A CLOUD-BASED REMOTE SENSING DATA SYSTEM FOR PROMOTING RESEARCH AND SOCIO-ECONOMIC STUDIES IN ARCTIC ENVIRONMENTS

Deliverable:

"D1.3 Data Management Plan"

Contributing Work Package	WP 1: Project Management and Coordination
Task	Knowledge and Information Management
Deliverable Lead	HUA
Authors	Patryk Grzybowski (CloudFerro)
Deliverable type	Deliverable type (Report)
Dissemination level	Public
Version number	1.0
	This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101086386

Document Log

Issue	Date	Summary of Changes	Author
0.1	31/05/2023	First version of Data Management Plan	Patryk Grzybowski (CloudFerro)
0.2	16/06/2023	Internal revised and corrected version of Data Management Plan	Patryk Grzybowski (CloudFerro) Pol Kolokoussis (Internal revision) (NTUA)
0.3	23/06/2023	Quality review	Alkyoni Baglatzi (Cloudeo Hellas)
1.0	26/06/2023	Final version of Data Management Plan	Patryk Grzybowski (CloudFerro

Table of content

Acronyms and Abbreviations

ALOS2 - The Advanced Land Observing Satellite 2
DEM – Digital Elevation Model
DMP – Data Management Plan
DoA - Description of the Action
ECMWF - European Centre for Medium-Range Weather Forecasts
EO – Earth Observation
ESA – European Space Agency
FAIR – Findable; Accessible; Interoperable; Reusable
FMI - Finnish Meteorological Institute
GDP - Gross domestic product
GNSS - Global Navigation Satellite Systems
HUA - Harokopio University of Athens
IGS - International GNSS Service
JAXA - Japan Aerospace Exploration Agency
LU – Lund University
NASA - National Aeronautics and Space Administration
NLS - National Land Survey of Finland
NOAA - National Oceanic and Atmospheric Administration
NTUA - National Technical University of Athens
OECD - Organisation for European Economic Co-operation
RS – Remote sensing
SAR - Synthetic Aperture Radar
SMOS - Soil Moisture and Ocean Salinity
SSMIS - Special Sensor Microwave Imager / Sounder
TAR - Transgressive Action Research
TSInSAR - Time Series Interferometric Synthetic Aperture Radar
UB - University of Bucharest
UC – Use case
USGS - United States Geological Survey
WP – Work package

1. Executive summary

This deliverable aims to collect information about the data sources, data management and data itself that the EO-PERSIST project consortium will work with. This deliverable has been compiled with collaborative work among the consortium partners who are involved in data collection, production and processing.

The scope of the Data Management Plan (DMP) is to compile a description of the data management life cycle for all data sets and to specify which data will be findable, openly accessible, interoperable and re-usable (FAIR). It also includes information about legal issues, privacy and maintenance. More specifically, for each work package (WP), the partners have to define the data sources and formats they will use and create a Data Management Plan.

Following the FAIR template, datasets are analyzed, providing answers about how the data will be managed within the project and also describing how it will be providing as much open and re-usable data as possible from the execution with the purpose of facilitating to others the reuse of such data.

The DMP is a live document and will be regularly updated as more information becomes available and data issues are resolved.

The deliverable is structured in the following chapters:

- Chapter 1: Introduction
- Chapter 2: Methodology description;
- Chapter 3: DMP components in EO-PERSIST work packages;
- Chapter 4: Summary.

2. Introduction

The Deliverable D1.3 - Data Management Plan is the first version of the DMP for the purposes of EO-PERSIST project. The goal of the EO-PERSIST project is to establish a fertile collaborative research and innovation environment by means of staff exchanges, knowledge sharing and know-how transfer that will leverage existing services, datasets and novel technologies to a) create a continuously updated ecosystem of EO-based datasets suitable for permafrost applications, b) promote methodological advances in the field of permafrost by exploiting the huge volume of remote sensing (RS) datasets and c) provide indicators directly connected with socioeconomic effects from permafrost dynamics. The main research and innovation of the proposed project is the availability in a unique way of the collection, management and exploitation of the available Earth observation (EO) data suitable for permafrost socioeconomic studies and research via one single cloud-based system/platform.

3. Methodology

DMP methodology has been established by the European Commission - Directorate-General for Research & Innovation in February 2016 within "Guidelines on FAIR Data Management in Horizon 2020". There has been an updated version (version 3.0) was published on 26 July 2016. The EO-PERSIST DMP addresses the following issues:

- Data summary;
- FAIR data;
- Allocation of resources;
- Data security;
- Ethical aspects;
- Other issues.

3.1. Data summary

The Data Summary addresses the following issues:

- What is the purpose of the data collection/generation and how does it relate to the objectives of the project?
- What types and formats of data will the project generate/collect?
- Will you re-use any existing data and how?
- What is the origin of the data?
- What is the expected size of the data?
- To whom might the data be useful ("data utility")?

3.2. FAIR data

FAIR methodology is a Horizon 2020 methodology dedicated to Data Management Plan. It means that research data should be findable (F), accessible (A), interoperable (I) and re-usable (R). It does not imply any specific technology, standard, or implementation-solution. Within this part, there are going to be introduced specific elements of FAIR terms.

3.2.1. Making data findable, including provisions for metadata

The section "Making data findable, including provisions for metadata" aims to answer the following questions:

- Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?
- What naming conventions does it follow?
- Will search keywords be provided that optimize possibilities for re-use?
- Does it provide clear version numbers?
- What metadata will be created? In case metadata standards do not exist in the discipline, outline what type of metadata will be created and how.

3.2.2. Making data openly accessible

Within this section answers for the following questions are going to be provided:

- Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), why, clearly separating legal and contractual reasons from voluntary restrictions.
  - it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.
- How will the data be made accessible (e.g. by deposition in a repository)?
- What methods or software tools are needed to access the data?
- Is documentation about the software needed to access the data included?
- Is it possible to include the relevant software (e.g. in open source code)?
- Where will the data and associated metadata, documentation and code be deposited?
- Have appropriate arrangements with the identified repository been explored?
- If there are restrictions on use, how will access be provided?
- Is there a need for a data access committee?
- Are the terms of access well described (e.g. a machine-readable license)?
- How will the identity of the person accessing the data be ascertained?

3.2.3. Making data interoperable

The Making data interoperable addresses the following issues:

- Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organizations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?
- What data and metadata vocabularies, standards or methodologies will you follow to make the data interoperable?
- Will standard vocabularies be used for all data types present in your dataset to enable interoperability across disciplines?
  - In case it is unavoidable that it uses uncommon ontologies or generates project specific ontologies or vocabularies, will it provide mappings to more commonly used ontologies?

3.2.4. Increase data re-use (through clarifying licenses)

This section addresses the following issues:

- How will the data be licensed to permit the widest possible re-use?
- When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
- Are the data produced and/or used in the project usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
- For how long is it intended that the data remains re-usable?
- Are data quality assurance processes described?

3.3. Allocation of resources

This part of the DMP aims to answer the following questions:

- What are the costs of making data FAIR in your project?
- How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).
- Who will be responsible for data management in your project?
- Have the resources for long term preservation been discussed (costs and potential value, who decides and how what data will be kept and for how long)?

3.4. Data security

Data Security addresses the following issues:

- What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?
- Is the data safely stored in certified repositories for long term preservation and curation?

3.5. Ethical aspects

Ethical aspects are addressing the following issues:

- Are there any ethical or legal issues that can affect data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
- Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

3.6. Other issues

This section provides an answer for a question:

- Will other national/funder/sectorial/departmental data management procedures be used? If yes, which are they?

4. Data Management Plan (DMP) components in EO-PERSIST

4.1. DMP Components in WP1 - Project Management and Coordination

DMP component		Related issues
Data summary		Collecting data within this WP aims to: • establish management committees and guidelines for their operation, • provide financial and technical monitoring and reporting standards, • define risk management, • establish quality control of deliverables, • establish administration of Consortium Agreement, • establish the delivery of all necessary reports including periodic and final project reports. This WP will create and share documents in .docx (draft of final documents and reports) and .pdf formats (final version of the documents and reports). The expected size of the files is not applicable. Moreover, tables (.xlsx/.xls) and text files (.csv) containing auxiliary data used as a source of information for the preparation of specific documents will be created. The expected size of the files is not applicable. However, their size is not supposed to be meaningful. Data within this WP is going to be created by project's participants. Templates provided by European Commission via website are going to be re-used in order to prepare specific document (e.g. in purposes of preparing of this document template provided via https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm has been used). In general, data which is going to be created within this WP are deliverables. The data will be useful for project's participants and stakeholders due to provision of legible and clear way of naming and storing deliverables.
FAIR DATA	**Making data findable, including provisions for metadata**	Deliverables within this WP are sensitive (except D1.3 – Data Management plan) and they will be shared only

		among partners and European Commission. They will be collected using object storage (s3fs protocol) provided by CloudFerro (only project's participants will be allowed to allocate data on this kind of storage). Deliverables will also be collected using MS teams consortium channel managed by HUA. The naming conventions used for these data will be: • DX.Y TITLE where, DX.Y is the number of the deliverable while TITLE is deliverable's name. As a part of newly created data, metadata will be generated. It will include: • contributing Work Package names, • task, • deliverable Lead, • authors, • deliverable type, • dissemination level, • version number. It is supposed that within the process of generation data some new/additional parameters for metadata will be included. Appropriate keywords will be included.
	Making data openly accessible	Data desired to be shared to public will be made publicly available as part of the aforementioned deliverables and through the EO-PERSIST website.
	Making data interoperable	N/A
	Increase data re-use	Data will be publicly available as part of the deliverables and could be accessed and re-used by public indefinitely without a license.
Allocation of resources		No additional costs are foreseen for making this dataset FAIR.
Data security		N/A
Ethical aspects		There are no ethical aspects or legal conditions.
Other issues		N/A

4.2. DMP Components in WP2 - Communities of Practices (CoP), User Requirements and EO-PERSIST design

DMP component	**Related issues**
Data summary	Collecting data within this WP aims to: • identify end-users' needs, • establish specification of system requirements, • design of the overall system architecture. • inventory and describe data repositories. Within this WP documents in .docx (draft of final

		documents and reports) and .pdf (final versions of documents and reports) formats will be created and shared. The expected size of the files is not applicable. Moreover, tables (.xlsx/.xls) and text files (.csv) will be created, containing auxiliary data which is going to be used as a source of information in order to create specific documents. The expected size of the files is not applicable. However, their size is not supposed to be large. Data within this WP is going to be created by the project's participants. However, this data will be based on data provided by third parties like end-users (through questionnaires, feedback during workshops, interviews) in order to identify their needs. The data regarding the users' needs will provide information with respect to the characteristics of specific roles of the end users (e.g.: researchers, data scientist, data engineers, IT, coordinators, students). This data will be useful for software developers and researchers who wish to find and understand system requirements that form the core of the development process of platforms and tools for remote monitoring of permafrost areas. Inventory and description of data repositories is necessary to determine what kind of data will be used within the project as well as their geographic and temporal coverage. It is crucial to set the required storage of the EO-PERSIST system as well as requirements of the project's participants for enabling data processing. In general, data collected within this WP will be useful for project's participants due to provision of roadmap for creation of EO-PERSIST system (user needs -> system requirements -> system architecture) and setting hardware and software requirements for data processing and performing data analysis and system validation processes, respectively.
FAIR DATA	**Making data findable, including provisions for metadata**	The data produced either via the online survey or the workshops with the end-users will not be identifiable to the individual respondent. However, there will be metadata that will allow the institution and roles to be identified. This kind of data will be a source of identification of end-users' needs which is going to lead to specification of system requirements. These questionaries and single responses are not going to be shared with the public. Only summaries and statistics based on that data will be published. Deliverables within this WP are sensitive, and they will be shared only among partners and the European Commission. They will be collected using object storage

		(s3fs protocol) provided by CloudFerro (only project's participants will be allowed to allocate data on this kind of storage). Deliverables will be also collected using MS teams consortium channel managed by HUA. The naming conventions used for these data will be: • DX.Y TITLE where, DX.Y is the number of the deliverable while TITTLE is deliverable's name. As a part of newly created data, metadata will be generated. It will include: • contributing Work Package names, • task, • deliverable Lead, • authors, • deliverable type, • dissemination level, • version number. It is supposed that within the process of generation data some new/additional parameters for metadata will be included. Appropriate keywords will be included.
	Making data openly accessible	The datasets contain raw data obtained by surveys, interviews and questionaries will not be publicly available. All the insights, conclusions and solutions developed within this WP will be made publicly available as part of the aforementioned deliverables and through the EO-PERSIST website.
	Making data interoperable	The data collected and produced within this WP will be used both in deliverables within this WP and within other WPs. Data collected by surveys, interviews and questionaries is going to be used to identify end-users' needs which are going to be used for specifying system requirements. These findings will be crucial to designing the overall system architecture. Then, inventorying and description of the data repositories will lead to identify necessary data for further analysis, processing, modelling and sharing.
	Increase data re-use	Data will be publicly available as part of the deliverables and could be accessed and re-used by public indefinitely without a license.
Allocation of resources		No additional costs are foreseen for making this dataset FAIR.
Data security		The data collected by interaction with end-users will be collected for internal use in the project, and not intended

	for long-term preservation.
Ethical aspects	There is no ethical aspects or legal conditions.
Other issues	It is supposed that during the project some data which is not included in the inventory could be used anyway. On the other hand, some data included in the inventory could be revealed as insignificant and they could be removed from the analysis. In such a case, storage of allocation could change.

4.3. DMP Components in WP3 - Socioeconomic impacts assessment related to thawing permafrost

DMP component	Related issues
Data summary	Collecting data within this WP aims to: • estimate the socio-economic impact of the thawing permafrost (modelling included); • conduct qualitative assessment by using a discourse analysis of media and relevant social media, together with online exchanges with key informants and Transgressive Action Research (TAR) techniques; • establish methodology to indicate land degradation and temperature anomalies; • establish methodology for coastal monitoring and modelling purposes. Inventory and described data (mentioned in chapter 4.2) as well as data related to population density, migration, distribution, income inequality and farm production will be used to assess socioeconomic impacts related to thawing permafrost. Within this WP data will be created and shared in: • .docx format - drafts of documents and reports; • .pdf format - final versions of the documents and reports; • .shp, .Json format – area of interest of land degradation, temperature anomalies, and costal monitoring; results of land degradation, temperature anomalies, and costal monitoring • .jp2, .tiff, .nc,. ascii – results of land degradation, temperature anomalies, and costal monitoring analysis; • .rdata, .svc – data analysis and data processing. The expected size of the files is not applicable. However, size of the single files and datasets within this WP can be very large (>1 GB and >10 TB, respectively), so any changes could imply significant increase of the required storage.

		Data will be an input to the EO-PERSIST system. It will be spatial data. It will cover the needs to implement and validate some of the desired products of the project. The final data outcomes of this WP are socioeconomic estimated coefficients. Data are also going to be used in order to perform three use cases:
		• UC 1: Land degradation:
		  ○ Assessing the relationship between the permafrost thawing and landslide occurrences,
		  ○ Mapping areas with possible thermokarst induced by permafrost thawing,
		  ○ Mapping the transformation of terrestrial ecosystems into aquatic ones;
		• UC 2: Temperature Anomalies:
		  ○ Site monitoring/risk assessment for chemical gases and plume detection and/or identification,
		  ○ Area monitoring/risk assessment for gas, hot water and high voltage transportation infrastructure leak detection;
		• UC 3: Coastal monitoring & modelling.
		
		The origin of the data for WP3 will be from various sources such as ESA, Copernicus Programme, NASA, NOAA, USGS, Eurostat, OECD, JAXA, ECMWF, NLS and World Bank. Also, data of the project's participants is going to be used (e.g. in-situ measurements). The origin of the data will include SAR, optical and thermal satellite data, land, marine atmospheric data, meteorological conditions and climate risk data, in situ data for soil moisture and permafrost, soil moisture maps, land deformation maps, GNSS data, snow cover data, socioeconomic and demographic data. That original data is going to be used to create a pipeline to extract socio-economic indicators.
		
		This data will be useful to end-users as one of the outcomes of the project.

FAIR DATA	**Making data findable, including provisions for metadata**	Deliverables within this WP are sensitive and they will be shared only among partners and the European Commission. They will be collected using object storage (s3fs protocol) provided by CloudFerro (only project's participants will be allowed to allocate data on this kind of storage). Deliverables will also be collected using MS teams consortium channel managed by HUA. Processed results of this WP will be available via EO-PERSIST website.The naming conventions used for these data will be:

		• DX.Y TITLE where, DX.Y is the number of the deliverable while TITTLE is deliverable's name. Scientific papers based on the data generated within this WP, in particular the results of several regression models such as generalized least square models will be presented to analyze relationships between the socioeconomic variables such as GDP, population structure, employment and related thawing permafrost indicators, would be published in open access journals. Spatial data (describing socioeconomic impacts) will be published in EO-PERSIST service as a one of the crucial element of the project. As a part of newly created data, metadata will be generated. Deliverable's metadata will include: • contributing Work Package names, • task, • deliverable Lead, • authors, • deliverable type, • dissemination level, • version number. Spatial data's metadata will include: • date of generation, • author (institution), • version number. It is supposed that within the process of generation data some new/additional parameters for metadata will be included. Appropriate keywords will be included.
	Making data openly accessible	All the insights, conclusions and solutions developed within this WP will be made publicly available as part of the aforementioned deliverables and through the EO-PERSIST website. Data including methodology, description of modelling activities, its results and results' validation will be included in scientific open-access papers. The raw data that will be provided to the technical team will not be publicly available to third parties
	Making data interoperable	There will be used formats and standard vocabularies for all data types present in the datasets, so it could be used by the public.
	Increase data re-use	Data will be publicly available as part of the deliverables and could be accessed and re-used by the public

		indefinitely without a license. Insight published in scientific papers could be re-used in purposes of another research within the project or used by public.
Allocation of resources		No additional costs are foreseen for making this dataset FAIR.
Data security		Data is going to be regularly backed up and secured.
Ethical aspects		There is no ethical aspects or legal conditions.
Other issues		Origin of data could change during the project. Also, additional data format could be used. In such a case, storage of allocation could change.

4.4. DMP Components in WP4 - Scientific advances for permafrost monitoring

DMP component	Related issues
Data summary	Data within this WP is going to be used in order to: • develop innovative algorithmic approaches in the permafrost domain; • develop material for training, education and research; • develop algorithms code. There will be created and shared data in: • .docx format - drafts of documents, reports and training materials; • .pdf format - final versions of the documents, reports and training materials; • .shp, .json format – vector type data and results (data dedicated to points and polygons objects); defining area of interest; • .tiff format - raster type data and results (spatial continuously data); • .nc format – many dimensional spatial continuous data. The expected size of the files is not applicable. However, size of the single files and datasets within this WP is going to be very large (>1 GB and >10 TB, respectively), so any changes could imply significant increase of the required storage. The source of the data for WP4 will be from ESA, Copernicus services, NASA, NOAA, IGS, JAXA, ECMWF and NLS. It will include mainly: • meteorological data including reanalysis data (e.g. air temperature, wind speed and direction, snow cover); • land cover data including DEMs, land cover type, vegetation (forest biomass, canopy closure); • satellite observations including single-look-

		complex SAR imagery (Sentinel-1, ALOS2) and passive microwave data (SMOS, SSMIS); • GNSS data. Furthermore, availability of ALOS2 data for third parties is under investigation. The SAR data will be used to establish methodology for monitoring active layer freeze-thaw cycle and methodology for monitoring permafrost active layer dynamics and snow modelling. This data is going to be inventoried and described in chapter 4.2. To ensure the innovative aspect of the developed algorithm, information from GNSS and atmospheric models will be utilized for the development and the validation phase of the algorithm. The output of this step is going to be an TSInSAR approach which will exploit information from distributed scatterers and include spatiotemporal unwrapping functionalities. Data is also going to be used in order to perform two use cases: • UC 4: Active layer freeze-thaw cycle monitoring using TSInSAR; • UC 5: Exploitation of low frequency SAR observations for monitoring permafrost active layer dynamics and snow modeling. The generated data will be useful for end-users and will be an input to the EO-PERSIST system (as a raster layer, documentation and training materials). What is more, it will introduce an innovative approach – TSInSAR – in public.
FAIR DATA	**Making data findable, including provisions for metadata**	Deliverables within this WP are sensitive, and they will be shared only among the project's consortium and the European Commission. They will be collected using object storage (s3fs protocol) provided by CloudFerro (only project's participants will be allowed to allocate data on this kind of storage). Deliverables will also be collected using MS teams consortium channel managed by HUA. Processed results of this WP will be available via EO-PERSIST website. The naming conventions used for these data will be: • DX.Y TITLE where, DX.Y is the number of the deliverable while TITTLE is deliverable's name. Scientific papers consisting of data generated within this WP would be published in open access journals. Spatial data (raster data) will be published in EO-PERSIST

		service as a crucial element of the project. Training materials will be available in the project's website (www.eo-persist.eu) and they will be open to the public. Their naming conventions: • TRANING MATERIAL #AAA: BBBBBBB where, AAA is the number (ID) of the training material while BBBBBBB is the name of the training material. As a part of newly created data, metadata will be generated as extensible markup language files (.xml). Deliverable's metadata will include: • contributing Work Package names, • task, • deliverable Lead, • authors, • deliverable type, • dissemination level, • version number. Spatial data's metadata will include: • date of generation, • author (institution), • version number. Training materials' metadata will include: • date of generation, • author (institution), • version number. It is supposed that within the process of generation data some new/additional parameters for metadata will be included. Appropriate keywords will be included.
	Making data openly accessible	All the insights, conclusions and solutions developed within this WP will be made publicly available as part of the aforementioned deliverables and through the EO-PERSIST website. Data including methodology, description of modelling activities, its results and results validation will be included in scientific open-access papers. Training materials and algorithms code will be provided to the public through the EO-PERSIST website. The raw data that will be provided to the technical team will not be publicly available to third parties.
	Making data interoperable	There will be common formats and standard vocabularies for all data types present in the datasets, so that they can be used by the public. Training materials are going to be universal. They will

		include how to use established methodology and EO-PERSISTS related data for different areas of interests.
	Increase data re-use	Data will be publicly available as part of the deliverables and could be accessed and re-used by the public indefinitely without a license. Insight published in scientific papers could be re-used for other research purposes within the project or used by the public. Training materials aim to encourage end-users to use EO-PERSISTS related data to use within their own activities. These materials could be used many times by various institutions.
Allocation of resources		No additional costs are foreseen for making this dataset FAIR.
Data security		Data is going to be regularly backed up and secured.
Ethical aspects		There is no ethical aspects or legal conditions.
Other issues		Origin of data could change during the project. Also, additional data format could be used. In such a case, storage of allocation could change.

4.5. DMP Components in WP5 - Implementation of EO-PERSIST system

DMP component	**Related issues**
Data summary	Data generated within this WP is going to be used in order to: • describe data management module on the EOPERSIST system; • describe connectivity module to other platforms and external sources of the EO-PERSIST system; • describe data processing module of the EO-PERSIST system; • describe front-end interface of the EO-PERSIST system; • describe Overall System Integration of the EO-PERSIST system. There will be created and shared data in .docx and .pdf formats. The expected size of the files is not applicable. Integration framework, component descriptions and dependencies, API descriptions, information flow, diagram, internal and external interfaces, software and hardware requirements, use case scenarios and testing procedures will be described. In general, data which is going to be created within this WP are deliverables and technical documentations of the EO-PERSIST system. The system will consist of data generated within the work packages 3 and 4. Thus, it will include published scientific

		data from FMI and LU, remote sensing data from NTUA, UB and GUA, GIS modeling data from planetGIS SKY and CloudEO Hellas as well as socioeconomic data from HUA. Data will be useful for project participants and stakeholders by providing a legible explanation of how the EO-PERSIST system is designed and how it works. Further, the desired platform will be a source of spatial data generated during the project. It will also be a single-entry point to the computing environment and resources, meaning that users will be able to use - discover and process - all datasets from the EO-PERSIST repository in one place.
FAIR DATA	**Making data findable, including provisions for metadata**	Deliverables within this WP are sensitive and they will be shared only among partners and the European Commission. They will be collected using object storage (s3fs protocol) provided by CloudFerro (only project's participants will be allowed to allocate data on this kind of storage). Deliverables will also be collected using MS teams consortium channel managed by HUA. Processed results of this WP will be available via EO-PERSIST website. The naming conventions used for these data will be: • DX.Y TITLE where, DX.Y is the number of the deliverable while TITTLE is deliverable's name. As a part of newly created data, metadata will be generated. It will include: • contributing Work Package names, • task, • deliverable Lead, • authors, • deliverable type, • dissemination level, • version number. It is supposed that within the process of generation data some new/additional parameters for metadata will be included. Appropriate keywords will be included.
	Making data openly accessible	All the data will be made publicly available as part of the aforementioned deliverables and through the EO-PERSIST website.
	Making data interoperable	N/A
	Increase data re-use	Data will be publicly available as part of the deliverables and could be accessed and re-used by the public indefinitely without a license. It could be an example for teams who are interested in building this kind of systems.
Allocation of resources		No additional costs are foreseen for making this dataset

	FAIR.
Data security	Data is going to be regularly backed up and secured.
Ethical aspects	There is no ethical aspects or legal conditions.
Other issues	N/A

4.6. DMP Components in WP6 - Validation of EO-PERSIST system

DMP component	**Related issues**
Data summary	Data generated within this WP is going to be used in order to: • Calibrate and validate developed algorithms exploiting EO-PERSIST system resources; • Establish EO-PERSIST system Version 2.0. There created and shared data of this WP will have an analogous format to the WP3 and WP4: • docx format - drafts of documents, reports and training materials; • .pdf format - final versions of the documents, reports and training materials; • .shp, .json format – vector type data and results (data dedicated to points and polygons objects); defining area of interest; • .jp2, .tiff format - raster type data and results (spatial continuously data); • .nc format – many dimensional spatial continuous data. • .rdata, .svc – data analysis and data processing. The expected size of the files is not applicable. However, size of the single files and datasets within this WP can be very large (>1 GB and >10 TB, respectively), so any changes could imply significant increase of the required storage. Further, the source of the data would be the same as in WP3 and WP4. They will cover data from: ESA, Copernicus Services, NASA, NOAA, USGS, Eurostat, OECD, IGS, JAXA, ECMWF, NLS and World Bank. In addition, data own data of the project's participants is going to be used (e.g. in-situ measurements). The origin of the data will cover SAR, optical and thermal satellite data, land, marine atmospheric data, meteorological conditions and climate risk data, in situ data for soil moisture and permafrost, soil moisture maps, land deformation maps, GNSS data, snow cover data, socioeconomic and demographic data. Origin in-situ measurement data will be crucial in respect to the goals of these WP. Selection of area of interest (for validation purposes) will be dependable on availability of in-situ measurements and its quality. Methodology and data generated within the WP3 and WP4 will be used,

		too. In general, data which is going to be created within this WP are reports on uncertainty assessment of EO-PERSIST products and recommendations which are aimed to improve EO-PERSIST system functionality. For each EO-PERSIST socio-economic indicator, an uncertainty layer exploiting all available multisource datasets will be produced. Data will be useful for project's participants and stakeholders due to provision of legible explanation how EO-SYSTEM was improved and what issues occurred and were fixed. It will also point out any uncertainties and can serve as an input to the scientific papers.
FAIR DATA	**Making data findable, including provisions for metadata**	Deliverables within this WP are sensitive, and they will be shared only among partners and the European Commission. They will be collected using object storage (s3fs protocol) provided by CloudFerro (only project's participants will be allowed to allocate data on this kind of storage). Deliverables will also be collected using MS Team consortium channel managed by HUA. Processed results of this WP will be available via EO-PERSIST website. The naming conventions used for these data will be: • DX.Y TITLE where, DX.Y is the number of the deliverable while TITTLE is deliverable's name. Scientific papers based on data generated within this WP, in particular, TSInSAR approach, results of several regression models, such as generalized least square models will be presented to analyze relationships between the socioeconomic variables such as GDP, population structure, employment and related thawing permafrost indicators, would be published in open access journals. As a part of newly created data, metadata will be generated. It will include: • contributing Work Package names, • task, • deliverable Lead, • authors, • deliverable type, • dissemination level, • version number. It is supposed that within the process of generation data some new/additional parameters for metadata will be included. Appropriate keywords will be included.

	Making data openly accessible	All the data will be made publicly available as part of the aforementioned deliverables and through the EO-PERSIST website.
	Making data interoperable	N/A
	Increase data re-use	Data will be publicly available as part of the deliverables and could be accessed and re-used by public indefinitely without a license.
Allocation of resources		No additional costs are foreseen for making this dataset FAIR.
Data security		Data is going to be regularly backed up and secured.
Ethical aspects		There is no ethical aspects or legal conditions.
Other issues		Origin of data could change during the project. Also, additional data format could be used. In such a case, storage of allocation could change.

4.7. DMP Components in WP7 - Dissemination, Communications, Technology transfer and Exploitation activities

DMP component		**Related issues**
Data summary		Data generated within this WP is going to be used in order to: • develop and implement an effective dissemination and communication strategy; • improve the communication and dissemination activities; • prepare materials, presentations, tutorials, scientific papers etc. Within this WP there created and shared data will be in .docx, .xlsx, .csv, .pptx, .jpg. .png, .gif and .pdf formats. The expected size of the files is not applicable. However, it is going to be meaningful. Reports will be collected from the partners regarding their performed dissemination activities. The data that will be collected will be statistics related to the project website and social media like LinkedIn and Twitter, for tracking the progress and improve the communication and dissemination activities. Presentations (.pptx) are going to be used during conferences, workshops, training material etc. Images (.jpg, .png., .gif) are going to be used during presentations as well as on social media. Data will be useful for the project's participants and stakeholders due to the provision of consistent, meaningful and legible information on projects results. What is more, the constant contact through the social media is going to maintain public interest in the project.
FAIR DATA	**Making data findable, including**	Deliverables within this WP are sensitive (excepted D.7.2

	provisions for metadata	Project website) and they will be shared only among partners and the European Commission. They will be collected using object storage (s3fs protocol) provided by CloudFerro (only project's participants will be allowed to allocate data on this kind of storage). Deliverables will also be collected using MS teams consortium channel managed by HUA. Processed results of this WP will be available via EO-PERSIST website. The naming conventions used for these data will be: • DX.Y TITLE where, DX.Y is the number of the deliverable while TITTLE is deliverable's name. As a part of newly created data, metadata will be generated. It will include: • contributing Work Package names, • task, • deliverable Lead, • authors, • deliverable type, • dissemination level, • version number. It is supposed that within the process of generation data some new/additional parameters for metadata will be included. Appropriate keywords will be included. Regarding the personal data derived from workshops or other sources, they will not be publicly available and only project partners will have access to them upon request to the responsible controller.
	Making data openly accessible	Data will be made publicly available as part of the aforementioned deliverables and through the EO-PERSIST.
	Making data interoperable	N/A
	Increase data re-use	Data will be publicly available as part of the deliverables and could be accessed and re-used by public indefinitely without a license.
Allocation of resources		No additional costs are foreseen for making this dataset FAIR.
Data security		Data is going to be regularly backed up and secured.
Ethical aspects		There are no ethical aspects or legal conditions.
Other issues		Unplanned ways of dissemination could be performed, and additional promotional materials could be generated. Still, it will not implicate additional cost in respect to managing this new data.

5. Conclusions

The introduced DMP refers to the data that is going to be used and poses the data management strategy for each work package. The DMP will evolve over the lifecycle of the project and according to the project impact on data management and security. Updates of the DMP are thus planned in project-month 18 and 36 and will be submitted to the EU as an integral part of the Project Periodic Reports.