

Project Title AI-augmented ecosystem for Earth Observation data accessibility with Extended reality User Interfaces for Service and data exploitation

Project Acronym EO4EU

Grant Agreement No. 101060784

Start Date of Project 1.06.2022

Duration of Project 36 Months

Project Website <https://www.eo4eu.eu/>

D1.3 – Data Management Plan

Work Package	WP 1, Project Management & Technical Coordination
Lead Author (Org)	Dimitris Tsakalidis (NVCR)
Contributing Author(s) (Org)	Georgios Domalis (NVCR), Aris Gioutlakis (NVCR)
Due Date	30.11.2022
Date	28.11.2022
Version	V1.0

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Author	Notes
0.1	06.09.2022	Georgios Domalis (NVCR)	Initial version and TOC
0.1	20.09.2022	Dimitris Tsakalidis (NVCR)	Formatted the structure based on the DMP template
0.2	28.10.2022	Dimitris Tsakalidis (NVCR)	Addressed the questions of the DMP template
0.3	01.11.2022	Aris Gioutlakis (NVCR)	Review
0.4	09.11.2022	Kaklis Dimitris (DANAOS)	Consortium reviewer
0.5	14.11.2022	Vasileios Baousis (ECMWF)	Consortium reviewer
0.6	25.11.2022	Maria Luisa Quarta (MEEU), Stefano Natali (SISTEMA)	Provided input on the expected data size and made suggestions
0.7	26.11.2022	Dimitris Tsakalidis (NVCR)	Addressed the comments of the consortium reviewers
1.0	28.11.2022	Sarantis Paskalis (NKUA)	Final version

Disclaimer

This document contains information which is proprietary to the EO4EU Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to a third party, in whole or parts, except with the prior consent of the EO4EU Consortium.

The content of the publication herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

Table of Contents

D1.3 – Data Management Plan	1
Disclaimer	3
Table of Contents	4
List of Figures	5
List of Tables	6
Executive Summary	8
1 Introduction	9
2 Data Summary	10
2.1.1 Definitions	10
2.1.2 Approach	10
2.1.3 List of Datasets	12
2.1.4 Data types	13
2.1.4.1 Data formats	15
2.1.4.2 Data size	20
3 FAIR data	21
3.1.1 Making data findable, including provisions for metadata	21
3.1.2 Making data accessible	21
3.1.3 Making data interoperable	22
3.1.4 Increase data re-use (through clarifying licenses)	23
4 Other research outputs	24
5 Allocation of resources	25
5.1.1 Responsibilities for data management in EO4EU	25
5.1.2 Costs for making data FAIR	25
6 Data security	26
6.1.1 Archiving and preservation	26
7 Ethics	28
8 Other issues	29
9 Conclusion	30
10 References	31
11 Annex I - List of datasets	32
12 Annex II - Informed Consent	34

List of Figures

Figure 1 – SAFE data specification.....	17
Figure 2 – Sentinel 2 product physical format.....	17
Figure 3 – SENTINEL 3 XFDU package schema.....	17

List of Tables

Table 1 – Target groups.....	11
Table 2 – Data formats.....	15
Table 3 – Data size per type of product.....	20
Table 4 – Data size per type of data	20
Table 5 – Data size per data source	20
Table 6 – List of datasets	32

Terminology

Terminology/Acronym	Description
AI	Artificial Intelligence
API	Application Programming Interfaces
BOA	Bottom-Of-Atmosphere
CA	Consortium Agreement
DataOps	Data Operations
DMP	Data Management Plan
DoA	Description of Action
EAB	External Advisory Board
EC	European Commission
EO	Earth Observation
EO4EU	AI-augmented ecosystem for Earth Observation data accessibility with Extended reality User Interfaces for Service and data exploitation
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
HPC	High-Performance Computing
HTTPS	Hypertext Transfer Protocol Secure
ML	Machine Learning
NDA	Non-Disclosure Agreement
TFVC	Team Foundation Version Control

Executive Summary

This deliverable presents the initial data management action plan (DMP) and provides the general outline of the project policy for data management. The EO4EU DMP identifies the data types that will be used, collected and/or generated, and frames the overall guidelines regarding data collected and generated throughout the project implementation.

The DMP will describe the format and support the data management life cycle of all the data collected and generated following the FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles, as defined in the “Guidelines to the Rules on Open Access to Scientific Publications [1] and Open Access to Research Data in Horizon 2020 [2]” and in the “Guidelines on FAIR Data Management in Horizon 2020 [3]”. This document will present the methodology and standards to be followed, in cases where data will be shared and/or made open, and how it will be curated and stored.

The strategy for the data management identifies and classifies respectively the data generated and collected, along with their metadata to be used. Also, this document reports on the exploitation and availability of the aforementioned data/metadata, the required ethical and legal compliance issues, and the responsibilities in the implementation of the DMP.

The DMP is a living document that will be updated when important changes to the project occur, due to the inclusion of new data sets, changes in consortium policies or any other external factors. At least two (2) updated versions are expected during the project lifetime; one in month 18 and the final one in month 36.

The next version of the EO4EU DMP will emphasize on the definition of procedures to be implemented by the 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 that data have to be preserved and preservation preparation time).

1 Introduction

The EO4EU project aims to provide innovative tools, services, methodologies and approaches utilizing modern technologies to support the wider exploitation of Earth Observation (EO) data. The project will assist a wide spectrum of users, from domain experts to simple citizens unaware of the plethora of data and capabilities offered by EU services, to access and process data and utilize the existing and future offered services.

To further enhance the proposed approach, the project will utilize existing background technologies and will capitalize on available data sources and data exploitation initiatives, such as Copernicus, GEOSS, INSPIRE, DestineE, Galileo/ EGNOS. By these means, we aim to support the EU Digital Single Market strategy's "Access" pillar by lowering the barrier for the development and deployment of advance data-driven services in EO, contribute to the Europe 2020: creating smart growth through open source and interoperable services, and blend a socio-economical, and data driven framework facilitating Sustainable Development.

Our mission is to contribute to the creation of an extensive collaboration hub by creating synergies with R&I funded projects and the community of EO data that includes EO data providers, the EOSC ecosystem, the Destination Earth initiative, multipliers, third-party organizations and environmental bodies and initiatives.

This document is the EO4EU data management plan that describes the data management life cycle for all datasets to be collected, processed and generated in the project. It constitutes the first version of the DMP document of the EO4EU project and provides the baseline of the policy that will be followed by the EO4EU consortium with respect to the data management related activities. More specifically, it covers the following activities:

- What types of data will be collected and/or generated?
- What standards will be used?
- How will this data be exploited, shared, processed and made accessible?
- How will this data be curated, stored and preserved?
- Which tools and methodologies will be used to store this data and for how long?
- How are data restriction levels managed?

This DMP outlines how research data will be handled throughout the life cycle of the project, as well as after its completion, contributing to the following issues:

- Ensuring that project research data and records are accurate, complete, authentic, interoperable and reliable.
- Saving time and resources in the long run.
- Enhancing data security and thereby minimizing the risk of data loss.
- Ensuring research integrity and reproducibility by others.
- Preventing duplication of effort by enabling others to use the EO4EU project's data.

2 Data Summary

2.1.1 Definitions

Open Access Data: Open access refers to unrestricted access to research results. Commonly, the open access characterization is given to open-source peer-reviewed publications, datasets, tools and source code. Open access focuses on building a community and enables scientists, researchers, individuals and enthusiasts to:

- Build and enhance existing research results
- Avoid redundancy
- Participate in Open Innovation activities
- Get in contact with high-ranking results of the EO4EU project

Open Research Data: Open research data refers to the disclosure of the linked research data which are needed to assess, validate and replicate the results presented in research publications. Complementary to the concept of open access, open research data enables the online availability of data resources towards promoting research.

The open research data concept focuses on enabling researchers and individuals to:

- understand, assess, reconstruct and further expand scientific publications
- build innovative concepts on top of existing research data
- establish a continuous improvement mechanism of research

2.1.2 Approach

The general strategy for data management, according to the Guidelines on Data Management in Horizon 2020 [3] will be based on the identification and classification of data generated and collected, standards and metadata to be used, exploitation and availability of data, data sharing and archiving, the preservation of the information, as well as on the ethical and legal compliance and the responsibilities in the implementation of the DMP.

Interpreting Environmental Observation data is of great significance on assessing climate change towards meeting sustainability criteria and supporting the Green Deal and the climate and ecological transition. EO4EU brings together a strong partnership that will focus on digesting EO information from different sources, ensuring multi-actor EO data accessibility across the EU and sectors and interoperability of systems while it aspires to establish strong links with the Copernicus Programme, the European Space Agency's (ESA) Earth observation programme, the Global Earth Observation System of Systems (GEOSS) and DestinE programme, promoting a unified data interoperable ecosystem. EO4EU will promote pre-operational European services like DestinE and will utilize existing platforms and services in a consolidated manner through the extensive use of disruptive technologies. The large amounts of data produced by DestinE naturally contribute to the European Green Deal data space, facilitating data flows inside the EU and across sectors, for the common good. Existing European programs like Copernicus, Galileo, EGNOS, INSPIRE already provide a significant amount of invaluable EO data currently being used from many organizations and SMEs to deliver their value-added services.

EO4EU will demonstrate its operational and technical capacity on seven (7) distinct pilots that cover the thematic areas of (i) personalized health care, (ii) sea route planning, (iii) ocean monitoring, (iv) food security, (v) food ecosystems, (vi) soil erosion, (vii) environmental pest, and (viii) crisis (responders) management.

These thematic areas will engage a wide spectrum of involved stakeholders, presented in the Table 1, to join efforts and provide their multidisciplinary expertise to support the Commission’s strategic goals towards the further exploitation of EO data. To further enhance the proposed approach, the project will utilize existing background technologies and will capitalize on available data sources and data exploitation initiatives, including Copernicus, GEOSS, INSPIRE, DestineE, Galileo/EGNOS.

Table 1 – Target groups

Target Groups	EO4EU contribution
Research and Academia	Support research requiring more accessible EO data
Policy makers and actors	Support decision-making and climate action through new user-friendly tools and enhanced EO capabilities
EO data providers	Promote further usage of their data through value added tools and services
Private sector	Encouraging new innovations through more accessible EO data and tools for non-technical users
Citizen scientists and general public	Enable citizen scientists to use more accessible EO tools and show how EU investments reduce impacts of climate change and disasters
Standards development organizations	Recommend revisions of existing standards or contribution of new standards
Civil society, NGOs, people’s associations	Civil society organizations can use the EO4EU tools to efficiently manage and empower their communities, by better monitoring, representing, storing and reusing earth related information.

EO4EU vision stands on enhancing the FAIRness of environmental observation data through an open, transparent, and ease to access DataOps Center. More specifically, EO4EU follows an ML-augmented dynamic annotation methodology that labels existing datasets. By these means, we can increase their searchability and discoverability by end users, transforming them into open and accessible knowledge sources while also facilitating their interpretability and re-usability. Labels and relevant metadata will be deposited within the knowledge graph providing an extended searchable capacity to various data resources. Raw and processed data will be accessible through dedicated Application Programming Interfaces (APIs), while a pool of visualization services will focus on advancing stakeholders’ knowledge capacity both at regional and local level, enabling an evidence-based decision support mechanism for decision/policy makers towards realizing the EU’s climate adaptation strategy. We aim to establish an open access data ecosystem that bridges the gap of the individual data “silos” that communicate and directly interact with numerous sources of EO data. Such an ecosystem targets on increasing the usability and exploitability of EO information, acting as a liaison between various initiatives and Data Pools. The proposed DataOps will act as a public-common EU data fetching mechanism for consolidating digital systems and services, focusing on facilitating stakeholders’ participation to global environmental observation data, actionable information and knowledge.

Each participating organization will examine whether open access can be granted without affecting any legal and ethical requirements, including the Intellectual Property Rights as per the dissemination access level of each dataset produced.

Such information consists of:

- A general data summary, including origin, types and formats of files, purpose, size, and utility;
- FAIR Data:
 - **Findable:** discoverability, naming convention, search keywords, version numbers, metadata;
 - **Accessible:** availability, software tools, repositories, restriction and/or conditions for access;
 - **Interoperable:** description of interoperability, metadata vocabularies, standards and/or methodologies;
 - **Re-Usable:** license, data quality, time frames for availability and storage.
- Allocation of resources (by Consortium agreement, each partner has to individually identify and allocate resources for data storage and management);
- Data security;
- Ethical aspects, if needed;
- Other issues.

Towards formulating an effective DMP, and ensuring to effectively keeping track of the varieties of data generated and/or collected by the project, it is of significant importance to categorize the datasets according to their form, source and type. For this reason, the EO4EU partnership classifies the numerous data collections according to their origins and purpose, as described in detail below. An initial breakdown of the project datasets in five (5) categories is as follows:

- **Dataset Category 1:** Consortium information
- **Dataset Category 2:** Project files
- **Dataset Category 3:** Research activities
- **Dataset Category 4:** Development data, packages and source code
- **Dataset Category 5:** Demonstration activities

2.1.3 List of Datasets

The data that will be collected, extracted and generated throughout the activities of the EO4EU project are classified in five (5) datasets, as follows:

Dataset Category 1 - Consortium Information (A): This dataset category includes personal and sensitive information of the consortium partners, such as emails and online accounts. This information will be treated with confidentiality and only for internal communication for the purposes of the project.

All the information will be stored in a private and secure storage area (Sharepoint [4]). Access will be restricted only to the members of the consortium and the External Advisory Board (EAB) members, which are also considered as members of the EO4EU partnership.

Dataset Category 2 - Project files (B): It includes all the gathered information from internal plenary, technical and review meetings as well as workshops. It will be securely stored within the private cloud area (Sharepoint) and will be only restricted to consortium members.

The project outcomes that do not contain any personal or sensitive information will be publicly disseminated through the relevant channels of the project (social media, website, etc.).

Confidential outcomes along with any linked files and documents will be encrypted and stored in the internal cloud area of the project (Sharepoint). The encrypted data will only be accessible by the data owners and specific consortium partners.

Dataset Category 3 - Research activities (C): It refers to the project's research related outputs such as deliverables, white papers, publications, survey data, etc. This dataset category will adopt an open access approach policy, being publicly accessible through open access platforms such as Zenodo [5]. State-of-the-art methods will be utilized, if and where necessary, to carry out the project's research and development activities. In specific research tasks that involve sensitive information (personal data, confidential information from public authorities, etc.), all relevant documentation, files and any other type of source will be stored and protected locally by the corresponding organization.

If any material related to the project needs to be shared, proper anonymization and encryption tools will ensure the integrity of personal and sensitive information. These procedures will comply with the ethical guidelines, as defined by the EU.

Dataset Category 4 - Development data, packages and source code (D): This dataset category concerns data generated throughout the life cycle of the project's implementation process, from research prototypes to development codes and deployment scripts. The aforementioned material will be stored and preserved in a private repository of version control tools, such as Team Foundation Version Control (TFVC) or Git repositories. Backup versions of each module will be extracted periodically and stored in the Sharepoint platform.

Dataset Category 5 - Demonstration activities (E): It includes the data generated in the pilot activities (WP5) of the project demonstrators including data generated and collected from stakeholder workshops and training (WP6). Other types of data, as well as users and participants that might be considered during the project lifetime, as set and updated by the user/ system requirements, will be included in the second version of this deliverable.

2.1.4 Data types

- **No personal data:** such as information which is not affected by Data Protection legislation.
- **Personal data:** Information that is clearly about a particular person such as email address, telephone number, passwords, etc.
- **Sensitive information:** A specific set of "special categories" information that must be treated with extra security. This includes information pertaining to:
 - Racial or ethnic origin;
 - Political opinions;
 - Religious or philosophical beliefs;
 - Trade union membership;
 - Genetic data; and
 - Biometric data (where processed to uniquely identify someone).

According to EU directives, personal/ sensitive information refers to the information that might lead to the identification of an individual, either:

- directly from this data, or
- indirectly, through information that is in the possession, or is likely to come into the possession, of the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

Since EO4EU follows the European directives, any personal data that might be included in the research and the implementation activities will be anonymised, according to the Community's guidelines, to ensure that no ethical or privacy issues with either the EU or any National legislation arise. Such activities are carried out as part of WP1 and WP5, namely demonstrators' activities and ethics requirements, respectively.

In the project's demonstration and pilot activities, along with the workshops and stakeholders' engagement activities planned to be carried out, each participant, including members of the EO4EU consortium, will be informed prior to the data collection process. In particular, to ensure the application of the ethics principles and avoid any ethics related implications, all participants will be invited to sign a consent form.

2.1.4.1 Data formats

The aforementioned identified datasets can include different data formats. A manually gathered dataset concerning impact assessment and user acceptance of the demonstrators can consist of interview notes, pictures from pilot sites, as well as questionnaires' responses. Both the collected and generated data will be delivered anonymous as open research data. EO4EU will use widely accepted data formats, as presented in Table 2 and onward.

Table 2 – Data formats

#	Name	Acronym	Extensions	Services
1	Standard Archive Format for Europe	SAFE	.safe	Copernicus
2	General Regularly distributed Information in Binary form	GRIB	.grib/.grb/.gb	ECMWF model, DestinE
3	Network Common Data Form	NetCDF	.nc	Copernicus
4	Geo Tag Image File Format	GeoTIFF	.tif	Copernicus
5	Shapefile	SHX	.shp/ .shx/ .dbf / .esri	Copernicus
6	Keyhole Markup Language	KML	.kml	Copernicus
7	Portable Document Format	PDF	.pdf	Copernicus
8	Text	TXT	.txt	EO4EU
9	Document	DOC/DOCX	.doc/.docx	EO4EU
10	JavaScript Object Notation	JSON	.json	Copernicus
11	Comma - Separated Values	CSV	.csv	EO4EU
12	Spreadsheets	XLS/XSLX	.xls/.xlsx	EO4EU
13	Extensible Markup Language	XML	.xml	Copernicus
14	Joint Photographic Experts Group	JPEG	.jpg/ .jpeg	EO4EU
15	Portable Network Graphics	PNG	.png	EO4EU

Standard Archive Format for Europe (.safe): SENTINEL data products are distributed using a SENTINEL-specific variation of the Standard Archive Format for Europe (SAFE) format specification. The SAFE format has been designed to act as a common format for archiving and conveying data within ESA Earth Observation archiving facilities. SAFE format wraps a folder containing image data in a binary data format and product metadata in XML and allows the format to be scalable enough to represent all levels of SENTINEL products. A SENTINEL product refers to a directory folder that contains a collection of information and includes:

- a 'manifest.safe' file which holds the general product information in XML
- subfolders for measurement datasets containing image data in various binary formats
- a preview folder containing 'quicklooks' in PNG format, Google Earth overlays in KML format and HTML preview files or GML-JPEG2000
- an annotation folder containing the product metadata in XML as well as calibration data
- a support folder containing the XML schemes describing the product XML.

SENTINEL 2 and LEVEL 2A: Prototype product is an orthorectified product providing Bottom-Of-Atmosphere (BOA) reflectances, and basic pixel classification (including classes for different types of cloud) [6]. Image data product uses the same tiling, encoding and filing structure as Level-1C [7].

SENTINEL 3 - OLCI: Data format follows the format defined for each Sentinel-3 product in the PDGS product specification and is based on SENTINEL-SAFE. SENTINEL-SAFE is a profile of the XML Formatted Data Units (XFDU), for specific utilization in the EO domain, providing semantics in the same domain to improve interoperability between ground segment facilities.

SENTINEL 3 - SLSTR: Data format follows the format defined for each Sentinel-3 product, i.e. the PDGS product specification, and is based on SENTINEL-SAFE and has Level-1 and 2 products [8].

SENTINEL 3 - SYNERGY: OLCI, in conjunction with the SLSTR instrument, provides the SYN products, providing continuity with SPOT VEGETATION. Syn products objective is the monitoring of land use and provide information relating to worldwide food security and contributes to the study of climate.

SENTINEL 3 - Altimetry: User Guide provides a high level description of the available instrument operating modes and products.

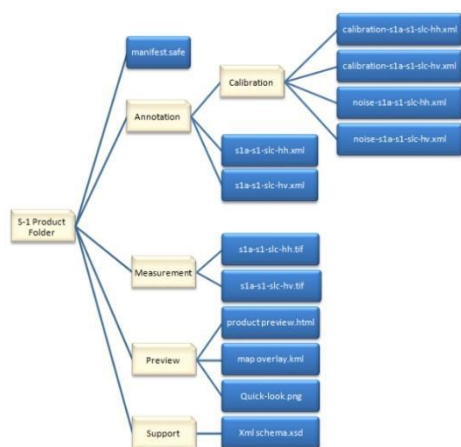


Figure 1: SAFE data specification

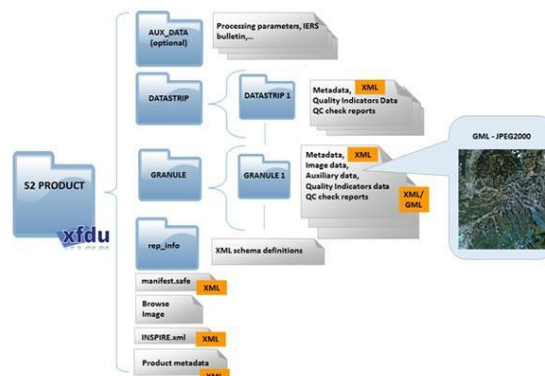


Figure 2: Sentinel 2 product physical format

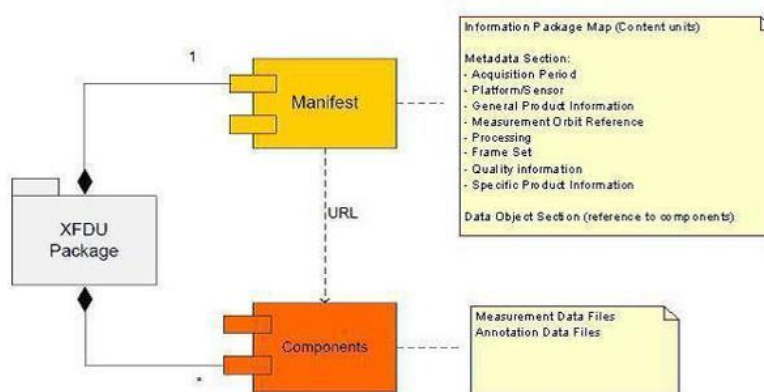


Figure 3: SENTINEL 3 XFDU package schema

GRIB (.grib, .grb or .gb): The GRIB file format is designed for storing and distributing weather data. GRIB files are widely used in meteorological applications. GRIB stands for "General Regularly distributed Information in Binary form" and is a WMO (World Meteorological Organisation) standard format for archiving and exchanging gridded data. GRIB is a binary format, and the data is packed to increase storage efficiency. GRIB messages are often concatenated together to form a GRIB file. GRIB files usually have the extension .grib, .grb or .gb.

Currently there are two different coding standards: GRIB edition 1 (commonly referred to as GRIB1) and GRIB edition 2 (GRIB2). The major differences are in the structure of the messages; in GRIB2, several variables are defined with more precision (e.g., in GRIB1, latitudes and longitudes are in milli-degrees while in GRIB2, they are in micro-degrees). Also, in GRIB2, longitude values must lie between 0 and 360 degrees, the encoding of the parameter is very different, and in GRIB2 the description of the data is template/table based. Note that a GRIB file can contain a mix of GRIB1 and GRIB2 messages.

The ECMWF model (the Integrated Forecasting System, IFS) currently outputs model-level fields in GRIB2 while pressure and surface level outputs are produced in GRIB1 [9]. For example, ERA-Interim (a climate reanalysis dataset provided by ECMWF) is produced in the GRIB edition 1 format. The ERA-

Interim data is then made available for download in its native GRIB format. In some cases, data is also available in NetCDF format as the result of the conversion of the GRIB file to NetCDF. Note that due to this conversion, not all the information in the GRIB file will be included in the NetCDF version, and this is particularly true for the GRIB file metadata. As a result, care should be taken when using these files. At this time, the NetCDF format is not formally supported by ECMWF. The DestinE first two Digital Twins will produce data in GRIB2 [10] format whereas the sentinel data will be in GeoTIFF

NetCDF (.nc): The NetCDF (Network Common Data Form) [11] is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. NetCDF is commonly used to store and distribute scientific data. The NetCDF software was developed at the Unidata Program Center in Boulder, Colorado, USA. NetCDF files usually have the extension .nc. To read NetCDF files there are tools with a graphical interface like Matlab, IDL, ArcGIS, NCVIEW, Xconv and developer (programming) tools like the Unidata NetCDF4 module for Python and Xarray.

For climate and forecast data stored in NetCDF format there are several (non-mandatory) metadata conventions which can be used (such as the CF Convention). CF compliant metadata in NetCDF files can be accessed by several tools, including Metview, NCVIEW, Xconv. The latest version of the NetCDF format is NetCDF 4 (aka NetCDF enhanced, introduced in 2008), but NetCDF 3 (NetCDF classic) is also still used. NetCDF files can be converted to ASCII or text.

GeoTIFF (.tif): It is a format extension for storing georeference and geocoding information in a TIFF 6.0 compliant raster file by tying a raster image to a known model space or map projection [12]. A GeoTIFF file is a TIFF 6.0 file, and inherits the file structure as described in the corresponding portion of the TIFF spec. The GeoTIFF format uses a defined set of TIFF tags to describe cartographic information associated with TIFF imagery that originates from satellite imaging systems, scanned aerial photography, scanned maps, digital elevation models, or as a result of geographic analyses. GeoTIFF can store a broad range of georeferencing information, catering to geographic as well as projected coordinate system's needs. Supported projections include UTM, US State Plane and National Grids, as well as the underlying projection types such as Transverse Mercator, Lambert Conformal Conic, etc. GeoTIFF uses a "MetaTag" (GeoKey) approach to encode dozens of information elements into just 6 private tags (33550, 34264, 33922, 34735, 34736, and 34737), taking advantage of TIFF platform-independent data format representation to avoid cross-platform interchange difficulties. GeoTIFF uses numerical codes to describe projection types, coordinate systems, datums, ellipsoids, etc.

Shapefile (.esri): The ESRI Shapefile stores non topological geometry and attributes information for spatial features in a data set. A Shapefile consists minimally of a main file, an index file, and a dBASE table. responding to the main file record from the beginning of the main file. Attributes are held in a dBASE format file. The shapefile format can support point, line, and area features. Area features are represented as closed loop, double-digitized polygons.

KML (.kml): KML is a file format used to display geographic data in a 2-D or 3-D earth browser, such as Google Earth, Google Maps. The markup specification was originally developed by Keyhole, Inc. The product became Google Earth. KML, version 2.2. was adopted as an international standard in 2008 by the Open Geospatial Consortium, Inc. (OGC). KML is an XML grammar used to encode and transport representations of geographic data for display in an earth browser (3D) or online map (2D). KML encodes what to show and how to show it, using a tag-based structure with nested elements and attributes. KML focuses on geographic visualization, including annotation of maps and images. KML files can pinpoint locations, add image overlays, and expose rich data through icons and captions. KML can support not only the presentation of data on the globe, but also control the user's navigation in

the sense of directing where to go and where to look. The vertical datum, used for altitude measurements is the WGS84 EGM96 Geoid.

Documents/Reports/Publications (.PDF/A, .txt, .doc/.docx): A TXT file is a standard text document that contains unformatted, plain text. It is recognized by any text editing or word processing program such as Microsoft Notepad or Apple TextEdit. TXT files are encoded in numerous non-proprietary encodings such as ASCII and UTF-8.

JavaScript Object Notation (.json): JavaScript Object Notation is a lightweight data interchange format. JSON files store simple data structures in an organized and easy-to-access approach. It is a collection of name/value pairs, while it is also completely language independent. JSON is built on two structures: (a) a collection of name/value pairs, (b) an ordered list of values. JSON is a text-based format accessed via any text editor, or command line editor (e.g., Vim) on other than Windows operating systems.

Comma - Separated Values (.csv): CSV text files use commas to separate values. Typically, they store tabular data (numbers and text) in plain text. They can be opened with any text editor, such as Notepad++ [8]. The CSV file format is not fully standardized, albeit there are some specific rules and recommendations. In cases where field data also contains commas or embedded line breaks, it is a common approach to use quotation marks to surround the field.

Spreadsheets (.xls/.xlsx): Spreadsheets is a program application that enables analysis and storage of data in tabular form. Spreadsheet is a file that consists of cells in rows and columns. In spreadsheets, data formats vary from numeric values to text, formulas, references and functions. Aside from storing tabular data and performing basic arithmetic and mathematical functions, spreadsheets such as Excel [9], WPS [10] provide built-in functions for statistical operations.

Extensible Markup Language (.xml): Markup Language defines a set of rules for encoding textual data such as documents. XML is widely acceptable and is used for the representation of arbitrary data structures. XML files are accessible with any web browser, editable via any text editor or website that offers a view, edit, and convert to other formats capacity. An XML file is encoded with the ASCII text standard.

Pictures (.jpg, .png): The abbreviation JPEG stands for “Joint Photographic Experts Group”. It is a universal- standard image format for compressed image data. JPEG files are accessed via an image viewer or image editor, such as Paint, Windows Photos, Apple Preview. Various web browsers such as Google Chrome, Firefox and Microsoft Edge can be utilized to access images. JPEG standard specifies the codec that defines how an image is compressed into a stream of bytes and decompressed back into an image.

2.1.4.2 Data size

The data assessment is performed on the first round of use cases requirements collection. At this state, most of the use cases have provided a detailed overview of the input data, while for few of them some details are still missing (e.g., duration / time frame). For the latter, a preliminary yearly data volume estimation is provided. A summary report on the currently available datasets is presented in Tables 3, 4 and 5, while a detailed report on the data in the use cases and the project in general, will be thoroughly presented in the next versions of the DMP.

Total number of datasets: 95

Total volume of input datasets: 1.266 TB

Table 3 – Data size per type of product

Type of products	n. datasets	size (GB)
forecast	23	21.00
re-analysis	19	74.97
analysis and forecast	11	402.49
observation	37	756.80
in-situ	0	-
ancillary	4	10.87
other	2	-
TOTAL	96	1,266.13

Table 4 – Data size per type of data

Type of data	n. datasets	size (GB)
model	50	498.14
satellite	14	523.60
in-situ	6	1.76
other	26	242.64
TOTAL	96	1,266.13

Table 5 – Data size per data source

Data source	n. datasets	size (GB)
ECMWF	5	11.16
Copernicus Atmosphere	12	8.93
Copernicus Climate	28	77.52
Copernicus Land	3	0.99
Copernicus Space	2	396.64
EUMETSAT	6	0.67
CMCC	6	400.00
ESA	1	117.00
Other	35	253.23
TOTAL	96	1,266.13

3 FAIR data

3.1.1 Making data findable, including provisions for metadata

Special emphasis is placed on enhancing the discoverability of the collected and generated data. EO4EU follows a knowledge graph-based approach to increase the searchability and the discoverability of the data, while also facilitating its understanding and re-use, ranking the output datasets and features based on the percentages of a metric (e.g., cosine), sorted in descending order. The extracted features of the data through the Machine Learning (ML) pipeline will be labelled and will act as metadata, while indexing methodologies will enable the discovery of the data according to the needs of the users. Binary content data, such as images, will be identified by a persistent identifier.

Metadata links information and data across the web and constitutes a powerful tool that helps individuals (researchers, developers, citizens, etc.) to discover, identify, and manage digital resources. Metadata refers to information about the data collected and/or generated. It is usually structured as textual information that describes the creation, content, or context of a digital resource. The most notably known types of metadata are names, dates, location, data types, relations and interdependencies to other data sets. In EO4EU, aside from the abovementioned, both features and labels will act as metadata towards enabling further discoverability of dataset and feature resources.

To further increase data discoverability, all data produced and categorized as open for sharing and re-use, will be accessible via open data repositories and EO4EU data storage units. Datasets that will be uploaded to open access repositories will be deposited in a searchable resource (the cloud web storage service of the project) and will be accessible via dedicated Application Programming Interfaces (APIs).

The naming conventions in the development of the project's data files can significantly increase their searchability. Towards this, EO4EU will design consistent data file names that properly describe their content, status and versioning, with a view on increasing their discoverability.

3.1.2 Making data accessible

FAIR open access to the data guide refers to making data accessible to all project partners, researchers and the public, following the privacy and anonymity guidelines of the EU and National regulations. Horizon 2020 Open Research Data Pilot states that all data generated and used, if possible, are publicly open and available. The EO4EU partnership will ensure the integrity of personal data and sensitive information prior to the dissemination of the datasets.

The project does not aim to replicate any data, but if needed, any extra information will be stored as temporary data to the internal infrastructure of CINECA or in the project partners' facilities. Since no data replication will take place in the project, there is no need for the establishment of a dedicated data access committee to evaluate or approve access requests to personal or sensitive data. Massive batches of EO datasets of historical and daily earth observation measurements along with open access data from national cohorts and pre-identified databases will be linked with EO4EU. Available resources will be interlinked with the project through Open APIs and will be publicly available.

The partners of EO4EU will utilize state-of-the-art methodologies towards ensuring the secure storage, delivery and access of all kinds of data and project related material, managing at the same time the clearance levels and access rights among the users.

During the execution of the project, each partner will provide detailed information on privacy/confidentiality and the procedures that will be implemented for data collection, storage,

access, sharing policies (especially when third party countries are concerned), protection, retention and destruction. The consortium will confirm that the project complies with national and EU legislation throughout its lifetime and after its completion.

Personal data, if any, will be treated confidentially and carefully, taking proper technical means of information protection (e.g., storing general and personal data separately, using encryption for personal data and identities, deleting personal data when it becomes unnecessary). Some examples include public-key encryption and symmetric encryption with session keys negotiation over HTTPS. Considering that some transmitted data may be regarded as highly sensitive, the highest security standards will be used (i.e., asymmetric cryptography with at least 1024-bit keys). Where necessary (e.g., sensitive and evidence information collected by public authorities and/or consortium members), data will be stored in a locked server, while all identification data will be stored separately. Moreover, no one outside the research team will have access to this information. The access to the database will be restricted to authorized personnel only, set by each consortium member. Moreover, an access log will be maintained to ensure the proper use of the accessed data.

For all data, especially those used in the Demonstrators (generated in WP2, WP5), the names of the participants are not pointing to real persons. All the direct and indirect (according to the EU legislation - GDPR) information of an individual will be fully masked in any printed materials, project reports or dissemination activities.

All identified dataset categories will be stored and backed up regularly through existing back-up mechanisms in place at Sharepoint. In particular, the open research datasets will be archived in the Zenodo and assigned a unique identifier (reference), to ensure open accessibility and discoverability. Qualitative data will be backed up and secured by the coordinator on a regular basis; metadata will include clear labelling of versions and dates. The data produced and generated in the context of the research and development activities of the project, from research prototypes to development codes and deployment scripts, will be stored and preserved in a private repository in version control tools such as GitLab and Sharepoint. Accordingly, backup versions of each module will be extracted periodically and stored in the Sharepoint platform. All data containing private information will be destroyed upon completion of the respective study/research task. In any case, all personal data will be destroyed at the end of the project and only anonymous or non-identifiable data will be retained after the completion of the final report.

3.1.3 Making data interoperable

Data interoperability refers to the ability of systems and services to access readable and editable data, in terms of their content, context and meaning. To achieve it, EO4EU will incorporate suitable standards and vocabularies for data and metadata creation.

The EO4EU vision of interconnecting various heterogeneous APIs/services, foresees the capitalization of High-Performance Computing (HPC) infrastructure, that ensures a continuous service of intercommunication and data exchange, supporting extreme data loads moving through the different "pipelines" of the architecture, providing also the capacity to control communications required to forward and deliver the processed/stacked data to each system and subsystem. A HPC infrastructure will be used for ML workflows and pipelines, and will also facilitate extreme data loads and data exchange among the integrated components of the solution as well as external sources, either to retrieve and/or send raw/processed data. A dedicated ML pipeline will utilize the available input data sources of EO data to process aiming to compress data towards optimizing performance on serving downstream tasks

The interoperability of the aforementioned publicly available data will be achieved, since the data and its respective metadata will be stored in JSON format according to a defined JSON schema. The JSON schema provides a collection of shared vocabularies that can be used to mark-up pages in ways that can be understood by the major search engines.

Common vocabulary and code lists of predefined values for harmonizing the descriptions of EO metadata and data are under definition and will be defined in the course of the project, using, when possible, the code lists and their values.

3.1.4 Increase data re-use (through clarifying licenses)

To ensure that interested third parties can re-use the EO4EU research datasets, all produced data will be released under the Creative Commons Licensing scheme. Licenses are the means that enable a third-party to copy, display, edit and build upon for the purposes set by a specific license.

Data availability after the end of the project depends highly on the type and content of data, taking into account sensitivity and anonymisation status. Data should be available for public reusability after being granted permission from their respective contributors, following the proposed legal and ethics requirements.

Rich metadata will enable proper discovery and identification of the data along with the appropriate licensing schemes facilitating their re-usability. In principle, it is expected that data will become available after the publication of the respective deliverables and will remain available after the completion of the project.

To safeguard the transparency, consistency, quality, completeness and accuracy of the data, EO4EU adopts a data quality assurance procedure. Peer-reviews of the data generation methods and/or data summaries will be applied to assess the quality of the dataset and identify any need for improvement.

4 Other research outputs

Other research data will be stored and backed up regularly through existing back-up mechanisms in place at Sharepoint. Qualitative data will be backed up and secured by the coordinator on a regular basis and metadata will include clear labeling of versions and dates.

5 Allocation of resources

5.1.1 Responsibilities for data management in EO4EU

The costs required for making the data collected/generated FAIR have been included in the budget of the project. The National Kapodistrian University of Athens (NKUA), as project coordinator and leader of WP1 and Novelcore (NVCR) as primarily responsible for the delivery of the data management strategy, have a particular responsibility in creating and updating the Data Management Plan. However, the datasets creators will be responsible for managing these individual datasets. In general, the EO4EU consortium as a whole will decide and contribute to relevant aspects of the data management cycle during and after the completion of this project. A specific table summarizing the research team leaders responsible for each dataset will be added in the future release of the DMP.

5.1.2 Costs for making data FAIR

The costs for making data FAIR and re-usable include the costs of the cloud infrastructure, maintenance, data storage, or charges for data archiving, conversion of proprietary data files into open formats and the costs of the personnel involved in collecting and managing data. At this state, the chosen repository (Zenodo), for long term deposit and preservation of searchable data intended for public use, does not apply fees for archiving and data curation. The next version of the DMP, will thoroughly describe the expected costs and the necessary actions needed to retain such information public, even after the completion of the project. Peer-reviewed publications costs related to open-access research data are eligible in Horizon Europe and will be covered by the EO4EU budget.

6 Data security

The EO4EU partnership pays a strong emphasis on ensuring the security of all the produced datasets, safeguarding them from unauthorized access and loss. To achieve so, dedicated technical and organization measures will be designed and applied, following the guidelines produced in the risk assessment plan formulated within *T1.3: Risk management & Quality Assurance*. More specifically, the project's data security plan will focus on minimizing the risks of a data breach during the implementation of the project, as well as after its completion. Both human, machine and hardware errors will be considered to cover a wide spectrum of potential failures.

All the information will be stored in a private and secure storage area, namely Sharepoint. The data will be backed up on a regular basis and access will be restricted only to the members of the consortium. The External Advisory Board (EAB) members, who have signed an NDA, are also considered as members of the EO4EU partnership, therefore full disclosure to the project files will be given to them.

To ensure that data is protected and secure, all project partners will incorporate the appropriate means in terms of both processing data as well as storing and preserving them, in private servers or cloud providers, according to the relevant legal data protection requirements (e.g., GDPR). In case of personal data collections, it is crucial that this data can only be accessible by those authorized to do so. To make the data publicly accessible in dedicated public repositories, storage environments will investigate in depth options such as Zenodo. Any personal data and sensitive information are stored after applying pseudo-anonymisation techniques and methodologies.

In case of data breach, the respective project partner is obliged to urgently notify (not later than 72 hours), the relevant national data protection authority and the subjects (partners, participants, citizens, etc.) that might be affected. The partner will develop a thorough report of the data breach. It will document the personal data breaches, including any information leaked, its effects and the remedial action(s) taken.

6.1.1 Archiving and preservation

The EO4EU partners will utilize state-of-the-art methodologies towards ensuring the secure storage, delivery and access of all kinds of project's data and related material, managing and setting clearly the access rights among the various user types.

All partners will contribute to the later versions of this document (D1.4, D1.5), by describing their privacy-confidentiality issues as well as the procedures to be implemented in terms of data collection, storage, sharing policies (especially when third party countries are concerned), protection, retention and destruction.

All the relevant EU legislation, such as GDPR and relevant national legislation, will be applied on information of an individual and any reference to personal data or sensitive information will be fully masked in any printed materials, project reports or dissemination activities.

Personal data, such as personal information from project partners members, will be treated confidentially, taking into consideration all the proper technical means. General and personal data will be stored separately. Where deemed necessary, encryption such as public-key encryption and symmetric encryption with session keys negotiation over HTTPS, will be applied on personal data. In cases of sensitive evidence and information collected by public authorities and/or consortium members the data will be stored in a locked server. Access to this information will be restricted to authorized personnel only and an access log will be maintained to record access to data.

All data containing private information will be destroyed upon completion of the respective study/research task. In any case, all personal data will be destroyed automatically anonymized at the end of the project and only anonymous or non-identifiable data will be retained after the completion of the final report.

7 Ethics

All details about ethics and legal compliance in terms of current EU legislative initiatives, anonymisation procedures, consent needed, restrictions on third parties, procedures for handling sensitive data and data owners will be included in the corresponding deliverables of WP1 (D1.10, D1.11). Procedures and clear protocols for collection and management of research data (gathering, processing and disseminating) will be defined and implemented with the support of the consortium, the EAB members as well as the Ethics Manager. All data that is transferred to the EO4EU repository shall be either pseudonymised or completely anonymized. The Data Owner/Data Provider is responsible for the anonymization or pseudonymization process and for ensuring that identifiable variables are not transferred to the EO4EU repository.

The procedures and criteria that will be used to identify/recruit research participants must be submitted as part of the DMP deliverable (D1.3). The informed consent procedures that will be implemented for the participation of humans and in regard to data processing must be submitted as part of this deliverable (see Annex 2). Templates of the informed consent/assent forms and information sheets covering the voluntary participation (humans) and data protection issues (in language and terms intelligible to the participants) must be submitted as part of this deliverable. Details on incidental findings policy must be also submitted as part of this deliverable. The applicant must clarify whether vulnerable individuals/groups will be involved, and the measures to protect them and minimize the risk of their stigmatization must be submitted as part of this deliverable, in the next versions.

Meanwhile, information about the expected future updates of this document will be thoroughly described in the next versions of the DMP, in months 18 and 36. Additionally, the Grant Agreement and the EO4EU Consortium Agreement are to be referred to for further details on the ownership and management of intellectual property and access.

8 Other issues

The procedures for data sharing, and requesting access to use the data will be described in the next version of this deliverable.

9 Conclusion

This deliverable provides the first iteration on the description of the data that EO4EU will generate, process and manage during its lifetime together with challenges and constraints that need to be considered to (i) ensure that the project's research data and records will be accurate, complete, interoperable and reliable, (ii) enhance data security and minimize the data loss risks, and (iii) prevent duplication of efforts allowing others to use some of the data managed by the project.

This document is not final, but evolves during the lifespan of the project. The next version of the EO4EU DMP will provide an elaborated description of data management policies and a log of actions performed to sensitive data collected (if any), which will follow the guidelines described in WP1. A final version of this document will be presented at the end of the project, once all data, their potential uses, management structure and available tools will be clear and well defined.

The EO4EU Consortium will ensure that all generated datasets do not infringe either partner IPR rules or regulations related to personal data protection. A clear and complete mechanism for systematic anonymisation of personal data will be defined and is planned to be in place before the first stage of pilots.

10 References

- [1] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- [2] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- [3] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- [4] <https://diuoagr.sharepoint.com/sites/EO4EU/Shared%20Documents/Forms/AllItems.aspx?newTargetListUrl=%2Fsites%2FEO4EU%2FShared%20Documents&viewpath=%2Fsites%2FEO4EU%2FShared%20Documents%2FForms%2FAllItems%2Easpx&viewid=58fb71e9%2Dd6b6%2D4f5e%2Db546%2D535f5ec41408>
- [5] <https://zenodo.org/>
- [6] <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/data-formats>
- [7] <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/product-formatting>
- [8] <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-3-slstr>
- [9] <https://old.wmo.int/extranet/pages/prog/www/WMOCodes/Guides/GRIB/GRIB1-Contents.html>
- [10] <https://old.wmo.int/extranet/pages/prog/www/WMOCodes.html>
- [11] <http://cfconventions.org/cf-conventions/cf-conventions.html>
- [12] <http://geotiff.maptools.org/spec/geotiff2.4.html#2.4>

11 Annex I - List of datasets

Annex I includes an extensive list of the datasets (rows), already delivered and to be developed in the context of the project's research and implementation activities. The list is composed of nine (9) columns each of which present the key aspects of each dataset. Aside from the expected deliverables, Annex 2 also includes all the supportive and complementary datasets (documents, excel, contacts, etc) structured towards delivering the goals of the project.

Table 6 – List of datasets

WP	Del	Task	Description	Lead	Type	Archiving	Diss. level	Est. Due Date
WP1	-	T1.1	Information about the consortium and legal documents (Grant Agreement, Consortium Agreement, mailing list)	NKUA	Project files	Sharepoint	Confidential	-
WP1	-	T1.1	Meetings minutes & recorded teleconferences	NKUA	Project files	Sharepoint	Confidential	-
WP1	-	T1.1	A dataset with the research activities that described the outcomes of the research activities, in terms of deliverables, publications as well as multimedia data collected during such activities.					
WP1	D1.1	T1.1	A management plan that will consist of a list with all the responsibilities of each partner, templates and set of documentation for the management process, quality checks and procedures for all deliverables and the internal and external review processes.	NKUA	Deliverable	Sharepoint	Public	31/08/2022
WP1	-	T1.1	Development data, implementations and codes	NKUA	Code	https://gitpcmp.di.uoa.gr/	Confidential	1/11/2022
WP1	-	T1.1	Bimonthly reports on the progress of the project	NKUA	Code	Sharepoint	Confidential	1/10/2022
WP1	D1.2	T1.3	A quality and risk management plan.	ENG	Deliverable	Sharepoint	Public	30/11/2022
WP1	D1.3	T1.4	This deliverable will provide the plan for managing the data generated and collected during the project.	NVCR	Deliverable	Sharepoint	Public	30/11/2022

12 Annex II - Informed Consent

Part A – Information Sheet

The following information is provided to inform you about the “EO4EU” project and help you to decide about your participation or not. The informed consent explains the purpose of the project in detail and how you can participate to provide feedback on the following aspects: **[add by project partners as needed]**.

Please read this consent form carefully and ask as many questions as you wish before you decide whether you want to participate in this research. Feel free to ask questions at any time before, during, and/or after your participation. You should be aware that even if you agree to participate, you are free to withdraw at any time without saying the reason and with no repercussions for you.

Project title: “AI-augmented ecosystem for Earth Observation data accessibility with Extended reality User Interfaces for Service and data exploitation - (EO4EU)”

Project partners:

1. ETHNIKO KAI KAPODISTRIAKO PANEPISTIMIO ATHINON – NKUA, with legal address 6 CHRISTOU LADA STR, 10561, ATHINA, EL, the Coordinator
2. EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS – ECMWF, with legal address SHINFIELD PARK, RG2 9AX, READING, UK,
3. D.TSAKALIDIS-G. DOMALIS OE – NVCR, with legal address PARODOS THEOFRASTOU 140, 1, 26443, PATRA, EL,
4. CINECA CONSORZIO INTERUNIVERSITARIO – CINECA, with legal address VIA MAGNANELLI 6/3, 40033, CASALECCHIO DI RENO BO, IT,
5. VILNIAUS UNIVERSITETAS – VU, with legal address UNIVERSITETO G. 3, 01513, VILNIUS, LT,
6. LATVIJAS UNIVERSITATE – LU, with legal address RAINIS BOULEVARD 19, 1586, RIGA, LV,
7. ILMATIETEEEN LAITOS – FMI, with legal address Erik Palmenin aukio 1, 00560, HELSINKI, FI,
8. FONDAZIONE CENTRO EURO-MEDITERRANEO SUI CAMBIAMENTI CLIMATICI – CMCC, with legal address VIA MARCO BIAGI 5, 73100, LECCE, IT,
9. SISTEMA GMBH – SIS, with legal address TIEFER GRABEN 19 TOP 2, 1010, WIEN, AT,
10. DANAOS SHIPPING COMPANY LIMITED – DANAOS, with legal address CHRISTAKI KOMPOU 3, 3011, LIMASSOL, CY,
11. KENTRO MELETON ASFALIAS – KEMEA, with legal address P KANELLOPOULOU 4 ST, 10177, ATHINA, EL,
12. EBOS TECHNOLOGIES LIMITED – EBOS, with legal address ARCH. MAKARIOU III AND MESAORIAS 1 OFFICE 101, 2322, NICOSIA, CY,
13. TRUST-IT SRL – TRUST-IT, with legal address VIA FRANCESCO REDI 10, 56124, PISA, IT,
14. COMMLA SRL – COMMLA, with legal address VIA FRANCESCO REDI 10, 56124, PISA, IT,
15. ENGINEERING INGEGNERIA INFORMATICA SPA – ENG, with legal address PIAZZALE DELL'AGRICOLTURA 24, 00144, ROMA, IT,
16. INTELLIGENCE FOR ENVIRONMENT AND SECURITY SRL IES SOLUTIONS SRL – IES, with legal address VIA MONTE SENARIO 98, 00141, ROMA, IT,
17. METEOROLOGICAL AND ENVIRONMENTAL EARTH OBSERVATION SRL – MEEO, with legal address CORSO ERCOLE I D'ESTE 6/A, 44121, FERRARA, IT,

18. FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV – IVI, with legal address HANSASTRASSE 27C, 80686, MUNCHEN, DE, acting as legal entity for and on behalf of its Fraunhofer-Institut für Verkehrs- und Infrastruktursysteme IVI

And the Associated Partners:

19. HAUTE ECOLE SPECIALISEE DE SUISSE OCCIDENTALE – HES-SO, with legal address Route de Moutier 14, 2800, DELEMONT CH

Project-Coordinator: Stathes Chatziefthymiades - NKUA

Technical-Coordinator: Vasileios Baousis - ECMWF

Local coordinator(s): [to be filled by the name of partner(s) in charge] Contact Information: [email and phone number]

1. Purpose of the research [to filled in by partners, adapted according to the objectives of data collection]

2. Description of EO4EU [what is the project about]

A vast amount of Earth Observation data is produced daily and made available through online services and repositories. Contemporary and historical data can be retrieved and used to power existing applications, to foster innovation and finally improve the EU citizens' lives. However, an undersized audience follows this activity, leaving huge volumes of valuable information unexploited. EO4EU aims to provide innovative tools, methodologies and approaches that would assist a wide spectrum of users, from domain experts and professionals to simple citizens to benefit from EO data. EO4EU strives to deliver dynamic data mapping and labelling based on AI adding FAIRness to the system and data. EO4EU introduces an ecosystem for holistic management of EO data, bridging the gap among domain experts and end users, bringing in the foreground technological advances to address the market straightness towards a wider usage of EO data. EO4EU envisages to boost the Earth Observation data market, providing a digestible data information modeling for a wide range of EO data, through dynamic data annotation and a state-of-the-art serverless processing by leveraging important European Cloud & HPC infrastructures.

3. Participant selection [to filled in by partners, indicating why these people have been chosen to participate in the research with reference to the objectives of data collection]

4. Explanation of the procedure to be followed [description of the structure of the research and flow to be followed]

5. Reimbursement [State clearly, what you will provide the participants with as a result of their participation.]

Participation is voluntary and no remuneration is foreseen.

6. Duration and requirements [how much time participants will have to allocate to the research and whether any specific requirement is needed, e.g., ICT literacy, etc.]

7. Risks [Explain and describe any risks that you anticipate or that are possible.]

8. Benefits **[mention the individual benefits, the benefits to the community in which the individual resides, and benefits to society as a whole, as a result of finding an answer to the research questions.]**
9. Confidentiality **[Explain how the research team will maintain the confidentiality of data with respect to both information about the participant and information that the participant shares. If the research is sensitive and/or involves participants who are highly vulnerable, explain to the participant any extra precautions you will take to ensure safety and anonymity and avoidance of stigmatization. In case of focus groups: Focus groups provide a particular challenge to confidentiality because once something is said in the group it becomes common knowledge. Explain to the participant that you will encourage group participants to respect confidentiality, but that you cannot guarantee it.]**

We will not be sharing information about you with anyone outside of the research team. The information that we collect from this research project will be kept private.

10. Participants' right to the termination of research study

You are free to choose whether or not to participate in this study. There will be no penalty or loss of benefits to which you are otherwise entitled if you choose not to participate. In the event you decide to discontinue your participation in the study, there will not be any disadvantages for you. You can decide to discontinue your participation at any time and without saying the reason. In this case, please inform us by contacting **[add the name, the affiliation, and the contact details, email, phone number of the local partner in charge]** and we will take care of all necessary steps to remove your data!

11. Data processing **[add the name and affiliation of the person in charge of data processing and contact details]**
12. During the research the following personal data will be processed **[list all personal data, indicate whether video or audio recording will take place]**
13. How will your data be stored and transferred during the project?

- [1. Identify all locations where data will be stored, indicating for each location whether it will be used to store identifying information or de-identified research data, and providing details of access controls that will be applied.]**
- 2. Describe any administrative measures that you will take to control the risks of inappropriate disclosure, e.g., pseudonymization, and procedures for secure transfer between locations, e.g., using file encryption and encrypted channels.**
- 3. Specify who will be able to access the identifying information and how you will ensure they process the information securely, e.g., through training, supervision and adherence to secure data handling procedure.**
- 4. Confirm that all data requested are adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed, in accordance with Article 5 of GDPR]**

14. How will research data be preserved and shared on completion of the project?

- [1. Identify the research data that will be preserved and shared at the end of the project by deposit in a public data repository or other archiving solution.]**

2. Describe the measures that will be taken to ensure data are suitable for sharing, e.g., securing consent, anonymizing data prior to deposit/sharing, sharing confidential or high-risk information under a controlled access policy.

3. Identify data repositories or other solutions that will be used to preserve and share data]

15. How will retention and disposal of personal data and confidential information after project completion be managed?

[State how long you plan to retain personal data/confidential information after the end of the project. Specify under whose authority this information will be maintained and disposed of after the project]

16. Publications

Results of this study can be publicized with anonymized data only. This ensures that it will not be possible to identify you in any way. Your personal data will never be published!

The Ethics Manager, Mr. Sarantis Paskalis (NKUA), will oversee that the procedures run smoothly; in case of any non-compliance, the Ethics Manager in collaboration with the Ethics Committee will decide the timeline and mitigation measures, so as to take all the necessary actions in a timely manner and ensure full compliance with the legal and ethical requirements of the project. In all cases, all non-compliant actions will be immediately suppressed or suspended, and the partners will implement a mitigation action within 5 working days.

Part b- Consent Form

Project title: “AI-augmented ecosystem for Earth Observation data accessibility with Extended reality User Interfaces for Service and data exploitation - (EO4EU)”

Participant name: _____

Gender: _____ [plus the option I prefer not to say]

Age range: 18-25, 26- 35, 36-45, 46-55, 56+

Stakeholder role: [select the one(s) that apply: public servant, citizen, user of the service, case study participant, NGO representative, other (pls specify _____)]

Which is your highest degree in education:

Email address: _____

I have read very carefully and completely understood this consent form, and I volunteer to participate in this research study, as a choice under free will. I understand that I will receive a copy of this form. I have had the opportunity to ask, and I have received answers to, any questions I had regarding the study.

I was thoroughly informed about the aim and course of the study, my rights and obligations, and the voluntary nature of participation.

I agree with the processing of my personal data as part of the research project as described in the Participant Information Form. I was informed about my privacy rights, especially my right to cancel my participation.

Place, date _____

Participants' signature

Place, date _____

Signature of the partner in charge