



PIISA

Piloting Innovative Insurance
Solutions for Adaptation

D5.11 Project Data Management Plan – Version 2

Authors: Kaisa Juhanko



Funded by
the European Union

Disclaimer

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.



Document information

Grant Agreement	n°101112841
Project Title	Piloting Innovative Insurance Solutions for Adaptation
Project Acronym	PIISA
Project Coordinator	Hilppa Gregow, Finnish Meteorological Institute
Project Duration	1 June 2023 – 31 May 2026 (36 months)
Related Work Package	WP5 Project Management
Deliverable Title	D5.11 Project Data Management Plan – Version 2
Related Task(s)	5.5
Lead Organisation	Finnish Meteorological Institute
Contributing Partner(s)	All
Authors	Kaisa Juhanko
Due Date	30 Nov 2024
Submission Date	29 Nov 2024
Dissemination level	Public

History

Date	Version	Submitted by	Reviewed by	Comments
2.10.2024	1	Kaisa Juhanko		First draft
15.11.2024 – 27.11.2024	2	Kaisa Juhanko	Heikki Tuomenvirta (FMI) Luiz Galizia (AXA) Leone Cavicchia (CMCC)	Review and comments
29.11.2024	3	Kaisa Juhanko	Hilppa Gregow (FMI)	The final version of D5.11 for EC submission.

Table of contents

1	Introduction.....	7
2	Data Summary	8
3	FAIR Data	14
3.1	Making Data Findable	14
3.1.1	Naming of the files.....	14
3.2	Making Data Accessible	15
3.2.1	Repository.....	15
3.2.2	Data.....	16
3.2.3	Metadata	16
3.3	Making Data Interoperable.....	17
3.4	Making Data Reusable	18
4	Other Data Products.....	20
5	Allocation of resources.....	21
6	Data security.....	22
7	Ethics	23
7.1	Responsible Use of AI in Research	24
8	Other issues.....	25

List of tables

Table 1:	Initial outline of the novel open-source risk indicators integrated into the Risk Data Hub.....	9
Table 2:	Data types produced in PIISA according to the DMP survey.....	12
Table 3:	Data formats produced in PIISA according to the DMP survey.....	12

Keywords

Data Management, FAIR-principles, Data identification, Data security, EU Data Act, Metadata

Abbreviations and acronyms

Acronym	Description
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR	Findable, Accessible, Interoperable, Reusable
FMI	Finnish Meteorological Institute
KPI	Key Performance Indicators
NBS	Nature Based Solutions
PID	Persistent Identifier
WP	Work Package

Glossary

Data identification. Identifier for the data set to be produced.

Data set description. Descriptions of the data that are being generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

Data Acquisition. Explanation of how the data will be obtained or collected, including the methods, instruments, and protocols to be used. This can include details on data sources, sampling techniques, and data generation processes.

Metadata. Metadata is providing information about one or more aspects of the data; it is used to summarize basic information about data that can make tracking and working with specific data easier. “Data about the data”.

Data sharing. Description of how data is being shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access is being widely open or restricted to specific groups. In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Archiving and preservation (including storage and backup). Description of the procedures that are being put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

Data Ownership. Identification of the individuals or entities that own the data and any associated intellectual property rights. Clarification on how ownership may affect data sharing and access.

Data Quality Assurance. Description of the measures and procedures in place to ensure data accuracy, consistency, and reliability. This can encompass data validation, error-checking, and quality control processes.

Data Security. Explanation of data security measures to protect sensitive or confidential information. This may involve encryption, access controls, and data anonymization techniques.

Data Lifecycle. A comprehensive overview of the data's lifecycle, including its creation, storage, usage, sharing, and eventual disposal or archiving. This helps ensure proper management at each stage.

Data Citation. Guidelines on how to cite the dataset in publications or other research outputs. This should include information such as dataset title, creator(s), publication date, and a persistent identifier (e.g., DOI).

Data Documentation. Detailed documentation about the dataset, including data dictionaries, codebooks, and other explanatory materials that aid in understanding the data's structure, variables, and context.

Data Licensing. If applicable, information about the licensing terms under which the data can be used, shared, and modified. Specify whether open licenses (e.g., Creative Commons) or custom licenses are being used.

Data Ethics and Compliance. Detailed information on how the data management plan adheres to ethical guidelines, legal requirements, and institutional policies, particularly when dealing with sensitive data or human subjects.

Data Disposal. Procedures for securely disposing of data when it is no longer needed, including data destruction methods and compliance with relevant data protection regulations.

Data Monitoring and Auditing. Explanation of how data management activities will be monitored and audited to ensure compliance with the data management plan and to address any issues or deviations promptly.

1 Introduction

PIISA stands for “Piloting Innovative Insurance Solutions for Adaptation” and is a 3-year Research and Innovation Action funded under the Horizon Europe program within the call “HORIZON-MISS-2022-CLIMA-01”. The partnership brings together 12 organisations from 5 European countries, and will co-develop climate resilient insurance portfolios, as well as develop solutions for sharing losses and climate risk data. PIISA aims to develop and deploy a range of insurance innovations to cover at least 50% of losses attributable to climate change effects in Europe. The project hopes to support households, firms, and public authorities to set up adaptation and create adaptation promoting conditions. The PIISA project is described in more detail on the project website which is available at <https://piisa-project.eu/index.html>.

The purpose of the data management plan (DMP) is to document how the data generated by the PIISA project are handled during and after the project. It describes the basic principles for data management within the project. This includes standards for documentation at discovery and data levels as well as data sharing and preservation including life cycle management of datasets. The data management plan includes information on what type of data is being generated in the PIISA project, what standards are being used, how this data is being exploited and made accessible for verification and re-use, and how this data is being curated and preserved.

This document provides an updated outline of the PIISA project policies for data management. The updated DMP status report for the Mid-Term of PIISA (D5.11) includes **up-to-date information about the data produced and used** by partners as well as new, more robust, **instructions for naming the datasets and files** and instructions for responsible and transparent **use of AI in research**. Key updates can be found in chapters 2, 3.1 and 8. DMP is a living document and following updates will be made when deemed necessary. The next scheduled revision is made for the status report at the end of the PIISA project (D5.12) which is currently scheduled for April 2026 (M35). The Horizon Europe Model Grant Agreement requires that a data management plan is established and regularly updated.

The first DMP and the updated DMP is made based on the EU Grants Data Management Template since it is recommended for Horizon Europe beneficiaries ([HORIZON EUROPE Data-Management-Plan-Template.pdf \(openaire.eu\)](#)). In completing the template sections, we address the requirements for research data management of Horizon Europe as described in article 17 and analysed in the Annotated Grant Agreement, article 17. PIISA is following the principles outlined by the Open Research Data Pilot and The FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al. 2016). The data management plan is based on the OpenAIRE guidelines ([How to create a Data Management Plan \(openaire.eu\)](#)).

2 Data Summary

To achieve PIISA's aims and objectives, the project will process and produce information and data about climate risks in the financial sector, the impact of climate change on the insurance sector, and user preferences for different insurance services based on interviews and surveys. The PIISA project is addressing datasets, for example, observations of conditions in the European region, to support stakeholders involved in climate adaptation and resilience efforts. In addition, the project is collecting information on activities and initiatives that are generating valuable climate data, such as surveys, climate risk assessments, and loss records, among others. Updated WP-specific data descriptions are presented below. The updated description is provided by each WP Leader with help from partners.

WP1 reviews the current state of the market for climate risk insurance and analyses the challenges and opportunities to overcome market barriers through innovation in the design and delivery of insurance products and services. WP1 will produce literature reviews to identify gaps between the users' needs and the insurance sector in a changing climate, as well as the established practices for actuarial modelling. The objectives of WP1 draw upon data from the EIOPA (European Insurance and Occupational Pensions Authority) Dashboard on the insurance protection gap for natural catastrophes (https://www.eiopa.europa.eu/tools-and-data_en) and also, from the findings reported in scientific, grey and white literature. WP1 will share, in open access repositories (such as Zotero), the data extraction matrices from the literature reviews of D1.1 *Role and potential of insurance in accelerating climate adaptation in Europe* and D1.2 *Advancements in actuarial risk modelling*, together with the code to ensure the reproducibility of the analysis, upon publication in peer-reviewed journals of the scientific papers based on the deliverables. In addition, results from the Politecnico di Milano (Polimi) [survey](#) will be utilised in the D1.4 *Focused market reviews in WP3 pilot sectors/areas*.

WP2 is dedicated to co-design, co-develop and co-produce climate services. In the context of the PIISA project, WP2 transforms the needs of insurance for creating innovative products, as clarified in WP1, into enhanced data and climate indicators e.g., high-resolution gridded data and climate indices for parametric insurance, return period services, microclimatic solutions, etc. These solutions are developed in partnership with WP3 through a collaborative process.

WP2 will identify and develop a set of climatic indicators, considering the role of extreme events and compound extreme events. These indicators will be obtained based on different sources of climate information: from weather predictions, through seasonal predictions to climate change projections. A significant portion of the data used will be sourced from the open datasets provided by the Copernicus Climate Change Service (<https://climate.copernicus.eu>), ensuring the comprehensiveness and reliability of our climate data inputs. In addition to the Copernicus datasets, WP2 will also incorporate data and projections from the Coupled Model Intercomparison Project from CMCC ([CMIP6 - Coupled Model Intercomparison Project - CMCC](#)).

WP2 will propose specific methodologies to integrate different indices associated with specific risks in order to quantify the composite physical risk. This information can be used either for pricing or to produce hazard maps for underwriting policies, to ensure the profitability of the products, but also in terms of prevention and awareness of end-users. WP2 is also responsible for integrating novel piloted open-source risk indicators into the [Risk Data Hub](#) (KPI 6). The following table is an initial outline of the new risk indicators exported to the Risk Data Hub:

Dataset	Category	Short description	Format	Source	Area
Hail hazard		Hail hazard indicator based on ERA5 data, using a machine learning approach to select meteorological variables	Gridded netcdf files	WP 2.2	
Era5_SWI	Climate	index to map windstorm footprints and predict severe windstorms events computed with the ERA5 reanalysis	NetCDF	Amigo	FR, AT
Cerra_SWI	Climate	index to map windstorm footprints and predict severe windstorms events computed with the CERRA reanalysis	Netcdf	Amigo	FR, AT
Climate Dryness Index	Climate	Compound index to assess the combined occurrence of drought and heatwaves	Netcdf	Amigo	Lyon
Temperature	Climate		netcdf	Era5-Land	
Relative Humidity	Climate		netcdf	Era5-Land	
Wind speed	Climate		netcdf	Era5-Land	
Wind direction	Climate		netcdf	Era5-Land	
Wind damage	GG		shapefile	FORWIND	
Topography	tpi		tiff	SRTM-mTPI	
Topography	twi		tiff	dtm_twi_merit	
Topography	wei			SRTM-GL1 (DEM) X Era5 land	
Topography	soil		tiff	soilgrids-isric	
Topography	Slope		tiff	SRTM-GL1 (DEM)	
Topography	Elevation		tiff	SRTM-GL1 (DEM)	
Topography	Aspect		tiff	SRTM-GL1 (DEM)	
Vegetation index	Land cover		tiff	Copernicus Land services	
Vegetation index	Tree cover		tiff	Copernicus Land services	
Vegetation index	Tree canopy high		tiff	GEDI Lidar	
Vegetation index	tree density		tiff	10.1038/nature14967	
Vegetation index	canopy height (2020)		tiff	ETH Global Sentinel-2 10m Canopy Height (2020)	
Vegetation index	canopy height (2005)		tiff	GLAD	
Vegetation index	tree cover (2000)			Hansen, Goabal forest change 2022	
Vegetation index	canopy cover (2005)		tiff	GFCC	
Vegetation index	forest management		tiff		
Wildfire	Historical wildfires		netcdf	MODIS (MCD64A1)	

Table 1: Initial outline of the novel open-source risk indicators integrated into the Risk Data Hub

WP3 focuses on the development of new innovative concepts, advanced products, and services through piloting. The following themes will be explored in the pilots:

- Pilot 1 - Green Roof Insurances
- Pilot 2 - Addressing soil stability risks for homeowner insurance holders
- Pilot 3 - Insurance Services for Agriculture
- Pilot 4 - Forest Insurances against selected biotic and abiotic risks
- Pilot 5 - Wildfire insurance enhancing adaptive actions

Within WP3, partners are engaged in the generation and analysis of various types of data that serve the development of insurance pilots. In general, WP3 will produce (end of loop) **surveys, conduct interviews and workshops** to find out customers' preferences related to different designs of insurance including parametric insurance and stakeholders' views on adverse weather phenomena and how to measure them (e.g. LIKERT scale opinion polls and embedded choice experiments). WP3 team will identify and process the required data input for index-based insurance products and forecast-based finance mechanisms. For example, one of the objectives is to evaluate a strategy of an insurance company (Interpolis) that is interested in stimulating the adoption of nature-based solutions (NBS) by its policyholders. The outcomes of the choice experiment that will value the relevant (co-)benefits of NBS will feed into a cost-benefit analysis. This data may be especially useful for insurance companies, researchers and policymakers.

To develop, improve and measure pilots, WP3 also uses a selection of open weather, satellite, yield and land use data from public sources such as MODIS ([MODIS | Earthdata \(nasa.gov\)](#) for the burned area), Sentinel ([Sentinel Online - ESA - Sentinel Online](#) for burn severity index), Fire Weather Index ([Fire Weather Index | Copernicus](#)), Land use data ([Copernicus Land Monitoring Service](#)) and potentially other climate data from Copernicus, Nasa or Global Forest Watch (<https://www.globalforestwatch.org/>). Also, other weather datasets (ERA5 land: [ERA5-Land | ECMWF](#)), GIS data (Sentinel 1: polarization HH, HV, VV & Sentinel 2 level 2A (hyperspectral) to extract features

for modelling), For comparison ground truth data sources will be explored such as FORWIND (European dataset for wind throw damages <https://doi.org/10.6084/m9.figshare.9555008>). In the case of Task 3.4.2, we are planning to use modelled and observed data; modelled data will be built from simulations of wildfire spread in different scenarios to assess the impact of risk mitigation actions in decreasing the wildfire risk. Observed data is going to be retrieved from **open-source satellites** and be used in the parametric insurance framework to trigger the pay-out of a wildfire event. In addition, we will explore data facilities and simulations of the **Destination Earth Climate Digital Twins (DestinE)** to be utilised in wildfire risk assessments.

WP4 is dedicated to communicating and disseminating the activities and results generated in PIISA to target audience as well as to promoting a dialogue with key stakeholders and exploiting the outcomes. Activities and data created in WP4 serve all other work packages as well as PIISAs visibility and dissemination of the project outcomes. To enhance dissemination, WP4 will be working on the public communication aspects of the project, including producing content for the newsletter and creating continuous social media engagement (website, X and LinkedIn). Data on PIISA website visitors as well as social media account followers, likes and shares will be collected. The data will be used to follow the KPIs and to evaluate the visibility of the project. Content for the newsletter and social media platforms is collected from the events, case studies, external engagement tasks and other research done within PIISA. As part of these tasks, WP4 will create a registry for stakeholders, events and newsletter subscribers:

- **Stakeholder registry** includes data on stakeholders: Name, organization, email, country, stakeholder type, level of collaboration, geographic coverage, potential role and WPs associated, who added the information and date of adding the information. The registry is in the PIISA Teams WP4 folder. It will be used to plan stakeholder collaboration activities especially in WP4, but also in other work packages.
- **Events registry** includes data on the (PIISA) events, the partners attended and attending. The data included: dates, event name, location, description, target audience, strategic level, size of the event, names of attendees from PIISA, poster or oral presentation, event web page. The registry is in the PIISA Teams WP4 folder. It will be used to follow the KPI for event attendance, to coordinate event attendance and to coordinate and report PIISA events.
- **Newsletter subscribers registry** will include the contact details of the persons that have subscribed to the PIISA newsletter. It will be used to send out the newsletter as well as to follow the KPI. The dataset on subscribers of the newsletter is only accessed by LGI Consulting.

Moreover, WP4 has collected data on expectations and will collect feedback from pilots in each loop. The questions are specified in D4.10 *Survey for mapping Loop feedback for collecting iteration needs – Version 1*. The surveys are sent via the **EUSurvey** platform. The first surveys on expectations have been conducted. The surveys included information on the respondents' role in the pilot and their expectations for the pilot. The results can be found in the Teams folder WP3 under "Pilot expectation and end of the Loop surveys" folder. WP4 will also conduct interviews with partners to co-develop their plans for exploiting and replicating pilot results. To support this, Task 4.5 we will set up an "IP Repository" which will be an Excel spreadsheet that documents the intellectual property of all partners in the project and their expected routes to exploit their results.

WP5 is responsible for the administration and overall coordination of the project. WP5 is creating documents and guides for managing the project to ensure efficient, timely and seamless project coordination and collaboration between partners. We aim to help with organising, guiding and managing all the processes and information shared in the PIISA project. WP5 has set up the Teams page and will produce deliverables according to Gantt. The deliverables of WP5 are mainly intended to support the administration and management of work, timeline, risks, ethics, equality, collaboration and data within PIISA. Some deliverables are not intended to be published outside the PIISA project because they are mainly relevant for internal use.

Data utility. Data and information generated by the PIISA project can have significant utility and relevance to various stakeholders and organizations outside the project, including insurance companies, government agencies, environmental organizations, academia, businesses, investors, local communities, and the general public. New innovative solutions in the insurance sector are contributing to a more informed and coordinated approach to addressing climate change and its associated risks. Sharing this data can foster collaboration and support evidence-based decision-making in the face of climate adaptation and resilience. Here are some potential beneficiaries and ways in which the data can be useful and valuable:

- **Insurance Companies:** Insurance companies and underwriters can benefit from the data to improve their risk assessment models and develop climate-resilient insurance products. This data can help them better understand and price climate-related risks.
- **Government Agencies:** National and regional government agencies responsible for climate adaptation and disaster management can use the data to inform policy decisions, allocate resources, and develop strategies for mitigating the impacts of climate change.
- **Environmental Organizations:** Environmental NGOs and research institutions can leverage the data for climate research, advocacy, and public awareness campaigns. It can provide valuable insights into the real-world impacts of climate change.
- **Academic and Research Institutions:** Researchers and scholars working in climate science, economics, and related fields can use the data for academic studies, modelling, and to advance our understanding of climate-related risks and adaptation strategies.
- **Businesses and Industry Sectors:** Companies in various sectors, such as agriculture, construction, and real estate, can benefit from the data to assess and manage climate risks in their operations, supply chains, and infrastructure investments.
- **Investors and Financial Institutions:** Investors and financial institutions can use the data to evaluate climate-related risks associated with their portfolios and investments. It can help them make more informed decisions and allocate resources accordingly.
- **Local Communities:** The data can be valuable for local communities, helping them understand their vulnerability to climate change and make informed decisions about adaptation measures.
- **General Public:** Providing access to some of the data to the general public can raise awareness and education about the real-world impacts of climate change, potentially leading to more climate-conscious behaviours and choices.

Summary of the data types planned to be produced and used in the PIISA project is provided in table 2 below. Table also includes the percentages for each data type based on their occurrence out of the total (57) data types reported in the DMP survey. In the DMP survey it was also noted that 62 % of respondents stated that they're using and/or producing some dynamic datasets which means that their data will be regularly updated.

Type	No.	Amount%
Numerical	10	17.5 %
Textual	6	10.5 %
Graphical	6	10.5 %
Visual	2	3.5 %
Born-digital	4	7 %
Digitized	0	0 %
Quantitative	7	12.3 %
Qualitative	4	7 %
Raw	4	7 %
Cleaned	6	10.5 %
Simulated	5	8.8 %
Processed	7	12.3 %

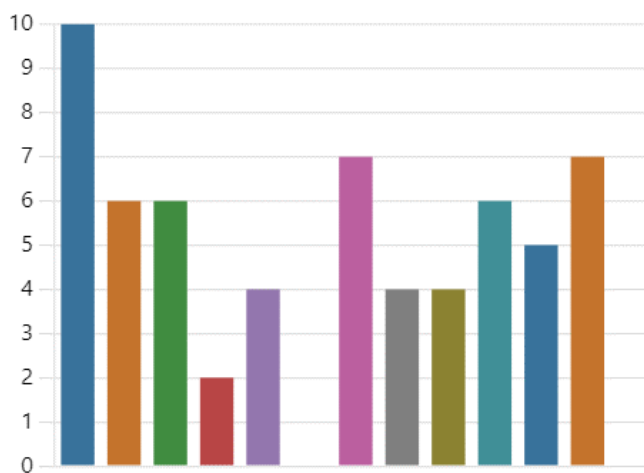


Table 2: Data types produced in PIISA according to the DMP survey

Summary of the data formats planned to be produced and used in the PIISA project as well as the amount and percentages for each file format based on the survey is provided below. Over 50 % of the reported data formats used or produced are csv, xlsx or netCDF files. In “other” box, use of Word (doc) and PowerPoint (pptx) were mentioned. The percentages represent the occurrence of each file format out of the total count of 41 answers.

Format	No.	Amount%
csv	9	22 %
ascii	1	2.4 %
xlsx	7	17.1 %
json	2	4.9 %
geojson	2	4.9 %
netCDF	5	12.2 %
zarr	1	2.4 %
grib	0	0 %
pdf	4	9.8 %
shp	4	9.8 %
tiff	4	9.8 %
geotiff	4	9.8 %
kml	2	4.9 %
gml	0	0 %
other	2	4.9 %

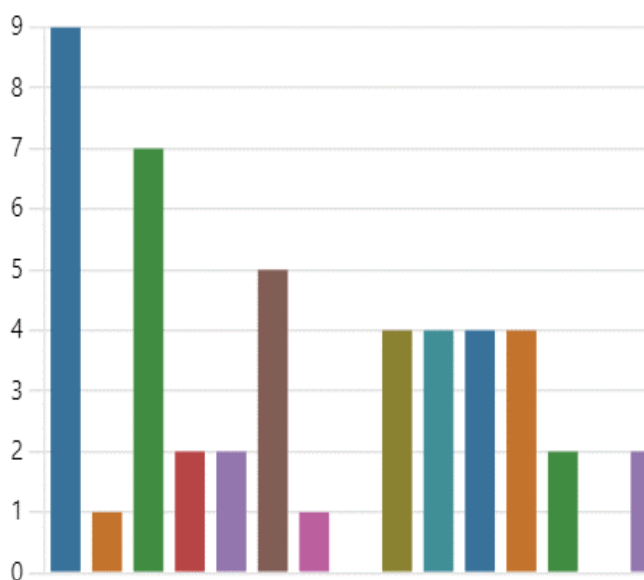


Table 3: Data formats produced in PIISA according to the DMP survey

Size of the data. Based on the provided information from partners in the first round of DMP survey, we can make first indications for the size of the data to be generated or used in the PIISA project. To estimate the total size of the data produced, we can multiple the number of answers to the assigned approximate values:

3 x Less than 1Gb: "Less than 1Gb" can be considered as 0.5 GB on average for each answer.

1 x 1–10Gb: "1–10Gb" can be considered as 5 GB on average for the one occurrence.

2 x 10Gb–500 Gb: "10Gb–500Gb" can be considered as 250 GB on average for each answer.

2 x 500Gb–1Tb: "500Gb–1Tb" can be considered as 750 GB on average for each occurrence.

6 x I'm not sure yet: "I'm not sure yet" doesn't provide a specific data size = unknown

With these estimates, the total estimated data size produced is:

$$\begin{aligned}\text{Total Data Size} &= (3 \times 0.5 \text{ GB}) + (1 \times 5 \text{ GB}) + (2 \times 250 \text{ GB}) + (2 \times 750 \text{ GB}) + \text{unknown} \\ &= 1.5 \text{ GB} + 5 \text{ GB} + 500 \text{ GB} + 1500 \text{ GB} + \text{unknown} \\ &= 2000 \text{ GB (or approximately 2 Terabytes, TB)} + \text{unknown}\end{aligned}$$

We would like to stress that this is only a preliminary estimate and probably well underestimated as 6/14 respondents stated they don't know yet. Because of this, the calculation formula presented above is more important, as it will be used to estimate the size of the data produced as the project continues.

3 FAIR Data

Horizon Europe emphasizes the management of data and other outputs in accordance with the FAIR principles, which means making data Findable, Accessible, Interoperable, and Reusable.

3.1 Making Data Findable

In the context of the PIISA project, ensuring data findability is paramount, both during and after the project's duration. To achieve this, we employ the use of Persistent Identifier (PID) or Digital Object Identifier (DOI¹) for the datasets and deliverables generated. PIDs and DOIs provide a unique and persistent reference for data, making them easily discoverable even beyond the project's scope. For scientific articles and public deliverables, DOI is created by uploading the article to PIISA's Zenodo community. Data and metadata should be easy to find for both humans and computers. To facilitate efficient data discovery, findability, and reuse, we will provide essential metadata alongside the datasets. This metadata will include codebooks, readme files, and netCDF files with metadata information in the attributes section. It will also comprehensively describe data sources, characteristics, and extraction methods. All public datasets, deliverables, info cards and presentations produced in PIISA will be uploaded to [Zenodo / PIISA project](#) and receive a DOI number.

3.1.1 Naming of the files

To ensure document and data control, each document and dataset shall be uniquely identifiable and associated with a unique document name. The tag PIISA, and the deliverable number and name must be used in the deliverable filename. The data filename for publication shall be in the format: **PIISA_Deliverable number_Name of the data**. With WP4 and WP5 living documents also the version number at the end is suggested (see example).

Example: **PIISA_D5.11_ Project Data Management Plan_Version 2**

When re-using open data e.g. from the Climate Data Store (<https://cds.climate.copernicus.eu/#!/home>) or the JRC Risk Data Hub (<https://drmkc.jrc.ec.europa.eu/risk-data-hub#/>) the data (usually) already has a PID. Also, the metadata is already provided on the websites enabling the downloading of this data. This kind of observed data is going to be referenced as documented in the original source. Whenever reusing datasets, a reference to the original data source must be included. If the data produced is sensitive or not meant for sharing, metadata isn't provided alongside the data. For example, documents produced in WP5 won't be made available for harvesting by web browsers and search engines. Also, the feedback from stakeholders and some user surveys collected in WP3 and WP4 are not for widespread distribution and dissemination.

Note: The data used and produced by the project should have a use beyond the project. To ensure that the data don't get abandoned if the project doesn't have a future after the EU-funded phase, PIISA suggests that the PID (or DOI) for the generated data is provided by each data producer since it's the best way to prevent this. Another option is to use PIISA's Zenodo repository to indicate the unique DOI number for future reference and preservation.

3.2 Making Data Accessible

Once the user finds the required data, they need to know how the data can be accessed, possibly including authentication and authorization. This chapter is sectioned for information about the repository, data and metadata.

3.2.1 Repository

The repository will ensure that data is stored safely and securely and in full compliance with European Union data protection regulations. The public data produced in the PIISA project will be deposited in a dedicated PIISA project community in Zenodo ([Zenodo / PIISA project](#)). Zenodo is a trusted and suitable repository for the PIISA project for the following reasons:

- **Open Access:** Zenodo is an open-access platform, which means that anyone can access and download the data you deposit without any restrictions or paywalls. This is important for promoting the sharing of research data and making it available to the wider scientific community and the public.
- **Long-Term Preservation:** Zenodo is operated by CERN, the European Organization for Nuclear Research, and supported by the European Commission. This backing ensures the long-term preservation and sustainability of the platform. It's designed to store and provide access to research outputs for the long haul.
- **Interdisciplinary:** Zenodo is not limited to a specific field or discipline. It accepts and hosts research data and other research outputs from a wide range of fields, making it a versatile option for researchers in various domains.
- **Integration with ORCID:** Zenodo allows you to link your ORCID iD to your Zenodo account, which can help increase the visibility and impact of your research outputs.
- **Version Control:** Zenodo provides version control for your data, ensuring that you can update or revise your datasets as needed while keeping previous versions accessible.
- **Metadata and DOI:** Zenodo assigns a Digital Object Identifier (DOI) to each deposited dataset, which makes it easy to cite and reference your data. Additionally, you can provide detailed metadata for your datasets, enhancing discoverability and usability.
- **Collaboration and Licensing:** Zenodo allows for collaborative data management and the inclusion of licenses, ensuring that you can define how your data can be used by others.
- **Integration with Other Services:** Zenodo is integrated with other research platforms and services, such as GitHub, making it easier to link your code and data together. This can enhance the transparency and reproducibility of your research.
- **Free to Use:** Zenodo is free to use, which is especially important for researchers and organizations with limited budgets.

The use of Zenodo also aligns with the new European Commission's regulation 'Data Governance Act' which entered into application on 24th of September 2023. The Data Governance Act aims to create a safe environment for sharing data across sectors and Member States to benefit society and the economy. The Data Governance Act also allows novel data intermediaries to act as trustworthy actors in the data economy. To read more information, please visit: <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>.



3.2.2 Data

All data that can be publicly shared will be made openly available and stored in PIISA's [website](#) and [Zenodo community](#). Datasets required for project purposes but not currently openly accessible, even though the data license permits open access, will be temporarily cached by FMI during the project's duration. These datasets will be made available to both internal and external potential users. Selected datasets will be preserved for future reference by FMI, and discovery metadata, as well as online access will be provided for these datasets. FMI complies with The EU data legislation.

In case certain datasets cannot be shared or need to be shared under restricted access conditions (they contain personal information and/or are collected for PIISA internal use only), they are handled accordingly. Access to documents and guidelines meant for internal use within PIISA will be facilitated through Teams and SharePoint. Sensitive data stemming from surveys and interviews will be securely stored in a non-open access repository.

For data subject to usage restrictions, PIISA will clearly outline how access to the data will be provided, both during and after the project. Legal and contractual reasons for intentional restrictions for data sharing will be clearly separated and stated regarding the sensitive data. For instance, data collected by WP4 (e.g. feedback from stakeholders, information of newsletter subscribers etc.) is intended solely for internal project use and will not be made publicly available. Further details about the data shareability will be updated during the project.

Note: in multi-beneficiary projects, it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

3.2.3 Metadata

Within PIISA, newly produced data will be provided along with metadata and it will be made openly available and licenced under a public domain dedication CC0², as per the Grant Agreement. PIISA will support partners in providing metadata which contains information to enable the user to access the data. The metadata will be available for re-use through the PIISA website where the annual reports and deliverables will be stored and will be published on [CORDIS](#)³. Metadata will be made available for harvesting by web browsers and search engines. If documentation or a reference to software is required to access or read the data, relevant software (e.g. open source) is included.

For how long the data and metadata will remain available and findable, will be discussed at the end of the project. The aim is to preserve the data in Zenodo as long as it might be useful, valid or valuable for research use and/or the general public. Metadata will remain available after data is no longer available.

2) CC0 (Creative Commons Zero) is a public domain dedication tool that removes legal doubt about whether researchers can use data in their projects. For more information see [CC0 - Creative Commons](#).

3) CORDIS (Community Research and Development Information Service) is the European Commission's primary source of information on research and innovation activities funded by the European Union (EU). CORDIS provides a comprehensive online platform that offers access to a wealth of information related to EU-funded research projects, programs, and initiatives.

3.3 Making Data Interoperable

In research, the data usually need to be integrated with other data. To support the reuse of data, combining information and datasets needs to be easy. Best way to ensure data compatibility and comprehensibility for other researchers and systems is to provide the data in a standardized and well-documented manner. In most cases, this means that the APIs/interfaces must be machine-readable, and as automated as possible so that formats are interoperable. Interoperability of data makes it more easily accessible for analysis and research objectives.

To ensure interoperability, PIISA requires that data generated in the project are published in a repository that allows machine access using standard interfaces and information objects for discovery metadata. Zenodo provides machine access to its data and metadata through standard interfaces and information objects. Zenodo supports the use of well-established protocols and formats for data discovery, making it accessible for automated access, searching and retrieval. Some of the key features and standards used for machine access in Zenodo:

- **OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting):** Zenodo supports OAI-PMH, which is a widely recognized protocol for harvesting metadata from repositories. This allows automated systems and search engines to discover and harvest metadata from Zenodo for indexing and discovery purposes.
- **API (Application Programming Interface):** Zenodo offers a RESTful API that enables programmatic access to its data and metadata. Researchers and developers can use this API to search for content, upload data, and retrieve metadata. The API is well-documented, making it easier to integrate with other systems.
- **Standard Metadata Formats:** Zenodo uses standard metadata formats such as Dublin Core, DataCite, and others, ensuring compatibility with established metadata standards used in the scholarly and research community.
- **DOIs (Digital Object Identifiers):** Zenodo assigns DOIs to each dataset and research output. DOIs are a recognized and persistent means of identifying and citing digital objects, which enhances the discoverability and citation of the data.
- **Creative Commons Licensing:** Zenodo allows researchers to specify Creative Commons licenses for their data, making it clear how the data can be used and reused by others. This licensing information is embedded in the metadata.
- **Structured Metadata:** Zenodo provides structured and comprehensive metadata for deposited content. This metadata includes information about the dataset, contributors, funding sources, and more, making it easier for machines to understand and process.

In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing. This part is particularly relevant for machine action-ability, the ability of machines to “understand” information that is consistently structured. Machine-readable metadata is essential for the automatic discovery of datasets and services, so this is an essential component of the FAIRification process. The metadata formats will follow the convention of the hosting research data repository. Further information on metadata formats is set out to be reviewed in the next DMP update.

Vocabularies, standards and formats. Data providers within PIISA stated the following:

- All data generated in the pilot is going to be described and documented in the same format and uploaded in the same public repository. The full description and mapping of all datasets is going to be included in the repository.
- The idea is to create a data cube, time*long*lat* (nbr of features + prediction). It will be a xarray dataset.

- Metadata vocabularies will be used for our social media communications such as with the use of the decided hashtags and keywords for the project.
- All the data will be available through the NetCDF format, which follows the CF Conventions about metadata standards.

<https://cfconventions.org/Data/cf-conventions/cf-conventions-1.10/cf-conventions.html>

Note: the use of schema and controlled vocabularies is not compulsory in Horizon Europe, although it is highly recommended - if partners have created brand new schemas or controlled vocabularies for the project, please indicate this to the data manager (kaisa.juhanko@fmi.fi).

Methodologies data providers are following for data creation and management:

- Collecting data through API and Institute for Environmental Studies (IVM) VU University Amsterdam
- Collecting input data from Copernicus (ERA 5 and Sentinel 1 & 2), merging with ground truth data and building an ML algorithm on top of that
- Using climate analysis methods (bias correction, downscaling and methods to identify extreme events) that have been developed in other H2020 projects on climate services
- Collecting ground-based information about the risk mitigation measures that are planned to be implemented in the regions of the Pilot
- Processing remote sensing data
- Validating and storing data outputs in a public repository
- Handling data through Python, QGIS, excel and standard statistical software
- Implementing scenario simulations
- Integrating different types of data, including structured and unstructured data
- Ensuring high data availability and disaster recovery
- Governing how data is accessed and used by people and applications
- Data transfers between researchers will be made with end-to-end-encryption, using services like SURFfilesender (<https://www.surffilesender.nl>).
- Customer survey and some customer data will be collected from LocalTapiola's CRM

3.4 Making Data Reusable

The end goal of FAIR is to optimize the reuse of data. To achieve this, data and metadata should be standardized and well-described so that they can be found and replicated and/or combined in different settings. To ensure this, PIISA partners will produce documentation to validate data analysis and facilitate data re-use by readme files with information on methodology, codebooks and the netCDF files. Also keywords and hashtags will be provided in the metadata to optimize discovery and reuse. The provenance of the data will be thoroughly documented using the appropriate standards.

Licensing: Data need to be accompanied by a clear, machine-readable license. You can use tools such as <https://chooser-beta.creativecommons.org/> (for data and publications) or <http://ufal.github.io/public-license-selector/> (for software) to choose a suitable license. Permissive licences (such as CC0 or CC BY) are preferred in Horizon Europe, wherever possible. However, data should be 'as open as possible, as closed as necessary': if there are intellectual property rights on data, as well as sensitive or classified data they must be kept close (e.g., through access control mechanisms) and the reason must be clearly explained.

In the PIISAs DMP survey, partners indicated to release the data under the MIT, CC0 (1.0) and CC BY (4.0) licenses. Some WPs may have some restrictions in sharing the client data, but it's not sure yet. Other WP stated that if required, they will make the data available open via github. The IP repository will be closed to protect the interests of partners.

All data that can be made freely available will be published in Zenodo to permit the widest re-use possible. The research outputs will be also available for reuse through the PIISA website where the annual reports and deliverables will be published. Datasets and documents uploaded and stored in Zenodo repository will be openly accessible to the public. Potential users are expected to adhere to the Terms of Use of the repository. By contacting PIISA management via the website, a third party can request access to datasets with restrictions on use. PIISA will ensure that all data requests will be subject to scrutiny under GDPR and that no data which can identify individual sources or technology will be released. To support the reuse of data produced in PIISA (in particular after the end of the project) we need to further elaborate on our potential end users and who needs to be able to read the data.

Data quality assurance processes for PIISA include:

- Data Validation: Implement automated checks and validation rules to verify data accuracy and consistency during data entry.
- Data Cleaning: Identify and rectify data errors, inconsistencies, and outliers to ensure data integrity.
- Data Standardization: Establish and enforce standardized data formats, units, and naming conventions to improve data consistency.
- Data Documentation: Maintain comprehensive metadata and documentation to describe the data sources, transformations, and any relevant assumptions.
- Data Auditing: Conduct periodic data audits to identify and resolve data quality issues.
- Data Profiling: Profile the data to understand its characteristics, distribution, and patterns, helping to identify potential data quality problems as well as ensuring all relevant data quality assurance processes.

In the DMP survey partners stated the following:

- Data will be cleaned thoroughly based on standards used in our field for assessing, e.g., protest responses, random responses, etc.
- Data is tested and verified by at least 2 independent engineers.
- Monthly updates
- Data is delivered from national agencies, and we perform initial data checks to remove outliers and ensure data quality.
- Data is peer-reviewed by the data and science teams and by the local partner.
- Every document gets comments from three people inside PIISA to avoid any errors.
- We check regularly the data submitted by partners.
- It goes through an internal review process, including technical review.
- We validate incoming data to ensure that it matches specified criteria, such as value ranges for meteorological values and geographic position. Our data cleaning process checks the data and identifies and corrects discrepancies, missing values, and outliers.

4 Other Data Outputs

To support the research in the PIISA project and especially the work in the WP3 Green Roof pilot, **Politecnico di Milano (Polimi)** conducts a survey targeting a sample of adult citizens from six European countries (Finland, France, Germany, Italy, the Netherlands, and Spain) to assess the **level of awareness** and **adoption** of green roofs as well as the **incentives to wider adoption**. The survey collects data about:

- **Demographic data:** country, age, gender, education level, wealth level, living situation, housing situation, financial and insurance expertise
- **Green-roof data:** knowledge, adoption, willingness (with and without economic expense), incentives
- **Other data:** climate change awareness, insurance knowledge and insurance situation

The survey is funded by Polimi's own resources and thus the raw data is not openly available for all partners (see more in [Allocation of Resources](#)). The results of the survey will be available for collaborating partners and findings will be used and presented, for example, in the D1.4 *Focused market reviews in WP3 pilot sectors/areas*, as well as in all of the pilots.

The publication committee suggested establishing a **shared literature repository on Zotero**. The aim was that this repository would serve as a centralized resource for collecting, organizing, and sharing essential references and literature among project members. By consolidating resources in one accessible platform, partners could enhance collaboration, streamline access to key materials, and improve the efficiency of research and project development. Establishing a shared literature repository was discussed with all partners in the fourth General Assembly (13.-15.11.2024) of PIISA and it was decided that **due to GDPR and copyright matters the shared literature repository won't be enabled**.

In addition to the management of data, PIISA will consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.). In case, additional information is generated or re-used, PIISA will consider which of the questions pertaining to FAIR are applicable to the management of these research outputs. PIISA aims to provide sufficient detail on how other research outputs will be managed and shared or made available for re-use in accordance with FAIR principles.

5 Allocation of Resources

The costs of making scientific publications, maintaining a project website and open-access data repositories are included within the PIISA budget as eligible costs. The activities related to making the research outputs open-access and FAIR are anticipated to be covered within the allocated budget for each partner. The costs related to open access to research data are eligible as part of the PIISA Horizon 2020 grant. Zenodo is free to use repository by CERN which is one reason PIISA Consortium decided to use Zenodo. The long-term preservation strategy and resources, how data will be kept beyond the project and how long as well as the associated costs and potential value, will be discussed at the last Consortium's General Assembly (GA) meeting at the end of the project.

Overall responsibility for the Data Management Plan lies with WP5 which has allocated resources for cataloguing, serving and preserving data within the project period. Project-leading affirmation **Finnish Meteorological Institute and WP5 Project Management** have appointed the Data Manager (Kaisa Juhanko), in collaboration with consortium partners, to manage the data generated during the project. Any costs associated with sharing data with external partners or stakeholders will be clearly identified and discussed with the involved parties to ensure transparency and alignment. WP5 will be responsible for disseminating this DMP to all project partners and each project partner will be responsible for managing their data, metadata, and ensuring their data meets the quality standard set out in this DMP.

Tyrsky Consulting (Tyrsky) have been appointed as WP4 *Communication and Dissemination* leader and in collaboration with consortium partners, they will oversee the identification of which datasets will be disseminated and the most appropriate means of disseminating this data. This has been conducted and the defined strategy is presented in Deliverable 4.1. *Project communication, dissemination and stakeholder engagement plan* (CDSEP).

Politecnico di Milano's (Polimi's) independent survey to assess the uptake of green roofs in six European countries **is fully funded by Polimi's own resources** (not by PIISA) and thus, its initial raw data is unavailable to other partners. However, the results of the survey are open and accessible to all collaborating partners, and the findings will be used in PIISA's publications. The Green Roof pilot surveys funded by PIISA:

- **Vrije Universiteit Amsterdam (VU):** Green Roof survey with Dutch households (co-benefits of green roofs, barriers to adoption, the effectiveness of insurance and other incentives).
- **STICHTING Climate Adaptation Service (CAS):** Nature Based Solutions (NBS) survey with Dutch insurance companies.

6 Data Security

PIISA requires partners to comply with the EU data policy and the General Data Protection Regulation (GDPR) ([Data protection under GDPR - Your Europe \(europa.eu\)](#)). This includes guidelines on data recovery as well as the secure storage/archiving and transfer of sensitive data. By following these guidelines, we can ensure that data collection within the PIISA project is conducted in a manner that respects EU data policy, protects the rights of participants, and maintains data security and compliance. These practices are informed by legal requirements and widely accepted data privacy and security standards in the European Union.

For the duration of the project, datasets will be stored on the responsible partner's storage system. FMI and Valtori (Government ICT Centre) provide holistic data security standards in full compliance with European Union data protection laws. Every partner is still responsible for ensuring that the data are stored safely and securely internally. Individual partner's institutional online repositories will also host and preserve data until the end of the project. During the project, partners indicate that the data is going to be backed up in a safe environment such as local storage (Private Corporate Cloud), SURFdrive, OneDrive, Amazon Web Services (AWS) and by using automated saving and online sharing platforms, such as Teams. PIISA Consortium's private file-sharing folder in TEAMS is the main platform for data and information sharing during the project and it is only accessed by PIISA partners. This allows secure data share across partners and an archive for all the documentation produced along the Project lifespan. Within PIISA all data is stored electronically on password-protected computers. Electronic data is backed up periodically. The dataset is uploaded and preserved onto an internal shared cloud storage (SharePoint). The dataset is preserved indefinitely and there is no expected cost associated to its preservation. For large files shared with other partner(s), PIISA recommends using the [Funet FileSender](#) or [SURFfilesender](#). Both platforms are secure ways to share large files among PIISA partners.

At the end of the project, the data and research outputs will be safely stored in Zenodo which is a trusted repository for long-term preservation and curation. Only aggregated non-sensitive information will be available through the PIISA data catalogue in Zenodo. Data from surveys, interviews and workshops that could be sensitive by nature will not be publicly available. The project partners are obliged to protect the results where these can be expected to be commercially or industrially exploited.

In compliance with the General Data Protection Regulation (GDPR), all data processed by our software is anonymized and aggregated to prevent identification of individual consumers, thereby respecting privacy and eliminating any risks related to personal data handling. Our resources undergo regular backups, and an external audit of our data management practices is conducted by an independent company to ensure optimal security. In the event of a data breach, we have a robust incident response plan to identify and mitigate the impact swiftly. Lastly, we conduct regular data protection impact assessments to identify and reduce data protection risks, ensuring our compliance with GDPR.

Note: For example, The PIISA website operates without the use of cookies and does not collect any personal data, such as email addresses or IP addresses. For the newsletter subscribers explicitly consent to the use of their information (including name, email address, and organization) solely for receiving the PIISA newsletter.

7 Ethics

Most of the data used and produced in PIISA is openly available and doesn't have ethical or legal restrictions for sharing. Ethical or legal issues that may have an impact on data sharing include confidential data from third parties and personal data (if it cannot be anonymized).

Sensitive information from surveys, interviews and workshops is handled in a separate system adhering to the ethical and legal regulations for such data. When collecting data from surveys, interviews and workshops, it's important to address constraints and instructions on data collection, storage and security to ensure that the processes are in compliance with EU data policy and the PIISA data management plan. When collecting sensitive data, considering these aspects is necessary:

- **EU Data Policy Compliance:** all data collection and storage processes must be fully compliant with EU data policy regulations, such as the General Data Protection Regulation (GDPR) to protect the privacy and rights of survey participants.
- **Data Minimization and Limitation:** Data collection tools should be designed to collect the minimum amount of data necessary for the survey's purpose. This aligns with the GDPR principle of data minimization, which discourages the collection of excessive or unnecessary personal information.
- **Informed Consent:** Administrations need to obtain clear and informed consent from survey participants before collecting any data. Participants must be fully aware of the purpose of data collection, how their data will be used, and their rights regarding data protection.
- **Anonymization and Pseudonymization:** Use of anonymization and pseudonymization techniques to protect the identities of survey participants are recommended when sensitive or personal data isn't valuable to the objective of the survey. Anonymizing personal data ensures that individuals cannot be identified, while pseudonymization involves replacing identifying information with pseudonyms.
- **Secure Data Collection Tools:** Use of secure survey tools that comply with EU data protection regulations are recommended. Ensure that the chosen tools have appropriate security measures in place to protect the data collected, including encryption and access controls. Access to survey data should be restricted to authorized personnel only.
- **Data Encryption:** Choose data collection tools that support data encryption both during transmission and while at rest. Encryption ensures that data is protected against unauthorized access, reducing the risk of data breaches and complying with EU data policy requirements.
- **Data Retention and Deletion Policies:** clear data retention and deletion policies ensure that Data is not retained longer than necessary for the survey's purpose, and when it is no longer needed, the data is securely deleted.
- **Documentation and Record-Keeping:** Maintaining records of data processing activities is a GDPR requirement. The tools should facilitate the creation and maintenance of records that document data processing activities. GDPR requires data controllers to maintain records of data processing for compliance purposes.
- **Secure Data Storage and Access Control:** Select data storage solutions that offer robust security features, including access controls, authentication, and encryption. These tools should comply with GDPR standards for secure data storage and access.

By following these guidelines, we ensure that data collection within the PIISA project is conducted in a manner that respects EU data policy, protects the rights of participants, and maintains data security and compliance. The information provided is based on generally recognized best practices and guidelines for ensuring compliance with EU data policy and the General Data Protection Regulation (GDPR) when dealing with personal and non-personal data. These best practices are informed by legal requirements and widely accepted data privacy and security standards in the European Union. To find more official and up-to-date information from the EU Commission on data protection and GDPR



compliance, please visit the EU Commission's Data Protection page: [Data protection under GDPR - Your Europe \(europa.eu\)](https://europe.europa.eu).

Note: Compliance guidelines can vary depending on the nature of the survey, the tools and technologies being used, and the organisation's data protection policies. Therefore, while the information provided reflects common practices, it should be adapted to the specific context and requirements of the organization and its survey activities. When dealing with data protection and privacy matters, it's advisable to consult with legal experts or data protection officers to ensure full compliance with relevant regulations and guidelines.

7.1 Responsible Use of AI in Research

The European Commission, together with the European Research Area countries and stakeholders, has put forward a set of [guidelines](#) to support the European research community in their responsible use of generative artificial intelligence (AI). PIISA partners are obligated to follow these guidelines for the responsible use of generative AI in research. Key takeaways from the guidelines include:

- **Researchers** refrain from using generative AI tools in sensitive activities such as peer reviews or evaluations and use generative AI respecting privacy, confidentiality, and intellectual property rights.
- **Research organisations** should facilitate the responsible use of generative AI and actively monitor how these tools are developed and used within their organisations.
- **Funding organisations** should support applicants in using generative AI transparently

Generative AI can hugely boost research, but its use demands transparency and responsibility. As generative AI is constantly evolving, the EC/ERA guidelines will be updated with regular feedback from the scientific community and stakeholders. PIISA's data management will follow these updates to keep abreast of recent guidance.

PIISA's Consortium members agreed to include a disclaimer in PIISA's deliverable template about the utilization of generative AI tools to enhance clarity and quality (e.g. translation). The common disclaimer in PIISA documents to promote transparency states as follows:

"AI was used to enhance the clarity of certain aspects in this document. The AI tool(s) used in this document is/are... (*write here the name of the AI tool and version used*). AI was used in sections... (*write here the chapters/sections*). The purpose was to... (*write here the reason for use*). The authors retain full accountability for the content and its accuracy."

Authors are obligated to fill in the disclaimer with the information about the specific AI tool (and its version) used, certain sections where the AI tools were used (if the use of AI concerns clearly referable sections of the document) and the reason for utilizing AI.

8 Other issues

The lead partner is subject to the Data Protection Act; Freedom of Information Act and General Data Protection Regulation. All data must be collected, stored and disseminated in accordance with these Acts. Other issues will be updated later if needed.



References

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018.
<https://doi.org/10.1038/sdata.2016.18>