 4-13: Dataset 8 - FDDB

Template Field	Description
Overview	
Dataset Name	FDDB
Dataset Category	Publicly available dataset
Partner	CERTH
Provider (if different from partner)	University of Massachusetts – Computer Vision Lab
Work Package	WP4
Task/Deliverable	T4.2 Face detection and face recognition from video surveillance D4.1 Video surveillance techniques: Initial release
Details	
Short Description	This is a dataset of face regions designed for studying the problem of unconstrained face detection. This dataset contains the annotations for 5171 faces in a set of 2845 images taken from the Labelled Faces in the Wild dataset.
Existing already before the 7SHIELD project?	Yes
Use in 7SHIELD	Will be used to train and test the face detection module
Use beyond 7SHIELD	This is a well-known public benchmark for face detection, widely referenced in the related literature. It can be used for research purposes outside the 7SHIELD scope.
Storage and access details	
Is the data open?	Yes
Available to 7SHIELD partners?	Yes, from the original provider.
Access control	The dataset is available to download from the website: http://vis-www.cs.umass.edu/fddb/
What kind of processing is involved?	Facial feature extraction will be performed on the images.
What kind of derivative data is produced?	Facial features will be derived from the dataset images, which will be then used to train face detection models. The features and models will exist in the 7SHIELD storage.
How it becomes accessible to stakeholders outside the consortium?	From the original provider.

Data flows and views	The raw data may enter the system as an input to the face detection module. It will be used to train and test face detection models. The raw data will not be forwarded to other modules or system components.
How can be managed after the project?	The dataset will not continue to be stored or managed outside the lifetime of the project.
License	Public Domain
Type and format	JPG images, TXT files
Data Size	553.7 MB
Storage location	CERTH local server
Storing responsible	CERTH
Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	There are additional metadata in the form of bounding box annotations of human faces (pixel coordinates) for every image.
Ethics and Data Protection	
Dataset contains personal data?	The dataset contains personal data (i.e., photos of human faces). The data will be used for research purposes within the scope of the 7SHIELD project. The dataset was collected from another institution, other than CERTH, and was made publicly available to the web. No formal consent has been requested. The distributor has provided direct access to retrieve all the data through their website using a published download link, without requiring to fill a consent request. The dataset, as distributed originally, includes non-anonymised personal data.
DPIA required	No
Dataset ethics and legal requirements	Legal req.: the photos in the dataset are in the public domain and are free of any copyright restrictions. Proper citation will be added in the project's deliverables in order to credit the creators of the dataset. Ethics req.: • Privacy – the photos found in this dataset contain personal identifiers (i.e., faces) of people that have consented to the public use of their photograph. If

	withdrawal of this consent is communicated to the dataset provider, CERTH will also ensure the deletion of the indicated photograph from its training model. • Data minimisation principle - the data to be processed from this database will be strictly necessary for the training of the face recognition module and will not be used for any other purpose. The dataset's processing will be done in accordance with deliverables produced within the context of WP9 – Ethics requirements.

Table 4-14: Dataset 8 – Gallery of Authorized People

Template Field	Description
Overview	
Dataset Name	Gallery of Authorized People
Dataset Category	Primary data collected by partner in 7SHIELD
Partner	CERTH
Provider (if different from partner)	The End Users will provide the data
Work Package	WP4
Task/Deliverable	T4.2 Face detection and face recognition from video surveillance D4.1 Video surveillance techniques: Initial release
Details	
Short Description	This dataset will be compiled from volunteers which will appear on CCTVs acting as authorized personnel, during the pilot tests of the 7SHIELD project. The gallery will contain high resolution photos of their faces from various angles.
Existing already before the 7SHIELD project?	No
Use in 7SHIELD	During the pilot runs, the face detection and recognition module will continuously monitor specific locations covered by CCTV cameras. Real human volunteers will appear on the CCTV streams which will be either authorized personnel or unauthorized people trying to breach in secure locations. The face recognition module will try to match every detected face to an authorized person from the existing gallery using facial feature similarity metrics. If a match cannot be made with a level of certainty an alert indicating potential unauthorized access will be produced.

Use beyond 7SHIELD	The data will not be available to use beyond the lifetime or outside the scope of the 7SHIELD project.
Storage and access details	
Is the data open?	No
Available to 7SHIELD partners?	Available only to CERTH and the end user volunteers
Access control	The data and all derivatives should be stored in 7SHIELD raw data storage. Access credentials should be granted to CERTH and end user volunteers only.
What kind of processing is involved?	Facial feature extraction will be performed on the images.
What kind of derivative data is produced?	Facial features will be derived from the gallery images, which will be then used to match facial features of unknown detected faces.
How it becomes accessible to stakeholders outside the consortium?	Access of the volunteer image gallery will not be granted outside the consortium.
Data flows and views	The gallery of photos will be provided as input to the specific component of the face detection and recognition module which will be responsible for facial feature extraction. As soon as features are extracted, they will be stored on the same storage space with the images, enriching the gallery. The features and/or images will be available for retrieval from the face recognition component whenever a feature comparison of an unknown face to the ones existing on the gallery is required.
How can be managed after the project?	The dataset will not continue to be stored or managed outside the lifetime of the project.
License	Proprietary
Type and format	TBC
Data Size	TBC
Storage location	TBC
Storing responsible	TBC
Secure storage procedures	TBC, Subject to storage medium.
Metadata	Metadata existing along with the images
Ethics and Data Protection	
Dataset contains personal data?	The data will contain personal data (photos of volunteers)

	Formal consent will be requested from the volunteers upon the use/reuse of their photos for the pilot tests as well as for research purposes. Personal data will not be anonymised.
DPIA required	No
Dataset ethics and legal requirements	Legal req.: All the photos in this dataset will be used within the scope of testing the 7SHIELD face detection and recognition module during the project's pilot runs. The volunteers' photos will be collected and used only after the volunteer's formal consent is given. Ethics req.: • Privacy – the photos in this dataset will contain personal identifiers (i.e., faces) of people that have consented to the use of their photograph within the scope of 7SHIELD pilot testing. • Data minimisation principle - the data to be processed from this database will be strictly necessary for the evaluation of the face recognition module and will not be used for any other purpose. The dataset's processing will be done in accordance with deliverables produced within the context of WP9 – Ethics requirements.

4.9. DATASET 9: Object detection and activity recognition from video content

Table 4-15: Dataset 9 – Microsoft COCO (Common Object in Context) 2017

Template Field	Description
Overview	
Dataset Name	Microsoft COCO (Common Object in Context) 2017
Dataset Category	Publicly available dataset
Partner	CERTH
Provider (if different from partner)	Microsoft
Work Package	WP4
Task/Deliverable	T4.3 Object detection and activity recognition from video content D4.1 - Video surveillance techniques: Initial release
Details	

Short Description	COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features: • Object segmentation • Recognition in context • Superpixel stuff segmentation • 330K images (>200K labeled) • 1.5 million object instances • 80 object categories • 91 stuff categories • 5 captions per image • 250,000 people with keypoints
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	Will be used to train the object detection module (and possibly activity recognition module)
Use beyond 7SHIELD	This is a public dataset which contains objects from the first person perspective and is widely used in object detection (and other tasks).
Storage and access details	
Is the data open?	Yes – Public
Available to 7SHIELD partners?	YES
Access control	Can be downloaded from the original publisher
What kind of processing is involved?	Deep learning network processing which will involve image classification and regression as outputs.
What kind of derivative data is produced?	A deep network model which will produce labels and bounding boxes around object of interest inside images.
How it becomes accessible to stakeholders outside the consortium?	From the original provider
Data flows and views	Images with annotated objects inside will be forwarded to train the module (s). The annotated output of the inference images and or videos will be propagated to the dashboard and any relevant module for demonstration/informative purposes.
How can be managed after the project?	The original dataset is publicly available so it will exist after the project life circle end. Any output of the module will not be used outside the projects.
License	Annotations are released under the Creative Commons Attribution 4.0 License. Use of the images must abide by the Flickr Terms of Use.

Type and format	Images in jpeg format, annotations in json format
Data Size	The whole dataset ~20.2 GB
Storage location	CERTH local server
Storing responsible	CERTH
Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	There are no additional metadata.
Ethics and Data Protection	
Dataset contains personal data?	Yes, it contains.
DPIA required	No
Dataset ethics and legal requirements	Please describe any Ethics and/or legal requirements relevant to the dataset - If you want you can refer to relevant Ethics deliverables.

Table 4-16: Dataset 9 – Pascal VOC (Visual Object Classes)

Template Field	**Description**
Overview	
Dataset Name	Pascal VOC (Visual Object Classes)
Dataset Category	Publicly available dataset
Partner	CERTH
Provider (if different from partner)	PASCAL2 Network of Excellence on Pattern Analysis, Statistical Modelling and Computational Learning
Work Package	WP4
Task/Deliverable	T4.3 Object detection and activity recognition from video content D4.1 - Video surveillance techniques: Initial release
Details	
Short Description	Pascal VOC is an object detection dataset containing images from 20 different classes from mostly first person perspective.
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	Will be used to train the object detection module (and possibly activity recognition module)

Use beyond 7SHIELD	This is a public dataset which contains objects from a first person perspective and is widely used in object detection (and other tasks).
Storage and access details	
Is the data open?	Yes – Public
Available to 7SHIELD partners?	YES
Access control	Can be downloaded from the original publisher
What kind of processing is involved?	Deep learning network processing which will involve image classification and regression as outputs.
What kind of derivative data is produced?	A deep network model which will produce labels and bounding boxes around object of interest inside images.
How it becomes accessible to stakeholders outside the consortium?	From the original provider
Data flows and views	Images with annotated objects inside will be forwarded to train the module (s). The annotated output of the inference images and or videos will be propagated to the dashboard and any relevant module for demonstration/informative purposes.
How can be managed after the project?	The original dataset is publicly available so it will exist after the project life circle end. Any output of the module will not be used outside the projects.
License	Use of the images must abide by the Flickr Terms of Use.
Type and format	Images in jpeg format, annotations in xml format
Data Size	The whole dataset ~2.9 GB
Storage location	CERTH local server
Storing responsible	CERTH
Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	There are no additional metadata.
Ethics and Data Protection	
Dataset contains personal data?	Yes, it contains.
DPIA required	No

Dataset ethics and legal requirements	Please describe any Ethics and/or legal requirements relevant to the dataset - If you want you can refer to relevant Ethics deliverables.

Table 4-17: Dataset 9 - VisDrone

Template Field	Description
Overview	
Dataset Name	VisDrone
Dataset Category	Publicly available dataset
Partner	CERTH
Provider (if different from partner)	AISKYEYE team at Lab of Machine Learning and Data Mining, Tianjin University, China
Work Package	WP4
Task/Deliverable	T4.3 Object detection and activity recognition from video content D4.1 - Video surveillance techniques: Initial release
Details	
Short Description	The benchmark dataset consists of **400** video clips formed by **265,228** frames and **10,209** static images, captured by various drone-mounted cameras, covering a wide range of aspects including location (taken from 14 different cities separated by thousands of kilometres in China), environment (urban and country), objects (pedestrian, vehicles, bicycles, etc.), and density (sparse and crowded scenes). Note that, the dataset was collected using various drone platforms (i.e., drones with different models), in different scenarios, and under various weather and lighting conditions. These frames are manually annotated with more than **2.6 million** bounding boxes or points of targets of frequent interests, such as pedestrians, cars, bicycles, and tricycles. Some important attributes including scene visibility, object class and occlusion, are also provided for better data utilization.
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	Will be used to train the object detection module (and possibly activity recognition module)
Use beyond 7SHIELD	This is a dataset which contain object from UAV perspective and can be used for similar purposes.

Storage and access details	
Is the data open?	Yes – Public
Available to 7SHIELD partners?	YES
Access control	Can be downloaded from the original publisher
What kind of processing is involved?	Deep learning network processing which will involve image classification and regression as outputs.
What kind of derivative data is produced?	A deep network model which will produce labels and bounding boxes around object of interest inside images.
How it becomes accessible to stakeholders outside the consortium?	From the original provider
Data flows and views	Images with annotated objects inside will be forwarded to train the module (s). The annotated output of the inference images and or videos will be propagated to the dashboard and any relevant module for demonstration/informative purposes.
How can be managed after the project?	The original dataset is publicly available so it will exist after the project life circle end. Any output of the module will not be used outside the projects.
License	Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License
Type and format	Images in jpeg format, annotations in text format
Data Size	~1.8 GB
Storage location	CERTH local server
Storing responsible	CERTH
Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	There are no additional metadata.
Ethics and Data Protection	
Dataset contains personal data?	It does not contain personal data and efforts have been made to exclude identifiable information from the data to protect privacy.
DPIA required	No

Dataset ethics and legal requirements	Please describe any Ethics and/or legal requirements relevant to the dataset - If you want you can refer to relevant Ethics deliverables.

Table 4-18: Dataset 9- UAV123

Template Field	Description
Overview	
Dataset Name	UAV123
Dataset Category	Publicly available dataset
Partner	CERTH
Provider (if different from partner)	The Image and Video Understanding Lab (IVUL) at King Abdullah University of Science and Technology (KAUST), Saudi Arabia
Work Package	WP4
Task/Deliverable	T4.3 Object detection and activity recognition from video content D4.1 - Video surveillance techniques: Initial release
Details	
Short Description	UAV123 contains sequences from an aerial viewpoint, a subset of which is meant for long-term aerial tracking (UAV20L). Our new UAV123 dataset contains a total of 123 video sequences and more than 110K frames making it the second-largest object tracking dataset after ALOV300++. All sequences are fully annotated with upright bounding boxes.
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	Will be used to train the object detection module (and possibly activity recognition module)
Use beyond 7SHIELD	This is a dataset which contain object from UAV perspective and can be used for similar purposes.
Storage and access details	
Is the data open?	Yes – Public
Available to 7SHIELD partners?	YES
Access control	Can be downloaded from the original publisher

What kind of processing is involved?	A sampling of frames was used. Additional annotation of all object of interest was produced for the selected frames. Deep learning network processing which will involve image classification and regression as outputs.
What kind of derivative data is produced?	A deep network model which will produce labels and bounding boxes around object of interest inside images.
How it becomes accessible to stakeholders outside the consortium?	From the original provider
Data flows and views	Images with annotated objects inside will be forwarded to train the module (s). The annotated output of the inference images and or videos will be propagated to the dashboard and any relevant module for demonstration/informative purposes.
How can be managed after the project?	The original dataset is publicly available so it will exist after the project life circle end. Any additional annotation produces can be used for object detection purposes.
License	Unknown
Type and format	Images in jpeg format, annotations in text format
Data Size	Complete dataset ~13.7GB
Storage location	CERTH local server
Storing responsible	CERTH
Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	The dataset has been annotated for object detection purposes. These annotations will be used to train the object detector.
Ethics and Data Protection	
Dataset contains personal data?	It may contain some personal data but it's a public dataset and the publisher owns the right of the images.
DPIA required	No
Dataset ethics and legal requirements	Please describe any Ethics and/or legal requirements relevant to the dataset - If you want you can refer to relevant Ethics deliverables.

Table 4-19: Dataset 9 – UCF Aerial Action

Template Field	Description

Overview	
Dataset Name	UCF Aerial Action Data Set
Dataset Category	Publicly available dataset
Partner	CERTH
Provider (if different from partner)	Centre For Research in Computer Vision Lab from University of Central Florida (UCF)
Work Package	WP4
Task/Deliverable	T4.3 Object detection and activity recognition from video content D4.1 - Video surveillance techniques: Initial release
Details	
Short Description	This data set features video sequences that were obtained using a R/C-controlled blimp equipped with an HD camera mounted on a gimbal. The collection represents a diverse pool of actions featured at different heights and aerial viewpoints. Multiple instances of each action were recorded at different flying altitudes which ranged from 400-450 feet and were performed by different actors.
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	Will be used to train the object detection module (and possibly activity recognition module)
Use beyond 7SHIELD	This is a dataset which contain object from UAV perspective and can be used for similar purposes.
Storage and access details	
Is the data open?	Yes – Public
Available to 7SHIELD partners?	YES
Access control	Can be downloaded from the original publisher
What kind of processing is involved?	Frames were extracted from the videos. Annotations were transformed to xml format. Deep learning network processing which will involve image classification and regression as outputs.
What kind of derivative data is produced?	A deep network model which will produce labels and bounding boxes around object of interest inside images.

How it becomes accessible to stakeholders outside the consortium?	From the original provider
Data flows and views	Images with annotated objects inside will be forwarded to train the module (s). The annotated output of the inference images and or videos will be propagated to the dashboard and any relevant module for demonstration/informative purposes.
How can be managed after the project?	The original dataset is publicly available so it will exist after the project life circle end. Any additional annotation produces can be used for object detection purposes.
License	Unknown
Type and format	Videos in mpg format, annotations in VIPER format
Data Size	Final dataset used ~21MB
Storage location	CERTH local server
Storing responsible	CERTH
Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	The dataset has been annotated for object detection purposes. These annotations will be used to train the object detector.
Ethics and Data Protection	
Dataset contains personal data?	It may contain some personal data but it's a public dataset and the publisher owns the right of the images.
DPIA required	No
Dataset ethics and legal requirements	Please describe any Ethics and/or legal requirements relevant to the dataset - If you want you can refer to relevant Ethics deliverables.

Table 4-20: Dataset 9 – VIRAT v1

Template Field	Description
Overview	
Dataset Name	VIRAT v1
Dataset Category	Publicly available dataset
Partner	CERTH

Provider (if different from partner)	**VIRAT Video Dataset collection**
Work Package	WP4
Task/Deliverable	T4.3 Object detection and activity recognition from video content D4.1 - Video surveillance techniques: Initial release
Details	
Short Description	The VIRAT Video Dataset is designed to be realistic, natural and challenging for video surveillance domains in terms of its resolution, background clutter, diversity in scenes, and human activity/event categories than existing action recognition datasets. It has become a benchmark dataset for the computer vision community. **Ground and Aerial Videos: Both ground camera videos and aerial videos are collected released as part of VIRAT Video Dataset.**
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	Will be used to train the activity recognition module.
Use beyond 7SHIELD	This is a public dataset which contains objects from a first person perspective and is widely used in activity recognition and detection.
Storage and access details	
Is the data open?	Yes – Public
Available to 7SHIELD partners?	YES
Access control	Can be acquired from the original publisher
What kind of processing is involved?	Deep learning network processing which will involve image classification and regression as outputs.
What kind of derivative data is produced?	A deep network model which will produce labels for video segments.
How it becomes accessible to stakeholders outside the consortium?	From the original provider
Data flows and views	Images with annotated video segments will be used to train the module. The annotated output of the inference videos will be propagated to the dashboard and any relevant module for demonstration/informative purposes.

How can be managed after the project?	The original dataset is publicly available so it will exist after the project life circle end. Any output of the module will not be used outside the projects.
License	The V1 data is governed by the VIRAT Video Dataset Usage Agreement. The V1-training annotations are released CC-BY 4.0.
Type and format	Images extracted in png format, videos in mp4 format, annotations in yml format
Data Size	The mp4 and annotations files ~52 GB
Storage location	CERTH local server
Storing responsible	CERTH
Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	There are no additional metadata.
Ethics and Data Protection	
Dataset contains personal data?	Yes, it contains.
DPIA required	No
Dataset ethics and legal requirements	Please describe any Ethics and/or legal requirements relevant to the dataset - If you want you can refer to relevant Ethics deliverables.

Table 4-21: Dataset 9 - CHARADES

Template Field	Description
Overview	
Dataset Name	CHARADES
Dataset Category	Publicly available dataset
Partner	CERTH
Provider (if different from partner)	VIRAT Video Dataset collection
Work Package	WP4
Task/Deliverable	T4.3 Object detection and activity recognition from video content D4.1 - Video surveillance techniques: Initial release
Details	

Short Description	Charades is dataset composed of 9848 videos of daily indoors activities collected through Amazon Mechanical Turk. 267 different users were presented with a sentence, that includes objects and actions from a fixed vocabulary, and they recorded a video acting out the sentence. The dataset contains 66,500 temporal annotations for 157 action classes, 41,104 labels for 46 object classes, and 27,847 textual descriptions of the videos.
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	Will be used to train the activity recognition module.
Use beyond 7SHIELD	This is a public dataset which contains objects from a first person perspective and is widely used activity recognition.
Storage and access details	
Is the data open?	Yes – Public
Available to 7SHIELD partners?	YES
Access control	Can be acquired from the original publisher
What kind of processing is involved?	Deep learning network processing which will involve image classification and regression as outputs.
What kind of derivative data is produced?	A deep network model which will produce **labels for video segments.**
How it becomes accessible to stakeholders outside the consortium?	From the original provider
Data flows and views	Images with annotated video segments will be used to train the module. The annotated output of the inference videos will be propagated to the dashboard and any relevant module for demonstration/informative purposes.
How can be managed after the project?	The original dataset is publicly available so it will exist after the project life circle end. Any output of the module will not be used outside the projects.
License	License for Non-Commercial Use
Type and format	videos in mp4 format, annotations in csv format
Data Size	Original size ~55 GB
Storage location	CERTH local server
Storing responsible	CERTH

Secure storage procedures	The local sever is protected by the CERTH firewall. No direct access points will be created to access these data.
Metadata	There are no additional metadata.
Ethics and Data Protection	
Dataset contains personal data?	Yes, it contains.
DPIA required	No
Dataset ethics and legal requirements	Please describe any Ethics and/or legal requirements relevant to the dataset - If you want you can refer to relevant Ethics deliverables.

4.10. DATASET 10: Cyber-attack detection methods

Table 4-22: Dataset 10 - DDoS Evaluation Dataset (CIC-DDoS2019)

Template Field	Description
Overview	
Dataset Name	DDoS Evaluation Dataset (CIC-DDoS2019)
Dataset Category	Publicly available dataset
Partner	CeRICT
Provider (if different from partner)	Canadian Institute for Cybersecurity
Work Package	WP4
Task/Deliverable	T4.4 - Cyber-attack detection methods D4.4 - Cyber-attack detection methods
Details	
Short Description	CICDDoS2019 contains benign and the most up-to-date common DDoS attacks, which resembles the true real-world data (PCAPs). It also includes the results of the network traffic analysis using CICFlowMeter-V3 with labelled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols and attack (CSV files).
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	the dataset will be used to test the detection of network patterns in order to trigger alerts and alarms

Use beyond 7SHIELD	The dataset can be used via AI techniques for analysis. A different feature extractor can be used, using raw scanned files (PCAP) to extract features. Data mining techniques can be used to analyse the generated data.
Storage and access details	
Is the data open?	YES
Available to 7SHIELD partners?	Yes, from the original provider
Access control	Can be downloaded from the provider website
What kind of processing is involved?	attack pattern detection using pre-set rules and information correlation
What kind of derivative data is produced?	Alerts and alarms deriving from the detection
How it becomes accessible to stakeholders outside the consortium?	From the original provider
Data flows and views	The dataset will be processed by the cyber-attack detection module and possibly by the cyber-attack correlation solution to be displayed as alerts and alarms
How can be managed after the project?	The dataset is public and will continue to exist even after the end of the project
License	Any use or redistribution of the data must include a citation to the CICDDoS2019 dataset and related published paper.
Type and format	Raw data network traffic (Pcaps) and event logs. Several features were then extracted from raw data and saved in CSV.
Data Size	2.2 Gb
Storage location	CeRICT local server
Storing responsible	CeRICT
Secure storage procedures	N/A at this stage
Metadata	N/A at this stage
Ethics and Data Protection	
Dataset contains personal data?	NO
DPIA required	NO

Dataset ethics and legal requirements	The dataset does not raise any ethical or legal issues

Table 4-23: Dataset 10 - ISOT Ransomware Dataset

Template Field	Description
Overview	
Dataset Name	ISOT Ransomware Dataset
Dataset Category	Publicly available dataset
Partner	CeRICT
Provider (if different from partner)	ISOT Research Lab
Work Package	WP4
Task/Deliverable	T4.4 - Cyber-attack detection methods D4.4 - Cyber-attack detection methods
Details	
Short Description	ISOT ransomware dataset is a combination of the behaviour data for a collection of ransomware samples and benign applications.
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	The dataset is used to detect changes to data due to the effect of ransomware, in order to improve detection
Use beyond 7SHIELD	The dataset is public and can be used for research purposes
Storage and access details	
Is the data open?	YES
Available to 7SHIELD partners?	Yes, from the original provider
Access control	Can be downloaded from the provider website
What kind of processing is involved?	attack pattern detection using pre-set rules and information correlation
What kind of derivative data is produced?	Alerts and alarms deriving from the detection

How it becomes accessible to stakeholders outside the consortium?	From the original provider
Data flows and views	The dataset will be processed by the cyber-attack detection module and possibly by the cyber-attack correlation solution to be displayed as alerts and alarms
How can be managed after the project?	The dataset is public and will continue to exist even after the end of the project
License	The dataset should be used for research purposes only. The dataset must not be passed on to other researchers without the explicit permission of Dr. Issa Traore. The source of the dataset must be indicated in any publications by means of a citation.
Type and format	Network dump of the traffic (pcap) memory dump of the analysis machine (dmp) Meta information about all processes that touched the file, its original file path in the analysis machine, etc. (.json) raw logs (.bson)
Data Size	420 GB
Storage location	CeRICT local server
Storing responsible	CeRICT
Secure storage procedures	N/A at this stage
Metadata	N/A at this stage
Ethics and Data Protection	
Dataset contains personal data?	NO
DPIA required	NO
Dataset ethics and legal requirements	The dataset does not raise any ethical or legal issues

4.11. DATASET 11: Infrared and thermal image processing for the detection of man-made disasters

Table 4-24: Dataset 11 - Thermal images with vehicles, people and large animals

Template Field	Description
Overview	
Dataset Name	Thermal images with vehicles, people and large animals

Dataset Category	Publicly available dataset
Partner	INOV
Provider (if different from partner)	FLIR for annotated vehicle and people images and YouTube for large animal images
Work Package	WP4
Task/Deliverable	T4.5 Infrared and thermal image processing for the detection of man-made disasters D4.6 Infrared and thermal image processing techniques
Details	
Short Description	The dataset consists of thermal images containing vehicles, people and large animals. It was collected by FLIR by driving around in a car and collecting images from what was seen. The large animals are collected from YouTube videos posted by several different authors. All images are annotated for the presence of objects of the classes to be identified in the images.
Existing already before the 7SHIELD project?	YES (FLIR images), No (Large animals' data)
Use in 7SHIELD	To create deep learning neural networks for detection of vehicles, people and large animals.
Use beyond 7SHIELD	For the same purpose
Storage and access details	
Is the data open?	Yes – Public
Available to 7SHIELD partners?	YES (FLIR data and large animal data without annotation)
Access control	Open access (FLIR data and large animal data without annotation)
What kind of processing is involved?	The images are annotated to have the position of the objects from the classes to discriminate. The FLIR dataset was already annotated but the images from YouTube are annotated by INOV.
What kind of derivative data is produced?	YouTube annotated images for large animals.
How it becomes accessible to stakeholders outside the consortium?	The data is mostly available at FLIR site. Remaining data will be available in place yet to be defined.
Data flows and views	The data is used to train, validate and test the networks, it is not used once a neural network is created.

How can be managed after the project?	FLIR dataset is public. Annotated large animals' data is licensed. It is used for deep learning models creation. Where it will be available is yet to be defined.
License	Proprietary for large animals' data.
Type and format	JPEG, JSON, TXT
Data Size	Approximately 30 GB of data.
Storage location	Project Git
Storing responsible	INOV
Secure storage procedures	INOV's data-centre procedures for backup.
Metadata	Image annotations
Ethics and Data Protection	
Dataset contains personal data?	No. Even though people and vehicles are filmed, the faces and license plates were blurred.
DPIA required	Not Required
Dataset ethics and legal requirements	The dataset does not raise any ethical or legal issues

4.12. DATASET 12: Laser-based technologies for the detection of ground-based and aerial threat detection

Table 4-25: Dataset 12 – Coordinates of the detected intruder

Template Field	Description
Overview	
Dataset Name	Coordinates of the detected intruder
Dataset Category	Primary data collected by partner in 7SHIELD
Partner	DFSL
Provider (if different from partner)	NA
Work Package	WP4
Task/Deliverable	T4.6 Laser-based technologies for the detection of ground-based and aerial threat detection D4.2 Combined Physical and Cyber Threat detection D4.7 Combined Physical and Cyber Threat detection and correlation

Details	
Short Description	Coordinates (Range, Azimuth) of ground-based intruders, Coordinates (Range, Azimuth, Elevation) of aerial intruder drones; using Laser Based Technologies
Existing already before the 7SHIELD project?	YES
Use in 7SHIELD	Will be shared with 7SHIELD platform via JSON through message broker to 7SHIELD Platform for decision making
Use beyond 7SHIELD	NIL
Storage and access details	
Is the data open?	Yes – but restricted access,
Available to 7SHIELD partners?	YES
Access control	Via JSON
What kind of processing is involved?	NA
What kind of derivative data is produced?	NIL
How it becomes accessible to stakeholders outside the consortium?	NA
Data flows and views	Enters into 7SHIELD Platform via JSON through message broker
How can be managed after the project?	NA
License	NIL
Type and format	JSON
Data Size	Can be records, GB, etc.
Storage location	Not on DFSL systems
Storing responsible	NA
Secure storage procedures	NA
Metadata	NA
Ethics and Data Protection	
Dataset contains personal data?	No

DPIA required	No
Dataset ethics and legal requirements	NA

4.13. DATASET 13: Combined Physical and Cyber Threat Detection and Early Warning

Table 4-26: Dataset 13 - Cyber, Physical and Availability UAF alerts

Template Field	Description
Overview	
Dataset Name	Cyber, Physical and Availability UAF alerts
Dataset Category	Derived data
Partner	CSNov
Provider (if different from partner)	Cyber: CeRICT Physical: STWS Availability: CSNov
Work Package	WP4
Task/Deliverable	T4.7 Combined Physical and Cyber Threat Detection and Early Warning D4.2 - Combined Physical and Cyber Threat detection D4.7 - Combined Physical and Cyber Threat detection and correlation
Details	
Short Description	The dataset consists of UAF alerts from three different sources: - Cyber UAF alerts for Cyber detection - Physical UAF alerts for Physical detection - Availability UAF alerts for Availability detection These three sources are correlators in their specific domain.
Existing already before the 7SHIELD project?	No
Use in 7SHIELD	To create combined and correlated UAF alerts.
Use beyond 7SHIELD	For the same purpose
Storage and access details	Stored in 7SHIELD storage module
Is the data open?	No

Available to 7SHIELD partners?	YES, but under EU RESTRICTED mark
Access control	Authorized only 7SHIELD partners and end users
What kind of processing is involved?	The metadata of each UAF alerts are correlated together. The aim is to correlate one type of alert (example Cyber Alert) with another type (example: Physical alert) to identify combined attack scenarios/incidents.
What kind of derivative data is produced?	A new UAF alert which reference the correlated alerts
How it becomes accessible to stakeholders outside the consortium?	Not accessible. RESTRICITED alerts
Data flows and views	The data came from 7SHIELD correlators. The data is used to be correlated. The result is showed to the end user.
How can be managed after the project?	To be defined
License	Proprietary 7SHIELD
Type and format	UAF serialized in JSON
Data Size	Few Gigabytes
Storage location	Where 7SHIELD is deployed (on premise)
Storing responsible	TBD
Secure storage procedures	7SHIELD defined procedure
Metadata	Security alerts
Ethics and Data Protection	
Dataset contains personal data?	Final UAF alerts on a 7SHIELD deployment can contain personal data
DPIA required	Not Required
Dataset ethics and legal requirements	The dataset does not raise any ethical or legal issues

Table 4-27: Dataset 13 – Physical UAF alerts

Template Field	Description
Overview	
Dataset Name	Physical UAF alerts

Dataset Category	Derived data
Partner	STWS
Provider (if different from partner)	- UAV data: ACCELI - Face detection and face recognition from video surveillance: CERTH - Object detection and activity recognition from video content: CERTH - Infrared and thermal image processing for the detection of man-made disasters: INOV - Laser-based technologies for the detection of ground-based and aerial threat detection: DFSL
Work Package	WP4
Task/Deliverable	T4.7 Combined Physical and Cyber Threat Detection and Early Warning D4.2 - Combined Physical and Cyber Threat detection D4.7 - Combined Physical and Cyber Threat detection and correlation
Details	
Short Description	The component will correlate the physical events detected by sensors on the field. The result of this correlation process is a dataset consisting of the related UAF alerts.
Existing already before the 7SHIELD project?	No
Use in 7SHIELD	The dataset will contain the correlation of the physical events detected by the sensors on the field. Each correlation will be represented as an alert. These alerts will be disseminated to the rest 7SHIELD components, for further analysis.
Use beyond 7SHIELD	For the same purpose
Storage and access details	Stored in 7SHIELD storage module
Is the data open?	No
Available to 7SHIELD partners?	YES, but under EU RESTRICTED mark
Access control	Authorized only 7SHIELD partners and end users
What kind of processing is involved?	The physical sensors (UAV, cameras etc) on the field will survey the area, searching for unusual and critical situations. The detected events will be combined and correlated, producing actionable information that will be useful for the rest components. The dataset will be consisted by that actionable information.

What kind of derivative data is produced?	A new UAF alert that references the correlated physical events.
How it becomes accessible to stakeholders outside the consortium?	Not accessible. RESTRICITED alerts
Data flows and views	The data came from 7SHIELD physical sensors on the field. The physical events will be correlated. The correlated alerts will be disseminated to the rest 7SHIELD components for further analysis.
How can be managed after the project?	To be defined
License	Proprietary 7SHIELD
Type and format	UAF serialized in JSON
Data Size	Few megabytes
Storage location	The same as the main 7SHIELD platform.
Storing responsible	-
Secure storage procedures	7SHIELD defined procedure
Metadata	-
Ethics and Data Protection	
Dataset contains personal data?	No. The dataset contains the correlation among the different physical events. The dataset will contain only the information that is related to the correlation. The real physical events will be stored on the resalted sensors datasets. So, it is agnostic to the personal data that each physical event may contain.
DPIA required	Not Required
Dataset ethics and legal requirements	The dataset does not raise any ethical or legal issues

4.14. DATASET 14: Semantic Representation

Table 4-28: Dataset 14 – Semantic Representation

Template Field	Description
Overview	
Dataset Name	Semantic Representation

Dataset Category	Metadata
Partner	CERTH
Provider (if different from partner)	-
Work Package	WP5 Post-Crisis management for response and mitigation of physical and cyber threats
Task/Deliverable	Task 5.1 Semantic representation and linking for decision-making
Details	
Short Description	The dataset will contain the semantic representation of the annotations that are generated by the various 7SHIELD modules. The vocabulary follows the ontology that will be developed in the T5.1
Existing already before the 7SHIELD project?	No
Use in 7SHIELD	This dataset will be used in order to map the incoming data into 7SHIELD KB and to develop the reasoning mechanism
Use beyond 7SHIELD	The ontologies that will be developed with the use of this dataset will be available for further use in other projects in this domain of interest
Storage and access details	
Is the data open?	With restricted access
Available to 7SHIELD partners?	If it's necessary for their services
Access control	The dataset will be stored in 7SHIELD's KB repository. Only authorized users can have access to it.
What kind of processing is involved?	Input data from other 7SHIEID modules (JSON) will be formulated into RDF triples and stored in a native RDF triple store.
What kind of derivative data is produced?	The derivative data will help with the formulation of the 7SHILED knowledge base.
How it becomes accessible to stakeholders outside the consortium?	Need to be specified
Data flows and views	The data are uploaded to the KB in order to upgrade it with new knowledge.
How can be managed after the project?	Need to be specified

License	Apache v2.0
Type and format	RDF
Data Size	Need to be specified
Storage location	Triple Store: GraphDB
Storing responsible	CERTH
Secure storage procedures	Basic local authentication (username/password)
Metadata	-
Ethics and Data Protection	
Dataset contains personal data?	No
DPIA required	No
Dataset ethics and legal requirements	The ethical issues related to the 7SHIELD project will be described in WP9.

4.15. DATASET 15: Data Severity Level

Table 4-29: Dataset 15 – Data Severity Level

Template Field	Description
Overview	
Dataset Name	Data Severity Level
Dataset Category	Derived data (e.g., output from processing by 7SHIELD module)
Partner	CERTH
Provider (if different from partner)	7SHIELD modules: • Situational Picture Generation and Update (SPGU) • 7SHIELD Knowledge Base • Tactical Decision Support System
Work Package	WP5 Post-Crisis management for response and mitigation of physical and cyber threats
Task/Deliverable	T5.3 Security Risk Assessment Algorithms for Decision Support (KR13 Crisis classification module)
Details	

Short Description	This dataset encapsulates characteristics that are derived from the outcome processes of other 7SHIELD modules. In particular, features that describe the current situational picture, semantically information that is stored in Knowledge Base, and other information from the field obtained from TDSS are integrated.
Existing already before the 7SHIELD project?	NO
Use in 7SHIELD	The Crisis Classification module will use this dataset in order to assess the current severity level of the ongoing crisis event.
Use beyond 7SHIELD	Need to be specified
Storage and access details	
Is the data open?	Yes – but restricted access
Available to 7SHIELD partners?	Yes, by Integrated Command Control and Coordination System
Access control	The dataset will be stored in 7SHIELD's data storage repository. Only authorised users can have access to it.
What kind of processing is involved?	The dataset will be annotated in terms of the severity level. The annotated dataset will be used to train a machine learning model enabled to assess the severity level of a crisis event.
What kind of derivative data is produced?	The derivative data present the severity level of a crisis. It will comply with Common Alerting Protocol (CAP) severity code values: • "Extreme" - Extraordinary threat • "Severe" - Significant threat • "Moderate" - Possible threat • "Minor" – Minimal to no known threat • "Unknown" - Severity unknown
How it becomes accessible to stakeholders outside the consortium?	Need to be specified
Data flows and views	The dataset is used in order to train and test (evaluate) a machine learning model which will be enabled to assess the severity level of a P/C crisis.
How can be managed after the project?	Need to be specified
License	Creative commons

	The algorithmic part is public. However, in case this module needs to process and analyse EU RES input, then the output will be of type EU RES too
Type and format	CSV or JSON, need to be specified
Data Size	Depends on the frequency that generated data will be obtained from the 7SHIELD ecosystem. However, a rough estimation could be less than 100 MB.
Storage location	Need to be defined
Storing responsible	Need to be defined
Secure storage procedures	It will be stored in the project repository which is hosted in a server of the project coordinator's IT infrastructure. The repository supports version control which should be enough to ensure data recovery in case of accidental deletions. Data back-ups will be done according to the internal IT policy of the project coordinator as applicable to all other relevant digital data of the company. Access to the data will only be possible through authenticated access to the repository.
Metadata	Metadata standards would be determined in the 7SHIELD project
Ethics and Data Protection	
Dataset contains personal data?	No
DPIA required	No
Dataset ethics and legal requirements	The ethical issues related to the 7SHIELD project will be described in WP9 deliverables.

4.16. DATASET 16: Emergency Response Plan

Table 4-30: Dataset 16 – Emergency Response Plan

Template Field	Description
Overview	
Dataset Name	Emergency Response Plan
Dataset Category	Primary data collected by partner (KEMEA) in 7SHIELD Synthetic / generated data (Data generated from bibliographic research; analysis deriving from processed public data such as for example papers, documents, standards etc. processed)

Partner	KEMEA
Provider (if different from partner)	FMI, SPACEAPPS, SERCO, NOA, DEIMOS, HP
Work Package	5
Task/Deliverable	T5.4/D5.7
Details	
Short Description	Answers retrieved from questionnaires (to be) distributed to and interviews taken by stakeholders and operators from each EU Ground Segments Pilot (i.e. operators' security department, third party companies involved in security, Police, technical teams, building managers, etc.) relation to specific information (such as regulations, best practices and methodologies used).
Existing already before the 7SHIELD project?	NO
Use in 7SHIELD	To draft a generic Emergency Response Plan and further adapt it per Pilot Site, tailored to the specific needs of each scenario per pilot site as developed in Task 2.1.
Use beyond 7SHIELD	NONE
Storage and access details	
Is the data open?	No (Classified Information: RESTREINT UE)
Available to 7SHIELD partners?	YES (the D5.7 report)
Access control	Following handling EUCI procedures (e.g. use o encryption software to exchange and read the report)
What kind of processing is involved?	Processes on EUCI handling, based on Commission Decision 2015/444/EC and the developed PSMP by the SAB in order to support the consortium (e.g. encryption/decryption process through approved software, in order to exchange the classified deliverable between th-e partners involved)
What kind of derivative data is produced?	No derivative data.
How it becomes accessible to stakeholders outside the consortium?	Only after need-to-know basis and approval by the SAB
Data flows and views	A model plan will be created and adapted, used per use case, integrating steps and guidelines that will describe

	into the 7SHIELD system, supporting end user decision making
How can be managed after the project?	Only generic data that are not linked to CIs and are not contradictory with the Commission Decision 2015/444/EC and the developed PSMP
License	Proprietary
Type and format	.pdf, .doc
Data Size	~5mb
Storage location	Partners' PCs and project MS TEAMS
Storing responsible	KEMEA
Secure storage procedures	Classified information in electronic format will be stored encrypted through ZED! or following any other procedure indicated by the PSMP (e.g. secure room/administrative area for hardcopies)
Metadata	N.A.
Ethics and Data Protection	
Dataset contains personal data?	• Deliverable to be submitted in August 2022. Data expected to be collected: name/surname & email. • Informed consent will be given for use/reuse. • Personal data will be anonymised.
DPIA required	Data Privacy Impact Assessment Required - NO
Dataset ethics and legal requirements	D9.1, Recruitment of participants, Informed Consent. Requirement respected.

4.17. DATASET 17: Social Awareness

Table 4-31: Dataset 17 – Social Awareness

Template Field	Description
Overview	
Dataset Name	Social Awareness (official and public)
Dataset Category	Primary data collected by partner in 7SHIELD
Partner	CENTRIC
Provider (if different from partner)	

Work Package	WP5
Task/Deliverable	T5.5 / D5.4
Details	
Short Description	This dataset will collect information from social media sites, groups and other formal online communication. The 'official' dataset will include data from public authorities, GSS operators and other organisations communicating in an official capacity. The 'public' dataset will include a set of community messages from citizens about specific incidents
Existing already before the 7SHIELD project?	NO
Use in 7SHIELD	The 'official' data will be used to inform the development of guidelines for crisis communications and be analysed to understand what makes certain methods of communication more effective. The public dataset will support this understanding to what extent social communications may be able to inform early warnings of future incidents.
Use beyond 7SHIELD	The data could be used by other researchers in the field of crisis communications.
Storage and access details	
Is the data open?	Possibly – data would require strict anonymisation procedures and would also have to comply with social media platforms' terms of service
Available to 7SHIELD partners?	YES
Access control	Data will be available in a tabular format (e.g., csv, xlsx or similar)
What kind of processing is involved?	Statistical processing and text processing techniques
What kind of derivative data is produced?	Keywords, potential warning indicators and similar may be extracted
How it becomes accessible to stakeholders outside the consortium?	TBC
Data flows and views	Data is not part of the formal 7SHIELD system, it is for research and analysis purposes
How can be managed after the project?	TBC – the raw data will be reviewed to understand the potential security context, extent of personal data and level

	of anonymisation required, and terms of services from collection platforms
License	TBC
Type and format	CSV or similar tabular format
Data Size	TBC – expected to be 1000s of records spread across multiple use cases
Storage location	Sheffield Hallam University Research Data Archive
Storing responsible	Task Leader in CENTRIC
Secure storage procedures	Strict access procedures will be defined on deposit
Metadata	TBC
Ethics and Data Protection	
Dataset contains personal data?	Dataset may contain personal data (e.g. names / usernames). Such information will be at least be pseudonymised, although full anonymisation would be sought before any wider sharing of datasets beyond the task.
DPIA required	Yes – a DPIA will be carried out prior to any data collection activity
Dataset ethics and legal requirements	Ethical approval for the research has been obtained through the Sheffield Hallam University Research Ethics Committee under application number ER27797704

4.18. DATASET 18: Pilot Critical Operation

Table 4-32: Dataset 18 - Pilot Critical Operation

Template Field	Description
Overview	
Dataset Name	Pilot Critical Operations Dataset
Dataset Category	Primary data collected by partner in 7SHIELD
Partner	KEMEA, ENG, INOV, and RG as needed
Provider (if different from partner)	FMI, NOA, DEIMOS, SPACEAPPS, SERCO
Work Package	WP5
Task/Deliverable	T5.6

Details	
Short Description	Data in standardized format on (i) how operations are conducted by each pilot case study, (ii) available infrastructure (cyber and physical), (iii) available man-power, (iv) available on/off-site physical & cyber resources that can be leveraged to mitigate impacts
Existing already before the 7SHIELD project?	NO
Use in 7SHIELD	Data will be employed to determine optimal business continuity scenarios per partner in T5.6
Use beyond 7SHIELD	This primary data cannot be used beyond 7SHIELD due to its highly sensitive nature.
Storage and access details	
Is the data open?	No
Available to 7SHIELD partners?	YES, subject to end-user confidentiality requirements
Access control	Verified access only to partners through project data curation platform.
What kind of processing is involved?	Standardization of data through the adoption of a common taxonomy for operations, resources and infrastructure.
What kind of derivative data is produced?	Business continuity scenarios per T5.6
How it becomes accessible to stakeholders outside the consortium?	Only the end-user to which each dataset pertains can provide access to it
Data flows and views	Data enters the system via detailed surveys and interviews with key personnel of each end user
How can be managed after the project?	None
License	Proprietary
Type and format	CSV, JSON
Data Size	Maximum 10MB per end-user
Storage location	Project internal data curation platform
Storing responsible	ENG, RG
Secure storage procedures	Encrypted storage, access only via verified login

Metadata	None
Ethics and Data Protection	
Dataset contains personal data?	No
DPIA required	None
Dataset ethics and legal requirements	Confidentiality per end-user requirements

Table 4-33: Dataset 18 - Pilot Critical Operations - Anonymized

Template Field	Description
Overview	
Dataset Name	Pilot Critical Operations Dataset - Anonymized
Dataset Category	Primary data collected by partner in 7SHIELD
Partner	KEMEA, ENG, INOV, and RG as needed
Provider (if different from partner)	FMI, NOA, DEIMOS, SPACEAPPS, SERCO
Work Package	WP5
Task/Deliverable	T5.6
Details	
Short Description	This is data similar in nature to the "Pilot Critical Operations Dataset", only anonymized & virtualized to remove any references and information that pertain to any specific partner, thus becoming a benchmark generic dataset, applicable to all but specific to no end-user.
Existing already before the 7SHIELD project?	NO
Use in 7SHIELD	Data will be employed for dissemination of the findings of T5.6
Use beyond 7SHIELD	This primary data is fully usable beyond 7SHIELD to help researchers and other operators derive business continuity scenarios and optimize their operation and emergency response planning.
Storage and access details	
Is the data open?	Yes – Public;

Available to 7SHIELD partners?	YES
Access control	How can the data be accessed? (software, techniques, credentials, key pair, etc.)
What kind of processing is involved?	Standardization of data through the adoption of a common taxonomy for operations, resources and infrastructure. Anonymization of data to avoid any relation with specific end-users
What kind of derivative data is produced?	Business continuity scenarios per T5.6
How it becomes accessible to stakeholders outside the consortium?	Through public archives and project website
Data flows and views	Data provided by anonymization and virtualization of he "Pilot Critical Operations Dataset"
How can be managed after the project?	Data to be fully and publicly available after the project
License	Creative commons
Type and format	CSV, JSON
Data Size	Maximum 10MB
Storage location	Zenodo, GitHub
Storing responsible	ENG, RG
Secure storage procedures	None
Metadata	None
Ethics and Data Protection	
Dataset contains personal data?	No
DPIA required	None
Dataset ethics and legal requirements	None

5. 7SHIELD IPR Plan

The management of IPR is strictly ruled by the Consortium Agreement (CA) which includes all provisions related to the management of IPR including ownership, protection and publication of knowledge, access rights to knowledge and pre-existing know-how as well as questions of confidentiality, liability and dispute settlement. The IPR Management focuses on the careful handling of IPR issues in 7SHIELD project, that are of strategic importance. It aims to create a favourable environment for respecting intellectual property rights (IPR) and ensuring a uniform approach by the Consortium, in conjunction with a permanent IPR monitoring during the project.

5.1. IPR Strategy

For a successful project, it is important to have an **IPR strategy** in place, so that all partners of the consortium work collaboratively towards the achievement of common objectives. The 7 SHIELD IPR strategy is focused and concise with the aim to protect the results developed within the timeframe of the project with a set of agreements that are planned (D8.6, D8.11) to clearly identify:

1) IPR ownership;
2) Exploitation rights (and royalties) by owners and co-owners;
3) IPR protection (e.g. copyright law, patent law, trademarks law, etc.) and filing of applications where applicable;
4) IPR exploitation strategy at consortium and partner level.

This will help in maximizing the returns on the human, capital and intellectual investments.

Regarding the management of knowledge, intellectual property and other aspects of innovation two types of activity are foreseen:

1) IPRs for new systems and solutions that are prepared by consortium partners;
2) Information disseminated within the project and to external bodies, such as publications, presentations and regulatory and standards bodies, only after the necessary steps for ensuring the protection of IPRs have been made. This ensures that IP will be secured in the interest of project partners. Contributions to external bodies will have an impact on global harmonisation of concepts and systems. The dissemination of information and the influence, e.g., on standards bodies, are the prerequisites for the economic success of IPRs.

5.2. IPR Management

IPR will be handled in line with general EU policies regarding ownership, exploitation rights, confidentiality and commercial utilization of results to other EU funded projects and disclaiming rules. Specific actions will be taken in order to satisfy the basic intellectual property regime: IPRs are attributed to those who generate the results, and results are owned by those who generate them); publications will be made by the owners of the background or results for their background or results and publications are owned by those who write the article; 7SHIELD will aim at open access publications; Some of the project software modules will be Open Source, allowing operators to start using the 7SHIELD services without a major up-front investment.

Following the CA, the Results are owned by the Party that generates them. Background remains with the respective party.

In case of Joint Ownership the joint owner shall be entitled to use the jointly owned Results for non- commercial research activities on a royalty-free basis, and without requiring the prior consent of the other joint owner(s), and each of the joint owners shall be entitled to otherwise Exploit the jointly owned Results and to grant non-exclusive licenses to third parties (without any right to sub- license), if the other joint owners are given a 45 calendar days advance notice (at least) and Fair and Reasonable compensation.

Access Rights to Results and Background are granted on a non-exclusive basis (sublicensing is excluded) and on a royalty-free basis, if they are used for the performance of the work of a Party under the Project (implementation). After the completion of the project, Access Rights to Results for exploitation are granted on Fair and Reasonable conditions (unless access is made for internal research activities in which case access is granted on a royalty-free basis). Access Rights to Background for Exploitation, including for research on behalf of a third party, shall be granted on Fair and Reasonable conditions.

The CA also foresees clauses in relation to rights and obligations for affiliated entities of a party; such entities shall enjoy the access rights on Fair and Reasonable conditions, upon written bilateral agreement and under the condition that they all confidentiality and other obligations accepted by the Parties under the GA and CA (Article 9.5). Access may be refused if such granting is contrary to the legitimate interests of the Party which owns the Background or the Results.

Specific IPR agreements will be released later in the project but before its end to ensure the exploitation strategy is identified before the exploitation activities will begin.

5.3. Background Data

Background means any data, know-how or information – whatever its form or nature (tangible or intangible), including any rights such as intellectual property rights – that is:

- Held by the beneficiaries before they acceded to the Agreement

- Needed to implement the action or exploit the results.

In relation to access rights to Background, the consortium will comply with the H2020 default rules (European IPR helpdesk 2015); it means that:

- On the one hand, the partners/beneficiaries will provide each other access — on a royalty-free basis — to Background needed to implement their own tasks under the action/project;
- On the other hand, the partners will give each other access — under fair and reasonable conditions — to background needed for exploiting their own results.

At proposal stage it has been confirmed that all partners have the access they need to each other background; in the Consortium Agreement (CA) signed between partners, the conditions regarding the need to implement exploitation actions are outlined.

5.4. Results

In relation to protection of the results, the partners will aim to protect all results that can be commercially or industrially exploited as appropriate (i.e. the form which offers the most adequate and effective protection for each type of result). In particular:

- Results shall be owned by the participant generating them,
- Results generated jointly (e.g. 7SHIELD platform), will be jointly owned,
- Access rights to Results. In keeping with the H2020 default rules the partners will give each other access — on a royalty-free basis — to results needed for implementing their own tasks under the action. The partners will give each other — under fair and reasonable conditions — access to results needed for exploiting their own results.

6. Conclusion and Future Outlook

This deliverable constitutes the initial Data Management Plan (DMP) and outlines the datasets which have already been collected / generated or are foreseen to be collected / generated in 7SHIELD with detailed information about their content and their FAIR ability. It provides the description of data types and sources, explains how the FAIR principles will be respected in the project, explores the security measures, allocation of resources, ethical and legal issues related to the data collection and generation in the project. It will describe in more details the procedures to be applied further in the 7SHIELD project to efficiently manage its research data in terms of storage and backup (backup provision, recovery procedure), selection and preservation (which data will be retained/shared/ preserved, length of time data to be preserved and preservation preparation time). It will also provide a clearer vision on the data management with respect to the technologies developed in the project. Moreover, the 7SHIELD IPR plan that focuses on the careful handling of IPR issues in the project has been also illustrated.

This document can evolve along the project and this is the initial version of the project's DMP and it will be updated in D1.4 Self-assessment & data management plan v2 (M18). The update will include new sets of data and changes in consortium policies and datasets management.

7. References

[1] Stiefelhagen, R., Bernardin, K., Bowers, R., Garofolo, J., Mostefa, D. and Soundararajan, P. (2006). "The CLEAR 2006 evaluation," presented at the Internat. evaluation workshop on classification of events, activities and relationships, pp. 1–44.

[2] Raad, J., & Cruz, C. (2015). A Survey on Ontology Evaluation Methods. https://doi.org/10.5220/0005591001790186

[3] Compton, M. (2012). The SSN ontology of the W3C semantic sensor network incubator group. Web semantics: science, services and agents on the WWW, 25.

[4] Hobbs, J. R., & Pan, F. (2006). Time ontology in OWL. W3C working draft, 27, 133.

[5] Perry, M., & Herring, J. (2012). OGC GeoSPARQL-A geographic query language for RDF data. OGC Implementation Standard, ref: OGC.

[6] Brickley, D., & Miller, L. (2014). FOAF Vocabulary Specification 0.99. Namespace Document 14 January 2014-Paddington Edition. Retrieved May, 3, 2016.

[7] Mascardi, V., Cord., V., & Rosso, P. (2007, September). A Comparison of Upper Ontologies. In WOA (Vol. 2007, pp. 55-64).

[8] European IPR Helpdesk Fact Sheet Open Access to scientific publications and research data in Horizon 2020: Frequently Asked Questions (FAQs), p.2

Annex I - Dataset Management Template

The following template has been developed (in line with the DMP H2020 template requirements) to specifically gather information on the various 7SHIELD datasets.

Table 8-34: Dataset Management Template

Template Field	Description
Overview	
Dataset Name	A unique and descriptive name for the dataset
Dataset Category	A category the Dataset: - Project management data - Primary data collected by partner in 7SHIELD - Secondary data (not publicly available) - Derived data (e.g., output from processing by 7SHIELD module) - Publicly available dataset (e.g. training / benchmark data) - Synthetic / generated data
Partner	The responsible partner for the dataset
Provider (if different from partner)	
Work Package	
Task/Deliverable	If applicable
Details	
Short Description	A short description for the dataset and how it was collected
Existing already before the 7SHIELD project?	YES/NO
Use in 7SHIELD	How the data is/will be used in 7SHIELD
Use beyond 7SHIELD	How the data could be useful to other researchers beyond 7SHIELD
Storage and access details	
Is the data open?	Yes – Public; Yes – but restricted access, No TBC
Available to 7SHIELD partners?	YES/NO

Access control	How can the data be accessed? (software, techniques, credentials, key pair, etc.)
What kind of processing is involved?	
What kind of derivative data is produced?	
How it becomes accessible to stakeholders outside the consortium?	
Data flows and views	Describe the data flows and views, i.e. how the data enters the system
How can be managed after the project?	Specify which part of the data can be made publicly available beyond the project end, for which purpose, under which license, and through what kind of infrastructure.
License	Specify license (e.g. Creative commons, Proprietary, etc.)
Type and format	e.g. CSV, TSV, XML, JSON, etc.
Data Size	Can be records, GB, etc.
Storage location	Where will it be stored; (if open, specify repository)
Storing responsible	Who is responsible for storing the data
Secure storage procedures	Which secure storage procedures are used for storing the dataset?
Metadata	Any metadata standards used in the dataset. What metadata has been created and how is this managed
Ethics and Data Protection	
Dataset contains personal data?	Does the dataset contain personal data, and if YES, which? (For details about personal data please see: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en) Was informed consent given for use/reuse? Is personal data anonymised?
DPIA required	Data Privacy Impact Assessment Required
Dataset ethics and legal requirements	Please describe any Ethics and/or legal requirements relevant to the dataset - If you want you can refer to relevant Ethics deliverables.

Annex II – Participant Consent Form

7SHIELD: Safety and Security Standards of Space Systems, ground Segments and Satellite data assets, via prevention, detection, response and mitigation of physical and cyber threats

I volunteer to participate in this research conducted as part of the research project 7SHIELD, coordinated by Engineering Ingegneria Informatica and funded by the European Commission. The information sheet has been made available to me.

Please place an "X in the boxes" to affirmatively consent to the following statements.

☐ I confirm that I have read and understood both this form and the accompanying Information Sheet. I had the time and opportunity to ask questions as needed.
☐ I understand that I am free to withdraw my consent at any time without giving reason.
☐ My personal data can be gathered to be used, stored and shared in the ways described on the accompanying Information Sheet.
☐ Data from my participation can be used to inform 7SHIELD system requirements, revise design, develop 7SHIELD technologies and subsequent evaluation activities.
☐ Data from my participation may be used to in articles for peer-reviewed journals and relevant industry magazines, for presentations at conferences and workshops,
☐ *Data from my participation may be used in the promotion of 7SHIELD in general.
☐ *7SHIELD may take research notes or audio recordings of my activities
☐ *I give my consent to be identified in any public reports.
☐ *I consent to having photos or videos taken of me for research purposes.
☐ *I consent to having photos or videos taken of me for communication purposes.
☐ *I agree to be quoted directly.
☐ I would like to receive updates on the progress and findings of the project.
☐ I agree to voluntarily take part in the 7SHIELD research.
☐ I agree to voluntarily take part in the 7SHIELD research.

*Optional

Participant Name		Participant Signature	
Participant Email*		Date	

Researcher Name		Researcher Signature	
Researcher Email		Date	
Researcher Organization			

Project Coordinator		Coordinating Organization	
Coordinator Email		Date	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883284