



Reinforcing the AI4EU Platform by Advancing Earth Observation Intelligence, Innovation and Adoption

## D2.1 Data Management Plan

<b>Grant Agreement ID</b>	101016798	<b>Acronym</b>	AI4COPERNICUS
<b>Project Title</b>	Reinforcing the AI4EU Platform by Advancing Earth Observation Intelligence, Innovation and Adoption		
<b>Start Date</b>	01/01/2021	<b>Duration</b>	36 Months
<b>Project URL</b>	<a href="https://ai4copernicus-project.eu/">https://ai4copernicus-project.eu/</a>		
<b>Contractual due date</b>	30/06/2021	<b>Actual submission date</b>	30/06/2021
<b>Nature</b>	R = Document, report	<b>Dissemination Level</b>	CO = Confidential
<b>Author(s)</b>	S. Auriol (TAS), M. Aubrun (TAS), A. Nousias (NCSR), A. Koukourikos (NCSR), X. Ziouvelou (NCSR), V. Karkaletsis (NCSR)		
<b>Contributor(s)</b>	M. Lazzarini (SatCEN), V. Baousis (ECMWF), G. Weikmann (UNITN), R. Hall (Equinor), J. Konarski (CF)		
<b>Reviewer(s)</b>	O. Barrilero (SatCEN), R. Hall (Equinor)		



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

**Document Revision History (including peer reviewing & quality control)**

Version	Date	Changes	Contributor(s)
v1.0	11/05/2021	First version of the document	S. Auriol, M. Aubrun
v1.1	25/06/2021	Completed section 2, 3, 4 et 6	A. Koukourikos, A. Nousias, X. Ziouvelou, V. Karkaletsis
v1.2	29/06/2021	Reviewed version	O. Barrilero, R. Hall, M. Aubrun
v1.3	30/06/2021	Incorporation of review comments <b>(pending review by EC)</b>	M. Aubrun, A. Koukourikos

Draft Version

## Executive Summary

The deliverable describes the Data Management policy that will be followed in the course of the project. In more detail, the report lays out the AI4Copernicus Research data management to elaborate; presents an overview of the datasets used by the project; specifies FAIRness to achieve for the methods, tools and data produced by AI4Copernicus; list the data management activities to take care; and it introduces GDPR principles.

The Data Management Plan will be a living document that will be regularly updated by the project partners in order to align with the needs and requirements of the project. As such updated versions of this document will be provided during M18 and M36 (in time with the periodic evaluation/assessment of the project) updating all the necessary sections in light of the needs and findings of the project.

D2.1, as noted in section 5.1 'Ethics' of Grant Agreement Part B, includes also 'Annex A: Ethics Requirements' to provide additional information with respect to the requirements issued in the Ethics Summary Report.

Draft Version

## Table of Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Purpose and Scope	6
1.2	Approach for Work Package and Relation to other Work Packages and Deliverables	6
1.3	Methodology and Structure of the Deliverable	6
<b>2</b>	<b>Data Summary</b>	<b>7</b>
2.1	WP4 datasets	7
2.2	WP5 datasets	7
2.3	Open Call datasets	10
<b>3</b>	<b>FAIR principles</b>	<b>11</b>
3.1	Making data findable, including provisions for metadata	11
3.2	Making data openly accessible	11
3.3	Making data interoperable	12
3.4	Increase data re-use (through clarifying licenses)	12
<b>4</b>	<b>Allocation of Resources</b>	<b>13</b>
<b>5</b>	<b>Data Security and Protection</b>	<b>14</b>
5.1	Security scope in the context of AI4Copernicus	14
5.2	Data protection and GDPR compliance	14
5.2.1	GDPR scope	14
5.2.2	Data protection form	14
5.2.3	GDPR procedure	14
<b>6</b>	<b>Conclusions</b>	<b>16</b>
	<b>Annex A: Ethics Requirements</b>	<b>17</b>

## List of Tables

Table 2.1: WP5 input data.....	7
Table 2.2: WP5 output data .....	9
Table 2.3: Open Calls input data .....	10
Table 2.4: Open Calls output data.....	10
Table 4.1: Data Management activities .....	13

## List of Terms & Abbreviations

Abbreviation	Definition
<b>DMP</b>	Data Management Plan
<b>FAIR</b>	Findable, Accessible, Interoperable and Reproducible
<b>GDPR</b>	General Data Protection Regulation
<b>WP</b>	Work Package
<b>IPR</b>	Intellectual Property Right
<b>EC</b>	European Commission
<b>EU</b>	European Union
<b>AI</b>	Artificial Intelligence
<b>GA</b>	Grant Agreement
<b>CA</b>	Consortium Agreement
<b>PB</b>	Petabytes
<b>SAR</b>	Synthetic Aperture Radar
<b>ESA</b>	European Spatial Agency
<b>CISO</b>	Chief Information Security Officer
<b>NDA</b>	Non-Disclosure Agreement
<b>DPO</b>	Data Protection Officers

## 1 Introduction

### 1.1 Purpose and Scope

The purpose of deliverable **D2.1: Data Management Plan** is to describe the data management life cycle for the data to be collected, processed and/or generated by the AI4Copernicus project and the actions endorsed via its Open Calls programme. As part of making research data findable, accessible, interoperable and reusable (FAIR), the project's Data Management Plan (DMP) includes information on the handling of research data during & after the end of the project; what data will be accessed, collected, processed and/or generated; which methodology & standards will be applied; whether data will be shared/made open access; how data will be curated & preserved (including after the end of the project).

The information in this document does not supersede the rules and conditions laid out in the AI4Copernicus Grant Agreement (GA) and those in the AI4Copernicus Consortium Agreement (CA).

### 1.2 Approach for Work Package and Relation to other Work Packages and Deliverables

From an organisational point of view, the present D2.1 deliverable is a direct outcome of T2.1: User requirements and definition of use-case scenarios. However, its scope extends to most major AI4Copernicus activities and particularly those under WPs 4, 5 and 6. Its purpose is to ensure that data generated and published within the project are appropriately licensed, openly accessible as much as possible, adequately described, contextualised and documented, and in accordance with all major EU and national regulations related to data integrity, security and privacy.

As activities of the project evolve, all aforementioned aspects will likely be clarified or modified. To accommodate this reality, the DMP will be a living document, regularly reviewed and revised in order to incorporate and manage currently unforeseen cases and settings of data usage and production in the projects and the actions carried out in the context of the Open Calls programme.

### 1.3 Methodology and Structure of the Deliverable

The deliverable is structured in accordance with the template and guidelines provided by the EC, and is organised in the following sections: Section 2 provides a data summary addressing issues regarding the purpose of the data access/collection/generation and its relation to the objectives of the project, the types and formats of data the project will generate/collect, the origin of data, the expected size of data, and the data utility (i.e., to whom might it be useful).

Section 3 reports on the measures and directions to be adopted to ensure the compliance of AI4Copernicus with FAIR data principles. Section 4 summarises the allocation of resources and a preliminary cost coverage plan in the context of the project, in order to serve the aforementioned measures. Section 5 discusses the main concerns regarding data security and privacy and the proposed approaches to face them, while Section 6 concludes D2.1.

Annex A: Ethics Requirements, as noted in section 5.1 'Ethics' of Grant Agreement Part B, provides additional information with respect to the requirements issued in the Ethics Summary Report.

## 2 Data Summary

At the core of AI4Copernicus objectives and ambition is the combination of Copernicus data and services with domain-specific data sources towards the provision of innovative AI solutions tailored to the needs of each domain and exploiting the full potential of Copernicus data.

Copernicus and third-party data sources will be mainly accessed and processed by the consortium partners in the context of *WP4: Implementation, customisation, integration and testing* and *WP5: Bootstrapping AI4Copernicus with high-impact services*, and used for implementing AI solutions via the project's Open Calls, organised and monitored under *WP6: Technology transfer via AI4Copernicus Open Calls*.

The following subsections describe all types of data exploited and generated by each aforementioned WP, based on our current view and understanding of AI4Copernicus reach. These lists will be further expanded and concretised as the project evolves.

### 2.1 WP4 datasets

The tools, machine learning models and use cases implemented, integrated and tested in the context of WP4 will have access to the entire CREODIAS repository (over 22 PB of free data described under <https://creodias.eu/data-offer>), where data are stored unzipped and immediately accessible from the computing cloud. For particular use cases, sample data from new collections, products, particular regions outside of Europe or VHR data may be ordered on demand.

Use-case prepared for the project [WP4] has shown the ability to create new dataset using access to EO data on CREODIAS repository and pipelines designed for AI exploitation.

### 2.2 WP5 datasets

The bootstrapping services have been designed to facilitate the task of open call bidders, who have the possibility to retrieve the work already done to build their solutions. These bootstrapping services not only include services developed by WP5 partners but also third party datasets, which will be not processed but have been identified due to their relevance to address domain-specific problems.

Copernicus and third party datasets that will be referenced or used as input to the bootstrapping services are incorporated in table 2.1, whereas data that will be generated by these services are detailed in table 2.2.

**Table 2.1: WP5 input data**

Dataset	Summary	Format(s)
Sentinel-1 <sup>1</sup>	<p>Sentinel-1 satellite images are SAR datasets produced as part of the EU Copernicus program. These data are collected from Sentinel-1 mission that comprises a constellation of two polar-orbiting satellites.</p> <p>Images coming from the Sentinel-1 mission are acquired in single or dual polarization with different spatial resolution depending on the acquisition mode (e.g. 5 x 5 (SM Mode), 10 x 10 m (IW mode)). With both satellites (S1A and S1B), the revisiting time is 6 days with the same orbital parameters.</p> <p>Dataset from Level-1 SLC and Level-1 GRD formats will be used.</p>	ESA SAFE

<sup>1</sup> <https://sentinel.esa.int/web/sentinel/missions/sentinel-1>

Sentinel-2 <sup>2</sup>	<p>Sentinel-2 satellite images are optical datasets produced as part of the EU Copernicus program. These data are collected from Sentinel-2 mission that comprises a constellation of two polar-orbiting satellites.</p> <p>Images coming from the Sentinel-2 mission are characterized by 13 bands, a spatial resolution between 10 and 60 m and a revisiting frequency of 5 days at the equator.</p> <p>Datasets from Level-2A (with atmospheric corrections) will be used.</p>	ESA SAFE
OpenStreetMap <sup>3</sup>	<p>OpenStreetMap is a collaborative project to create a free editable map of the world with an open-content license. The OpenStreetMap License allows free (or almost free) access to the map images and all of the underlying map data.</p>	OSM XML
The European Marine Observation and Data Network (EMODnet)	<p>EMODnet is a long term marine data initiative from the European Commission Directorate-General for Maritime Affairs and Fisheries (DG MARE) underpinning its Marine Knowledge 2020 strategy. The main purpose of EMODnet is to unlock fragmented and hidden marine data resources and to make these available to individuals and organisations (public and private), and to facilitate investment in sustainable coastal and offshore activities through improved access to quality-assured, standardised and harmonised marine data which are interoperable and free of restrictions on use.</p>	Depending on the datasets
European Forest Fire Information System (EFFIS) database	<p>The European Fire Database contains the forest fire information compiled by EU Member States and other European countries. Since 2004 the forest fire data provided each year by individual EU Member States and other European countries are checked, stored and managed by JRC within EFFIS. At present the database contains fire data from 22 countries: Bulgaria, Croatia, Cyprus, Czech, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Latvia, Lebanon, Lithuania, Poland, Portugal, Romania, Slovakia, Spain, Sweden, Switzerland, Turkey. The public access to the database currently allows the users to retrieve general information such as maps of the number of fires for a selected year and for the countries for which data are available</p>	ESRI shapefile, DBF, GeoJSON, KML, MS office open XML
Data from the CAMS Atmosphere Data Store (ADS) and CAMS European air quality forecasts ADS web interface	<p>The Atmosphere Data Store (ADS, <a href="http://ads.atmosphere.copernicus.eu">http://ads.atmosphere.copernicus.eu</a>) is the operational data access portal of the Copernicus Atmosphere Monitoring Service (CAMS), which is implemented by ECMWF on behalf of the European Commission. It provides access to the CAMS portfolio of products through a web interface and API. The portfolio of products can be searched directly from the ADS. Evaluation and Quality Control information about the CAMS products is also available on the ADS.</p>	GRIB, NetCDF and CSV
Data and data API from Copernicus	<p>The Climate Data Store (CDS, <a href="http://cds.climate.copernicus.eu">http://cds.climate.copernicus.eu</a>) is the operational data access portal of the Copernicus Climate Change Service (C3S), which is implemented by ECMWF on behalf of the European</p>	GRIB, NetCDF and CSV

<sup>2</sup> <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>

<sup>3</sup> <https://www.openstreetmap.org>



Climate Data Store (CDS)	Commission. It provides access to the C3S portfolio of products through a web interface and API. It also offers a toolbox to develop applications. The portfolio of products can be searched directly from the CDS. Evaluation and Quality Control information about the C3S products is also available on the CDS.	
POD, the dissemination Platform for Operational Data	Service to access and request datasets from innovation project activities and operational assets in the renewable energy domain. The latters can be obtained from ORE Catapult	Depending on the datasets
TimeSen2Crop	TimeSen2Crop is a pixel-based dataset of more than 1 million samples of Sentinel 2 Time Series associated with 16 crop types. The dataset contains atmospherically corrected samples with information regarding the coverage (clean, snow, cloud, and shadow).	CSV

Table 2.2: WP5 output data

Dataset	Used tool	Summary	Output format
pre-processed S1 images	SNAP, GDAL	Pre-processing chain using operators (e.g., calibration, terrain correction) to create Analysis Ready Data from original Sentinel-1 products.	Geotiff images
pre-processed S2 images	SNAP, GDAL	Pre-processing chain using operators (e.g., resample) to create Analysis Ready Data from original Sentinel-2 products.	Geotiff images
Binary Map of Changes from S1	SatCen SAR change detection chain based on SNAP, GDAL, OTB	SAR change detection maps generated using Sentinel-1 images.	Geotiff images
Binary Map of Changes from S2	SatCen Optical change detection chain based on SNAP, GDAL, OTB	Optical change detection maps generated using Sentinel-2 images.	Geotiff images
Vector data of human features	SatCen vector converter	Vector files generated with OSM data and converted to a data dictionary relevant for Security.	Vector format
Pre-processing of pollution data and data workflow	CliMetLab, Python and JupyterLabs	Collect existing software solutions from the AI4EO pollution challenge and existing studies and apply them on DIAS platforms to provide vanilla solutions for machine learning studies	Vector format in Python framework
Pre-processing of greenhouse gas concentration data and data workflow	CliMetLab, Python and JupyterLabs	Collect and apply existing software solutions to handle greenhouse gas concentration data and emission forecasts and apply them on DIAS platforms to provide vanilla solutions for machine learning studies	Vector format in Python framework
Pre-processing of weather information and datasets for disease spread	CliMetLab, Python and JupyterLabs	Collect and apply existing software solutions to <b>disease</b> spread data handling and apply them on DIAS platforms to provide vanilla solutions for machine learning studies	Vector format in Python framework

Harmonization of pre-processed S2 data	Python and JupyterLabs	Creation of monthly composite from Sentinel-2 images to alleviate the cloud problem and compare different spatial area in a harmonized temporal dimension	Geotiff images
Crop Type Map from S2 harmonized data	Python and JupyterLabs	Crop Type Maps generated by Long Short-Term Memory using harmonized time series of images	Geotiff images, Shapefiles

### 2.3 Open Call datasets

Regarding Open Calls-related data, WP6 will collect data of the applicants participating in the project's open calls, through an online form within the Open Calls Management platform. The information gathered will serve to evaluate and select the most promising projects, depending on the type of Open Call. As such, it will be necessary to collect, store and process data from the online application forms submitted by project applicants. These data will be exploited for following main purposes: (a) Assessment of Eligibility Criteria, (b) Evaluation of proposals, (c) Impact/Policy assessments, (d) Research (e.g., analysis of usage statistics, etc.).

The anonymised datasets will be exploited through the creation of figures and charts that will be updated at the end of the selection process of each open call. The figures and charts generated will be publicly shown as part of the dissemination activities of the project (e.g., AI4Copernicus website, etc.). The full data set of anonymised data will also be available only to AI4Copernicus project partners and for the purpose of providing support to the selected projects and for research purposes. Access to anonymised data will be provided only to specific third parties, such as the AI4Copernicus Advisory Board, and for the purposes of the Open Call Activities.

**Table 2.3: Open Calls input data**

Dataset	Summary
Datasets that will be used by the Open Call participants	<p>The potential data sources that the Open Call participants may use, for example: Sentinel-1 and Sentinel-2 among others.</p> <p>Sentinel-1 satellite images are SAR datasets produced as part of the EU Copernicus program. These data are collected from Sentinel-1 mission that comprises a constellation of two polar-orbiting satellites.</p> <p>Sentinel-2 satellite images are optical datasets produced as part of the EU Copernicus program. These data are collected from Sentinel-2 mission that comprises a constellation of two polar-orbiting satellites.</p>
Applicant Data	<p>The datasets that will be collected during the open calls, will be primarily utilized in order to facilitate the analysis of submitted proposals including information that relates to (non-exhaustive list): Country; Organization name; Project name; Address; Number of team members; Name of the team members; Industrial Domains Targeted; Abstract; Brief description, Technological Focus, etc. This data will be provided by applicants in the application forms (a detailed overview of this information will be provided in the course of WP6, D6.1, 1st Round of Open Calls Documentation). This data will be anonymised and used only for the purposes of the project open calls.</p>

**Table 2.4: Open Calls output data**

Dataset	Used tool	Summary	Output format
Data that will be produced by the participants of the Open Calls	Unknown	Data that will emerge as the output of the Use-cases and experiments that the Open Call participants will use in the course of their projects.	Unknown

### 3 FAIR principles

As one of the main objectives of AI4Copernicus is the promotion and facilitation of the usage of Copernicus assets, it is critical to ensure that project data results adhere to the FAIR principles, that is they are Findable, Accessible, Interoperable and Reproducible through different means, including the AI4EU catalogue of resources. Additionally, and as described in the previous sections, AI4Copernicus will try to rely on input data (outside the project's sphere of control) that also adhere as much as possible to the FAIR principles

The following subsections describe the main actions and directions that will be examined in order to ensure that FAIRness is fulfilled to the maximum possible extent for data generated in the context of AI4Copernicus.

#### 3.1 Making data findable, including provisions for metadata

To enable and facilitate the fast and contextualised discovery of data and knowledge, AI4Copernicus will build the Copernicus Knowledge Graph and extend it via the project's Virtual Semantic Catalogue, that will also incorporate the YAGO2geo Knowledge Graph as well as knowledge extracted from EO images using machine learning algorithms (as detailed in deliverable D3.1, Architecture, semantics and discovery report). Furthermore, datasets will be described using the AI4EU Conceptual Schema<sup>4</sup> and published in the AI4EU resource catalogue, making them easily findable via the AI4EU platform<sup>5</sup>.

#### 3.2 Making data openly accessible

Several datasets that will be used as part of the project will be offered by third-party providers. Some of these datasets are already open to the public, while others are proprietary and have high commercial sensitivity. In the cases where private data are processed and aggregated (e.g., as part of a model, or functionality of a component) permission will be requested by the provider prior to making the altered data publicly available.

In reference to the nature of user data and private datasets brought in the project's Open Calls, some of the results that will be generated by each project phase will be restricted to authorised users, while other results will be publicly available, in collaboration with the actors involved in each project and in accordance with the rules set by the Open Calls. As per our Ethics commitment during the negotiation phase of the project, data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU.

Depending on their nature and size, datasets characterised as openly accessible will be published in the Zenodo repository, stored in DIASes and/or published in the AI4EU resource catalogue.

Furthermore, computational components developed in the context of the project will be containerized in the Docker configuration defined by the AI4EU Experiments<sup>6</sup> service and be made available via the AI4EU Experiments marketplace. We will also examine the benefits of publishing non-configured versions in Docker Hub.

When possible and desirable by the providing party (i.e., Open Call participants, the source code of AI4Copernicus resources and components will be openly available under appropriate licenses, to be selected per individual case. Regarding components developed by AI4Copernicus members, publishable source code will be accessible via a central repository hosted in popular relevant services like GitHub or GitLab, with the

---

<sup>4</sup> <https://github.com/ai4eu/ai-resources-ontology>

<sup>5</sup> <https://www.ai4europe.eu>

<sup>6</sup> <https://acumos-int-fhg.ai4eu.eu/#/home>

possibility that the consortium uses repositories accessible only to consortium members for development and testing purposes.

As the project evolves the DMP will incorporate details on the services and repositories used for distributing, documenting and supporting the source code.

### 3.3 Making data interoperable

As AI4Copernicus will be making use of variegating data sources along with the data generated by the project itself, interoperability is a major challenge to be tackled for making the provided components and datasets valuable and usable.

The challenge will be met primarily via the extensive usage of semantic web technologies and the incorporation of standards and widely used schemas, ontologies and vocabularies for describing and representing such data assets. Furthermore, data will be made available via standard-based APIs and brokers where applicable. In the same fashion, external data sources will be ingested via their APIs when such services are present.

Relevant documentation for the way data is represented will be included in all cases and will be made available following the same approach as the project's source code, i.e., via publicly accessible repositories.

### 3.4 Increase data re-use (through clarifying licenses)

If possible, the data set will be licensed under an Open Access license. However, this will depend on the level of privacy, and the Intellectual Property Right (IPR) involved in the data primarily by project partners and Open Call project implementors. A period of embargo will only be necessary if a data set contains specific IPR or other exploitable results that will justify an embargo. Therefore, the data will be licensed to permit the widest reuse possible when no limitations are identified by the key stakeholders.

The intention of AI4Copernicus is to make as much data as possible reusable for third parties. Restriction will only apply when privacy, IPR, or other exploitation grounds are in play. All data sets will be cleared of bad records, with clear naming conventions, and with appropriate metadata.

All data generated and collected in AI4Copernicus will undergo a quality check in order to analyse its individual plausibility and consistency, making sure that others can directly use it to perform assessments and validate the results produced by the project.

## 4 Allocation of Resources

Data Management is a core overarching activity for AI4Copernicus, as data access, processing and analysis inform all its technical activities. Per the project's work plan data management activities are the main responsibility of task 2.1, led by TAS, where usage requirements and consequently data needs are defined. In any case, effort on Data Management will be also dedicated in the context of Work Packages 4, 5 and 6 where the actual data collection, processing and generation primarily takes place. Regarding additional costs emerging from the adopted approach for data and knowledge publishing, the following table provides a high-level summary.

**Table 4.1: Data Management activities**

Issue	Action
Costs for making non-patented data and code FAIR	<ul style="list-style-type: none"> <li>- Fees associated with the publication of scientific articles containing project's research data in "Gold" Open access journals. The cost sharing, in case of multiple authors, shall be decided among the authors on a case-by-case basis.</li> <li>- Project Website operation: to be determined</li> <li>- Data archiving at OPENAIRE: free of charge</li> <li>- Copyright licensing with Creative Commons: free of charge</li> </ul>
Partner responsibilities	<p>Every partner is responsible for the data they produce. Any fee incurred for Open Access through scientific publication of the data will be the responsibility of the data owner (authors) partner(s) in compliance with the CA, Article 8.4.2.1: During the Project and for a period of 1 year after the end of the Project, the dissemination of own Results by one or several Parties including but not restricted to publications and presentations, shall be governed by the procedure of Article 29.1 of the Grant Agreement.</p>
Long-term preservation	<p>Data preservation of at least 1 year after the project is required by the Grant Agreement (Article 31.3). The associated costs for dataset preparation for archiving will be covered by the project itself. Long-term preservation of code and open datasets will be provided, and associated costs covered by a selected disciplinary repository. Proprietary and sensitive datasets will remain the property of the owners.</p>

## 5 Data Security and Protection

### 5.1 Security scope in the context of AI4Copernicus

AI4Copernicus uses established platforms and execution spaces that already adopt state-of-the-art security, authentication and authorization mechanisms. Partners and third parties (winners of the Open Calls bids) will be given access to these resources via the already existing processes foreseen by the collaborating DIASes and the AI4EU services.

Additionally, regarding the Open Call proposal submission platform, the following measures will be put in place:

1. **Roles & Permissions:** For data security, the proposal submission platform will have an authorized user management subsystem. The subsystem will allow access at different levels and with different rights depending on the role of the user. This ensures access to sensitive data only to authorized users.
2. **Encryption:** Additionally, the file encryption subsystem will create a folder where the proposal files will be stored encrypted. An encryption algorithm will be used that will use a specific platform key. This way the files cannot be accessed off-platform without the corresponding decryption key.
3. **Backups:** Finally, the platform will have a mechanism for creating daily backups of data. Backups will be done in two levels. In the first level the platform will receive a backup of the data stored in the platform database. In the second level the platform will back up all the folders where all the proposal files will be encrypted.

### 5.2 Data protection and GDPR compliance

AI4Copernicus project will process personal data both on its own behalf and on behalf of its clients.

In this context, AI4Copernicus will pay particular attention to the protection of personal data, to maximize its utility in compliance with applicable legislation relating to the protection of personal data. AI4Copernicus project shall ensure that no data will be shared without ensuring legal compliance and calculating legal risks.

In particular, the AI4Copernicus project shall ensure that personal data are protected according to the GDPR (General Data Protection Regulation) and that the use cases and experiments developed adhere to the GDPR.

This paragraph contains mandatory requirements that AI4Copernicus shall respect and defines a set of principles that AI4Copernicus shall be compliant with.

#### 5.2.1 GDPR scope

AI4Copernicus provides a 360° Privacy by Design and Privacy by Default data processing environment, following the GDPR, relevant best practices and applicable digital ethics requirements, while empowering individuals with informed decisions and free participation on a collective process for the public interest and benefit. All delivered information, templates, as well as measures and processes are designed to provide the means towards direct interaction and to some extent data transactions with individuals. As such it has a horizontal impact.

#### 5.2.2 Data protection form

AI4Copernicus enters into personal data processing activities at a limited scale and to the absolutely necessary extent. Details are provided in response to the Ethics Requirements (see Annex A).

#### 5.2.3 GDPR procedure

A full scale GDPR process is developed. Appointing a DPO is the starting point and appropriate technical and organisational measures, and safeguards are deployed throughout the personal data lifecycle as met within

the project. In regard to collection and use of personal data for every given task of the project, we address the following:

1. The personal data AI4Copernicus collects and the purpose(s) for which it collects that information;
2. How AI4Copernicus uses personal data internally;
3. Whether AI4Copernicus shares personal data with external entities, the categories of those entities, and the purposes for such sharing.

Details are provided in response to the Ethics Requirements (see Annex A).

Draft Version

## 6 Conclusions

The present report describes the data management lifecycle for the data to be collected, processed and generated in the context of AI4Copernicus. As part of the pursue to make research data findable, accessible, interoperable and re-usable (FAIR), it focuses 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; how data will be curated & preserved (including after the end of the project).

The current deliverable is the initial version of the project's DMP, which will be treated as a live document from hereon; as the project evolves, the report will be revised and updated providing further details and - if needed - amendments and corrections to make sure that all data-related aspects of AI4Copernicus are covered thoroughly and in accordance with the guiding principles of FAIRness, security and privacy.

Draft Version



## Annex A: Ethics Requirements

### A.1. Ethics Scope in the context of AI4Copernicus

Following the scope of the European Strategy for Data COM (2020), the post GDPR recently introduced set of proposals for regulation, namely the proposed regulation on Data Governance and the proposed regulation on AI, plus the proposals for the Digital Services Act and the Digital Markets Act provides proof of concept for the European Digital Strategy objectives. The implementation of this strategy at any given use case, especially in the research domain, requires a balancing test between the protection of personal data and the free movement of such data. Thus, a set of rules and means and a set of structures and processes to share data is imperative. The present document provides the ethical blueprint of AI4Copernicus by outlining a wider ethical framework that aims to set the necessary means, rules and structures by addressing. AI4Copernicus aims at building on a specialised, unique set of terms, a set of basic principles and case-specific (human) rights that do indeed address all post grant ethics requirements.

### A.2. Post-Grant Ethics Requirements

#### **Requirement 2.2. A description of the informed consent procedures that will be implemented for the participation of humans and sharing their personal data with third parties must be submitted as a deliverable**

Individual engagement in terms of personal data provision is limited in the context of WP6 and the AI4Copernicus Open Calls. The methodology underlying the collection and processing of personal data operates on a dual ground. On one hand the participation in the Open Calls by definition requires the provision of specific types of personal data like name, occupation, affiliation, contact details. These data types are necessary for the performance of the Open Calls. A special note on the 'data subjects' who by definition will mostly be constituted by legal entities thus outside the scope of the GDPR. However, all necessary GDPR safeguards are in place regardless of the volume of the inbound personal data streams.

Technically speaking, participation in the Open Calls equals to processing necessary for the performance of a contract which constitutes a legal basis for personal data processing according to GDPR Art.6.1(b). Open Calls are subject to specific annexes as served in the context of WP6. These annexes provide the terms under which a candidate may participate (the participation contract). On top of that FAQs do deliver very useful information on a number of issues including personal data processing. To the extent personal data is processed for the delivery of the Open Calls, no further legal basis is required whatsoever. In the event AI4Copernicus aims at further use and reuse for purposes outside the scope of WP6, a relevant consent form mechanism will be deployed.

Call for projects (use cases & small-scale experiments) process involves four distinct phases. Phase 4 (sustain) includes the delivery of relevant focus groups, where participants will be asked to provide their views on specific agenda items **solely on an anonymous format**. The content, language, format and overall architecture is presented in relevant templates herein. At this point we may say that all information is specific and provided in a plain language. The consent provision is explicit, and individuals are explicitly requested to do so 'in time', namely always prior to any processing activity thereof. Individuals may exclude themselves from any action at all times as things proceed.

Any other personal data processing is excluded as AI4Copernicus has no interest in the personal data per se, rather aggregated information, namely proposal thematics, demographics and related statistics directed for future audits.

**Requirement 2.3. Templates of the informed consent forms and information sheets (in language and terms intelligible to the participants) must be kept on file.**

When drafting consent forms the following elements need to be taken into account:

1. The means provided to individuals, as described in the consent script, need to be appropriate to make them understand the consequences of their authorization for the collection, use, dissemination and retention of their personal data.
2. Individuals need to be aware of and consent to all possible uses not initially described in the public privacy information notice that is in effect at the time of collection (when applicable).
3. The technical consent requirements as set by the GDPR are described as follows: *“any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data concerning him or her...”*.

As described AI4Copernicus will conduct focus groups and surveys, where solely anonymous data will be processed. Therefore, an informed consent template has been drafted based on the anonymous data script (see Appendix A(I)).

In regard to Open Calls, consent does not apply as a legal basis (see above, in response to Requirement 2.2.). However, the Open Call platform design and generated material provide proof of concept for up-to-date GDPR compliance and individual empowerment. A set of very straightforward FAQs provides in plain and intelligible language all personal data related issues. WP6 Annex 2 <Guidelines for applicants, delivered in plain language and comfortable layout, allows the candidate participants to navigate smoothly within diverse information, without embedding any dark patterns and misleading elements like language, ‘click design’, colour selection. Further information requested in WP6, Annexes 3, 4, 5, 6 are absolutely fundamental for the delivery of the Open Calls and will be kept on file following tested EU practices and procedures, throughout the AI4Copernicus, Open Calls personal data lifecycle.

**Requirement 11.1. A detailed description of the trust assurance mechanism and certification developed with regards to AI4Copernicus Trustmark must be submitted as a deliverable**

AI4Copernicus Trustmark will be awarded to the successful graduates of the project Open Calls process. The consortium will ensure a high level of transparency and will provide in detail the mechanisms for developing and granting the AI4Copernicus Trustmark.

This will be done as part of activities of WP6 where the consortium will provide: (a) an overview of the Trustmark process and the award criteria, which will be included in D6.2 (M28); (b) an overview of Trustmark granting results, which will be included in D6.5 (M36).

**Requirement 4.8. Description of the anonymisation/pseudonymisation techniques that will be implemented in relation to personal data collected in the project must be submitted as deliverable.**

Collected data, namely participant requested personal information solely in the context of the Open Calls, will be stored on password-protected databases to which only members of the appointed AI4Copernicus partners and personnel will have access on a need-to-know basis. The data will be used only within the project and will not be made accessible to any third party.

All other stored data will not include the names or addresses of participants, and identities of individuals will be known only by the research partners directly involved. Raw data, such as surveys, questionnaires, evaluations, will only be shared within the AI4Copernicus research group in conformity with all the rules set out in this document, EU legislation and national law. Unless informed consent has been obtained, information about participants will be edited for full anonymity before being processed, for example archived in file, used in project reports etc. Depending on the nature and utility of the personal data, appropriate techniques will be applied (data masking, pseudonymisation, perturbation) via the execution of relevant tools provided by AI4Copernicus partners. Off note, in terms of reporting for future audits at the levels of proposal a) submission, b) selection and c) evaluation solely aggregated data will be processed.

#### **Requirement 4.17. Description of the procedures that will ensure that GDPR will be respected by individual project members and as a Consortium**

AI4Copernicus partners do apply a high level of GDPR awareness and performance rooted to the internal policies, structures and safeguards in place. A project DPO has been appointed and specific guidance on the data quality requirements and AI4Copernicus practices has been circulated (see indicatively below Requirement 4.4.). At a higher level of abstraction, the general rules as agreed in the Consortium Agreement (art. 11.3) provide the following obligations on behalf of the partners:

*“In processing personal data pursuant to this Consortium Agreement, each Party shall:*

- *process, or permit to be processed, personal data only for the purposes of the performance of this Consortium Agreement. For avoidance of doubt each Party operates as Controller or Joint Controller to the extent it determines the purposes and means of a processing operation;*
- *ensure that its personnel are subject to an obligation of confidentiality in respect of the processing of personal data under this Consortium Agreement;*
- *ensure that appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data;*
- *not disclose or transfer personal data to any third-party other than where strictly necessary for the purposes of the performance of this Consortium Agreement;*
- *provide contact details of its DPO (if required to appoint one under applicable Data Protection Law) to the other Parties and the data subjects;*
- *notify the other Parties of any security incidents, events, weaknesses, data breaches or suspected data breaches impacting or capable of impacting the security of personal data processed under this Consortium Agreement;*
- *provide at no charge reasonable assistance to the other Parties to deal with any requests or complaints made by data subjects relating to the processing of their personal data in the exercise of their rights under applicable Data Protection Law.*
- *shall be individually responsible for its own processing of personal data pursuant to and/or in connection with this Consortium Agreement*
- *Upon termination or expiry of this Consortium Agreement, each Party serving as processor shall either delete or return to the controller all personal data processed under this Consortium Agreement, unless Data Protection Law requires otherwise.”*

In regard to the Open Calls related information (including personal data), proposal reviewers are appointed under a mutual agreement by the AI4Copernicus and provided with specific guidance. All AI4Copernicus partners are bound by confidentiality obligations as set forth in the Consortium Agreement (section 10). External advisory board members are bound by separate NDA agreement (see appendix B).

**Requirement 4.4. The beneficiary must explain how all of the data they intend to process is relevant and limited to the purposes of the research project (in accordance with the 'data minimisation' principle).**

WP6 as the sole field of personal data processing provides an adequate level of compliance in terms of a) relevant information provision to individuals b) data quality principles relating to personal data processing in accordance with GDPR article 5, and c) security measures and technical safeguards. The compliance items are analysed as follows:

- **information provision to individuals:** a full set of intelligible information in various levels of abstractions is provided via the annexes of WP6.
- **data quality:** all quality principles listed in GDPR article 5 are met.
  - **lawfulness, fairness and transparency:** all personal data processing activities are conducted under a solid legal basis (contract or consent);
  - **purpose limitation:** the processing purposes are explicitly identified a) for the execution and delivery of the Open Calls and b) for the performance of the focus groups in the context of phase 4 (sustain) of the Open Call for projects (use cases & small-scale experiments);
  - **data minimisation and accuracy:** the absolutely necessary personal data are obtained for all set purposes; the data is provided directly from the individuals;
  - **storage limitation:** personal data are kept solely for the delivery of the purpose of processing which is explicitly defined;
  - **integrity and confidentiality:** AI4Copernicus designs an Open Call process that aims to become the blueprint of similar endeavours. All partners are bound by confidentiality agreements and provide proof of concept in regard to security while maintaining no record of breach whatsoever;
  - **accountability:** roles and liabilities are explicitly defined in all operational documents (including the Consortium Agreement);

In the event during EO data analysis, personal data are directly or indirectly exposed, AI4Copernicus is committed to refrain from any kind of related processing and all downstream activities shall be excluded.

In the event selected third parties as approved in the context of the Open Calls seek to combine EO data with personal data or infer personal data information from EO data analysis, special guidance shall be provided based on the adopted GDPR strategy and overall ethics framework. As the case may be, selected third parties will be bound to specific GDPR obligations stated in the sub-grant agreement (see WP6 - Annex 7) to be signed between them and AI4Copernicus coordinator.

**Requirement 4.7. A description of the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing must be submitted as a deliverable**

All confidential information including the personal data at hand enter the AI4Copernicus ecosystem via the Open Calls platform managed by the AI4Copernicus coordinator subject to the following security controls:

- Appointment of a Chief Information Security Officer (CISO);
- Access on a need-to-know basis;
- Non-Disclosure Agreement (NDA) when necessary, in the event access is provided to external parties;

**Requirement 4.17. The templates of corresponding terms and conditions for 3rd parties regarding the use of personal data and incidental finding policy must be submitted as a deliverable.**

No personal data will be shared or made public. The sole personal information in public view shall be the names, affiliations and the proposal of the selected third parties, a processing activity in full alignment with the scope of the Open Calls and the participants' pursuits. No terms and conditions for 3rd parties regarding the use of personal data apply as such a component is irrelevant to the Open Calls designed process as in detail described in the context of WP6. The AI4Copernicus personal data lifecycle begins at the point of the proposal submission and ends a) at the selection phase for the non-selected participants and b) by the completion of Tasks 6.2., 6.3. and 6.4 (WP6) respectively for the selected participants. A relevant database, subject to all appropriate safeguards and purpose specific informed consent (see Appendix A(II)), shall be developed towards the formulation of an AI4Copernicus network community. No secondary uses are in the pipeline whatsoever.

Draft Version

## Appendix A(I)

### Consent to Participate in Research, Focus Group, Survey

#### Anonymous

Date: \_\_\_\_\_

Place: \_\_\_\_\_

**Full title of Project: AI4Copernicus.**

**A European funded (Horizon 2020) research project (Grant Agreement n° \_\_\_\_\_).**

[Research Activity] Title: \_\_\_\_\_

[Research Activity] Organiser: [Name, affiliation, position and contact info].

For any query regarding your personal data, your rights (access, information, edit, deletion, rectification, portability and further privacy issues you may contact with [\_\_\_\_\_, Event Owner Ethics Officer, @\_\_\_\_\_] at all times

Things I need to know

1. Purpose and objectives of research:

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_
4. \_\_\_\_\_
5. \_\_\_\_\_

2. Voluntary Participation: I understand that my participation is voluntary. I do not have to take part if I do not want to. I may leave the group at any time for any reason. I may also and/or revoke my consent at any time for any reason.

3. Risks: By participating in this research I may reveal opinions that may lead to stigmatization and or discrimination, if identified by any third party outside AI4Copernicus or combined with any third party information in a sophisticated way

4. Benefits: There are no direct benefits for participating in this research activity.

5. Third parties: \_\_\_\_\_

6. Storage: [Event Owner] shall store my anonymous data for [retention period,], on [storage asset,]

I hereby confirm that:

[1] I have read this consent form and understood the aims of the research activity.

Yes  No.

[2] I agree and provide specific and informed consent to participate in the research conducted by \_\_\_\_\_, who is a member of the AI4Copernicus research team.

Yes  No

[3] I consent to the [appropriated per research action, where consent is needed (e.g. audiovisual recording, photos)]

Yes  No

[4] I agree and provide specific and informed consent to share my anonymous data and/or content via [specify communication channel]

Yes  No

Name and email of the interviewee	Signature:
Interviewer:	Signature:

## Appendix A(II)

### Consent to subscribe to the AI4Copernicus Community

Date: \_\_\_\_\_

Place: \_\_\_\_\_

**Full title of Project: AI4Copernicus**

**A European funded (Horizon 2020) research project (Grant Agreement n° \_\_\_\_\_).**

Coordinator: [Name, affiliation, position and contact info].

For any query regarding your personal data, your rights (access, information, edit, deletion, rectification, portability and further privacy issues you may contact with [\_\_\_\_\_, Ethics Officer, @\_\_\_\_\_] at all times

Things I need to know

1. Purpose and objectives:

1. To develop an AI4Copernicus driven ecosystem;
2. To generate networking and collaboration channels within the AI4Copernicus ecosystem;
3. To actively participate in the AI4Copernicus ecosystem;
4. To inform you about future initiatives (Open Calls, workshops etc);
5. To inform you about new tools, technologies and datasets related to the EO domain;

2. By becoming member of the AI4Copernicus network and ecosystem I may be asked to provide: a) name; b) email address; c) institution; d) occupation/title e) technology interests and domains of expertise. AI4Copernicus collects, stores and processes this data to the extent necessary and the sole purpose to better evaluate Ai4Copernicus community and for communication reasons.

3. Voluntary Participation: I understand that my participation is voluntary. I do not have to take part if I do not want to. I may opt out at all times with no further implications.

4. Risks: By participating in this community I may reveal opinions that may lead to stigmatization and or discrimination, if identified by any third party outside AI4Copernicus or combined with any third party information in a sophisticated way.

5. Benefits: Active participation in a vibrant community of peers.

6. Third parties: Solely the members of the AI4Copernicus network and ecosystem.

7. Storage: [Event Owner] shall store my personal data for [retention period,], on [storage asset,]



I hereby confirm that:

[1] I have read this consent form and understood the aims of the present document.

Yes  No.

[2] I agree and provide specific and informed consent to become a member of the AI4Copernicus network and ecosystem.

Yes  No

[3] I agree and provide specific and informed consent to share my personal data solely within the AI4Copernicus network and ecosystem via [specify communication channel]

Yes  No

Name and email	Signature:
----------------	------------

## Appendix B

### Advisory Board Experts

### Non Disclosure Agreement

This Non-Disclosure Agreement (*the "NDA"*) in \_\_\_\_\_, (hereafter the "Effective Date") is entered into by and between

1. NCSR (National Centre for Scientific Research) Demokritos (hereinafter referred to as "NCSR-D"), with registered office in Agia Paraskevi, at Patr. Gregoriou E & 27 Neapoleos Str, 15341, hereby legally represented by Mr. Georgios Nounesis in his capacity as NCSR-D Director and Chairman of the Board and
2. Mr/Mrs \_\_\_\_\_, with her principal address in \_\_\_\_\_, \_\_\_\_\_ with ID \_\_\_\_\_, Tax Number \_\_\_\_\_ (the "External Advisor").

Hereinafter collectively referred to as "Parties"

Whereas:

- AI4Copernicus is an H2020 Innovation Action under grant agreement No 101016798 (the "AI4Copernicus").
- AI4Copernicus is consisted by: a) NCSR-D, b) Ethniko Kapodistriako Panepistimio Athinon (UoA), c) Thales Alenia Space France SAS (TAS), d) Institute European d'Administration des Affaires (INSEAD), e) Thales Six France SAS (THA), f) European Centre for Medium Range Weather Forecasts (ECMWF), g) Cloudferry SP ZOO (CF), h) Universita Degli Studi di Trento (UNITN), k) European Union Satellite Center (SatCen), l) Equinor ASA (EQUINOR), m) Blue Sight Conseil (BLUE SIGHT) (the "Consortium Partners")
- NCSR-D is the Project Coordinator of AI4Copernicus acting as the intermediary between the Action Partners and the External Advisor.
- Part of AI4Copernicus involves technology transfer via the AI4Copernicus Open Calls for projects and for citizen driven theme/social cause selection ("the Open Calls")
- The External Advisor is an expert on \_\_\_\_\_ [AN1], a high impact application area of earth observation and AI.
- The Parties have entered into a consultation service agreement in order to specify the terms under which the External Advisor assumes a consultation role to the Project Board of AI4Copernicus focusing on the Open Calls (the "Agreement").

- The Parties have agreed that Consortium Partners may disclose Confidential Information to the External Advisor in relation to AI4Copernicus and/or the Open Calls under the conditions of confidentiality set out in the Agreement.

## 1. Definitions

1.1. Confidential Information. For purposes of this Agreement, “Confidential Information” shall include all information concerning AI4Copernicus and/or the Open Calls and/or a Consortium Partner, whether commercial, financial, technical or otherwise that is disclosed to the External Advisor (whether disclosed orally, in documentary form, in electronic form, by demonstration or otherwise), and that: a) is designated as confidential either expressly or by necessary implication b) should be considered confidential given its nature or the circumstances surrounding its disclosure;

1.2. Disclosing Party means the Consortium Partner disclosing Confidential Information under this Agreement;

1.3. “Materials” means all material containing Confidential Information furnished by or obtained from the Open Calls and/or any Disclosing Party, including without limitation the Open Calls proposals as submitted, selection and evaluations thereof, Open Call project support and sustainability phases, use cases, experiment projects, ideas, know-how, trade secrets, documents, manuals, specifications, flowcharts, data (including personal data) and file storage media of any kind;

2. Exclusions from Confidential Information. External Advisor’s obligations under this Agreement do not extend to information that is: (a) publicly known at the time of disclosure or subsequently becomes publicly known through no fault of the External Advisor; (b) discovered or created by the External Advisor before disclosure by the Disclosing Party; (c) learned by the External Advisor through legitimate means other than from the Disclosing Party or Disclosing Party’s representatives; or (d) is disclosed by External Advisor with Disclosing Party prior written approval.

3. Obligations of the External Advisor. External Advisor shall hold and maintain the Confidential Information and any Materials thereof in strictest confidence. External Advisor shall not, without the prior written approval of the Disclosing Party, use for External Advisor’s own benefit, publish, copy, or otherwise disclose to others, or permit the use by others for their benefit or to the detriment of the Disclosing Party and/or the AI4Copernicus and/or the Open Calls candidate and/or selected applicants, any Confidential Information. External Advisor shall return to the Disclosing Party any and all records, notes, and other written, printed, or tangible Materials in its possession pertaining to Confidential Information immediately if the Disclosing Party requests it in writing.

4. Time Periods. The non disclosure provisions of this NDA shall survive the termination of both this NDA and of the Agreement and External Advisor’s duty to hold Confidential Information in confidence shall remain in effect for the duration of the AI4Copernicus as specified in the grant agreement No 101016798.

5. Relationships. Nothing contained in this Agreement shall be deemed to constitute either party a partner, joint venture or employee of the other party for any purpose.

6. Severability. If a court finds any provision of this NDA invalid or unenforceable, the remainder of this NDA shall be interpreted so as to best to affect the intent of the parties.

7. Integration. This NDA expresses the complete understanding of the parties with respect to the subject matter and supersedes all prior proposals, agreements, representations, and understandings. This NDA may not be amended except in a writing signed by both parties.

8. Waiver. The failure to exercise any right provided in this NDA shall not be a waiver of prior or subsequent rights.

9. Applicable Law and Jurisdiction: This NDA and any dispute or claim arising out of or in connection with or its subject matter or formation shall be governed by and construed in accordance with the laws of Belgium. The Parties irrevocably agree that the courts of Brussels shall have exclusive jurisdiction to hear or decide any claim, action or proceedings and to settle any disputes arising out or in connection with this NDA or its subject matter or formation.

This NDA and each party's obligations shall be binding on the representatives, assigns and successors of such party. Each party has signed this NDA through its authorized representative.

**NCSR-DEMOKRITOS**

Signature \_\_\_\_\_

Typed or Printed Name \_\_\_\_\_ Title \_\_\_\_\_

Date: \_\_\_\_\_

**EXTERNAL ADVISOR**

Signature \_\_\_\_\_

Typed or Printed Name \_\_\_\_\_ Title \_\_\_\_\_

Date: \_\_\_\_\_