Promoting and Incentivising Federated, Trusted, and Fair Sharing and Trading of Interoperable Data ASsets

D8.2
Data Management Plan

Editor(s)	Yury Glikman
Lead Beneficiary	FHG
Status	Final
Version	1.0
Due Date	30/06/2023
Delivery Date	11/07/2023
Dissemination Level	PU

Project	PISTIS – 101093016
Work Package	WP8 - Project Coordination, Scientific Technical and Innovation Management
Deliverable	D8.2 – Data Management Plan
Contributor(s)	SUITE5 , ALEGAL
Reviewer(s)	Sofianna Menesidou (UBITECH), Marios Zacharias (SPH)
Abstract	This deliverable is the first version of the Data Management Plan (DMP) for the PISTIS project outlining how the project handles its data.

Executive Summary

This deliverable presents the first version of the Data Management Plan (DMP) for the PISTIS project after its first 6 months. The plan is consistent with the projects' participation in the Open Research Data Pilot initiative and with the FAIR (Findable, Accessible, Interoperable and Reusable) approach to data management in Horizon Europe.

The document provides an overview of data used and generated in the project including the data used in the PISTIS demonstrators, project public deliverables, its scientific publications and referenced in them data. Then, the document depicts the PISTIS approach following the FAIR principles to data management, aiming at making research findings and content be discoverable, accessible, interoperable, and re-usable by all interested and legally permitted stakeholders (especially by the European research community), whilst preserving the adopted in the project legal and ethic policies. The plan follows the principle "as open as possible, as closed as necessary".

Table of Contents

List of Tables

Terms and Abbreviations

AODB	Airport Operational Database
AMS	Airport Management Solution
BSCW	Basic Support for Cooperative Work
DMP	Data Management Plan
DSU	Data Sharing Unit
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation
OA	Open Access
OECD	Organisation for Economic Co-operation and Development
P2P	Peer-to-Peer

1 INTRODUCTION

The purpose of this document is to provide a data management plan (DMP) for PISTIS supporting the data management life cycle for all research data that will be collected, processed or generated by the project. The plan explains how the research data associated to PISTIS will conform to the 'FAIR' principles, making it findable, accessible, interoperable and re-usable. It is a living document, which is going to be updated to address the needs of the project.

1.1 DOCUMENT STRUCTURE

The document follows the recommended by the European Commission DMP template for the Horizon Europe research programme[1]. It has the following structure:

- Section 2 provides a summary of the data collected, used and generated within the project.
- Section 3 provides the activities planned to make the research data generated in the project to conform to the FAIR principles.
- Section 4 presents the other than data research outputs of PISTIS and the plan for managing them.
- Section 5 provides an overview of the resources planned for managing the PISTIS data in accordance with the FAIR principles.
- Section 6 presents the provisions in place for data security.
- Section 7 addresses the ethical and legal aspects related with data collection, management and sharing in the project.
- Section 8 concludes the document.

2 DATA SUMMARY

PISTIS aims to create a reference federated data sharing/trading and monetisation platform for secure, trusted and controlled exchange and usage of proprietary data assets among the registered organisations. The platform is not restricted to a specific type of data. In contrary, it enables organisations to trade and exchange data of heterogeneous formats, types and sizes.

The purpose of collection and generation of data in the PISTIS project is to achieve the main goal and strategic objectives of the project. Specifically, the data collected from the sources described below is used to show how trading/exchange of this data can be realised among organisations using the PISTIS technologies and which benefits it brings to them. Apart from that, collection and generation of data are necessary to manage the project, disseminate the information about it, analyse and exploit its results.

---

[1] Horizon Europe. Data Management Plan Template, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx

2.1 DATA RELATED TO THE DEMONSTRATOR CASES

The developed in the project software is going to be evaluated in different use cases of 3 demonstration hubs. The following subsections provide a short preliminary overview of the data planned to be used in the demonstration use cases. Most of the data are proprietary and confidential. Nevertheless, the consortium will evaluate which data and under which conditions can be shared with the research community.

2.1.1 Demonstration Hub #1 – Mobility and Urban Planning

The mobility and urban planning demonstrator hub will enable data trading and sharing amongst stakeholders from aviation, public transports and public administration in the region of Athens, Greece for creating a value chain around mobility and urban planning data. The demonstrator hub includes AIA (Athens International Airport), GOLDAIR (Ground Handling Service Provider), OASA (Athens Public Transportation System, including bus and metro lines) and DAEM (City of Athens IT and Development division), that act as data consumers and/or providers depending on the use case.

The demonstrator hub members are going to use/exchange among them the following data:

- Flight tracker data (50KB/min);
- AMS SITA Personnel resources workflows & tasks staff planning (300KB/day);
- Transfer passengers (200KB/day);
- AIA AMS SITA AODB (historical data-40GB; flight schedules-200KB/day, flight allocations-200KB/day, real time flight updates-50KB/min);
- Weather data;
- Baggage data;
- Baggage journey data;
- Bus schedule data;
- Metro schedule data;
- Bus Fleet data,
- Bus Active Capacity (measured per stop);
- Metro Station Incoming/Outgoing Passengers (per minute);
- City Commercial Zone Data (businesses, types, turnover, etc.);
- City Visitors Load data (visitors per zone);
- OAG Schedules & Connections Data (5-7 million records/day)

2.1.2 Demonstration Hub #2 – Energy Sector

The energy demonstrator hub will work on ensuring the resilient operation of the distribution grid through the utilization of the flexibility capacity that can be offered by local prosumers and triggered by the aggregator (as the main actor involved in flexibility transactions and representing aggregated clusters of prosumers in energy markets), thus providing a real environment for validating the operational benefits of data sharing.

The hub involves 4 partners playing different roles in the energy sector value chain: CUERVA, as the Distribution System Operator and the owner of the Granada Living Lab infrastructure, providing a broad number of prosumers to be engaged in the project and real data from the electrical network; OMIE, as a market operator; BAMBOO, as an aggregator, to enable the provision of flexibility on the energy demand side and CARTIF supporting the technical needs of the demonstrator.

The demonstrator hub members are going to use/exchange among them the following data:

- Energy Consumption Data (Hourly Data Consumption from 5500 Supply Points – 5.5GB/hour in total, Spain (clients)),
- Real Time Data from the Secondary Substation - 29 MB per secondary substation (counting for 44 substations with advanced low voltage supervision, 1 record per 5 minutes, Spain (Secondary Substation)),
- Grid Topology data,
- Grid Energy Requirements data,
- DERs Real Time Data,
- DERs penalties,
- Market transactions,
- Dispatch data,
- Validation data from digital twins.

2.1.3   Demonstration Hub #3 – Automotive Sector

The automotive demonstrator hub will enable exchange of data between its members to perform traffic quality assessment in urban areas (led by TRA) and driving style and risk assessment (led by VIF). This will enable them to provide data-driven services to individual drivers (driver coaching), businesses (corporate mobility management for green driving), and public administrations (urban emissions' modelling). The main providers of the data will be CARUSO as the partner operating automotive data marketplace and VIF as the Europe's largest RTO for virtual vehicle technology. The weather data will be provided by UBIMET.

The demonstrator hub members are going to use/exchange among them the following data:

- Historical vehicle data (~50 GB historical time series data of recorded trips available from >100 different drivers),
- live driving data collected with a smartphone app, up to 100 Hz (GPS, IMU data),
- Connected vehicle data, 5MB/h per vehicle (speed, GPS, assistance system status, drivetrain status, etc),
- Historical weather data (geo-spatial data, - 40 MB per year per weather station),
- Data on current weather situation (low frequency),
- Air quality data (15 MB per year per measurement station),
- Trip data from Smartphones,
- Map data.

2.2   DATA RELATED TO SCIENTIFIC PUBLICATIONS

During the course of the project the consortium will present its research results in scientific publications and will make the referenced in the publications data available to the scientific community as curated and anonymized datasets. Those will be published in an open access data repository such as OpenAIRE[2] or Zenodo[3]. The main target groups for such data are the researchers working in projects that are similar to PISTIS or in the PISTIS demonstration domains.

2.3   PISTIS PUBLIC DELIVERABLES

The project deliverables contain valuable information for the community. Therefore, the publication of all project public deliverables is considered as part of the data management plan. The following table provides an overview of the planned public deliverables (reports):

Table 1: PISTIS Public Deliverables (reports)

#	Deliverable Name	Due Date
D1.1	PISTIS Operation Principles and Context Detailing	30 Jun 2023
D1.2	PISTIS Technical Requirements and MVP – v1	30 Sep 2023
D1.3	PISTIS Technical Requirements and MVP – v2	30 Nov 2024
D2.1	Data Interoperability, Management and Protection Framework	30 Sep 2023
D3.1	Data Valuation, Sharing and Trading Framework	30 Sep 2023
D4.1	PISTIS Reference Architecture and API Documentation	31 Dec 2023
D6.1	Dissemination, Communication, Liaison, Training and Living Lab Plan	31 Mar 2023
D6.2	Dissemination Activities Report, Training Material and MOOC - First Report	30 Sep 2024
D6.3	Dissemination Activities Report, Training Material and MOOC - Second Report	30 Jun 2026

All public deliverables (reports) will be published on the project web site in PDF format and will follow the naming format

PISTIS D#.# <Title> v#.#.pdf

3   FAIR DATA

3.1   MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

The discoverability of PISTIS open data will be ensured by the following actions:

- Publishing PISTIS open research datasets at open access platforms, as Zenodo[3].

---

[2] https://www.openaire.eu/
[3] https://zenodo.org/

- Accompanying datasets with properly structured and accurate metadata identifying their content, structure, potential use and license conditions.
- Making data identifiable by using standard identification mechanisms along with persistent and unique identifiers (e.g. DOI, where applicable).
- Advertising the published curated and anonymised datasets on the PISTIS website.
- Publishing papers and reports with references to those datasets.

3.2   MAKING DATA ACCESSIBLE

In this section we describe Open Access (OA) to data that are to be made openly available during the PISTIS project.

Open Access refers to the practice of providing online access to data such as scientific information or research data.

Open Access to scientific information during the PISTIS project will be achieved via two routes:

a. Self-archiving / "Green" open access): the published article or the final peer-reviewed manuscript is archived in an online repository before, at the same time as, or after publication. Some publications can be delayed, and open access be granted only after an embargo period has elapsed (usually 6 months)

b. Open access publishing / "Gold" open access) the article is immediately published in open access mode. Therefore, publication costs are not borne by subscribed readers, instead, the publication costs are usually borne by the entity (university, institute, etc.) funding the research. In other cases, the costs of open access publishing are covered by subsidies or other funding models.

The preference for all PISTIS scientific publications will be that of "Green" open access. In case that the consortium decides that certain publications should be made available under the "Gold Open Access" scheme to bring higher value to the project and/or the scientific community (for example publishing on a very prestigious journal that only provides this scheme), then this option can be also adopted and the costs will be covered by the partners involved.

To define which data should be made available using the Open Access strategy of the project, the following structured list of questions has been compiled which can guide any partner to understand if their data should be made available and how.

1. **Data's Significant Value**. Does a result provide significant value to others or is it necessary to understand a scientific conclusion?

If Yes, then the result is classified as public (i.e. granted for open access).

If No, the result is classified as non-public. (E.g.: code that is specific to a platform (e.g. a database initialization) is usually of no scientific interest to anyone, nor does it add any significant contribution.

2. **Inclusion of Personal Information**. Does a result include personal information that is not the author's name?

If Yes, the result is classified as non-public. Any personal information further than the name must be removed if it should be published according to the ethics management plan of the project.

3. **Identification of Individuals**. Does a result allow the identification of individuals even without the name?

If Yes, the result is classified as non-public. This is also covered by the PISTIS ethics management plan towards anonymizing a single user's identity, e.g. abstraction, dummy users, or non-intersecting features.

4. **Harm Potential**. Can a result be abused for a purpose that is undesired by society in general or contradict with the project's ethics?

If Yes, the result should be classified as non-public. This is also managed by the ethics management plan of the project (D9.1).

5. **Sensitive Business Information.** Does a result include business or trade secrets of one or more partners of the project?

If Yes, the result should be classified as non-public. Any business or trade secrets need to be deleted in accordance to all partners' requirements prior to being published.

6. **Ongoing Patent Information.** Does a result name technology that consist of an ongoing, project-related patent application?

If Yes, the result should be classified as non-public. The result can be published once a patent has been filed.

7. **Sensitive Security Information**. Does a result break security interests for any project partner?

   If Yes, the result should be classified as non-public.

Upon classification of the different data generated and collected by each partner, the following responsibilities should also be met towards disseminating the project's outcomes.

- Define how these data will be made available;
- Define how access will be provided in case there are restrictions;
- Define if and to which registry these data will be added;
- Define the methods or software required to access such data; whether a documentation is necessary about the software and if possible, include the relevant software (e.g. in open source code);
- Define the repository of the data and associated metadata, documentation and code.

In principle, datasets will be made available under restricted access on the PISTIS platform during the project, while at the end of the project, open data will be uploaded to a public repository.

The publicly available reports and other documents related to the project will be accessible from the PISTIS project website, as well as the CORDIS results website[4] (published by EC) upon acceptance by the EC. Publications will be uploaded on open access platforms, such as OpenAIRE[2] or Zenodo[3]. Open access is also possible directly through the organisations own channels, such as via direct links to papers and other project results.

For open-source code, PISTIS will make use of GitHub[5] as code repository.

3.3 MAKING DATA INTEROPERABLE

The data standards related to the project activities might vary depending on partners' internal tools and methodologies, always respecting the guidelines of this DMP and other regulations and legal documents (for example: GDPR, Directive 2002/58/EC on privacy and electronic communications, the Human Rights Law, including the European Convention of Human Rights and the Charter of Fundamental Rights of the European Union) that are of important relevance for the data. The consortium will publish the data in a machine-readable format following the best practices for its structure.

The data that are gathered and stored in the PISTIS platform will be accompanied by detailed metadata following the DCAT[6] vocabulary and a PISTIS specific application profile that will be developed as part of the PISTIS implementation in conformance with the GAIA-X initiative.

3.4 INCREASE DATA RE-USE

To increase data re-useability by other parties, it is recommended that partners generating these data accompany them with proper documentation (like readme files, etc.) that can help other parties. Nevertheless, the exact type of documentation accompanying the data will depend on the types of data to be generated.

The partners will choose which data will be made openly available, based on the above-described methodology and for those will ensure the widest possible re-use. In that case, the data will be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement.

Reuse of data by third parties, during and after the end of the project is something that is of interest of the PISTIS project, and therefore all consortium partners will be encouraged to offer their data as openly available. When it comes to data to be shared over the PISTIS platform, these will adhere to the licensing terms defined by the data owners, and as part of the business model of the platform, these data are through to be useable by third parties, but the usage and access terms will be governed by specific contracts and licenses.

Finally, data provenance will be properly addressed in PISTIS by including a dedicated software component in the PISTIS platform monitoring the evolution of datasets within the PISTIS ecosystem.

---

[4] https://cordis.europa.eu/project/id/101093016
[5] https://github.com/
[6] https://www.w3.org/TR/vocab-dcat-3/

4 OTHER RESEARCH OUTPUTS

One of the most important outputs of the PISTIS project is the software developed in it. The consortium is planning to reuse existing Open Source components and contribute to the Open Source community back. The responsible technical partners contributing to the development of the open source software component will decide which of them can be open sourced. The source code will be reviewed before its publication, while information about platform security (e.g. server names), user privacy, or business secrets will be removed. The documentation of the source code along with the associated metadata will be included in the code repository. The details of the open source datasets will be defined during the implementation and updated in the following versions of the data management plan.

5 ALLOCATION OF RESOURCES

During the course of the project, the necessary resources will be mobilised to cover the incurred costs for making PISTIS data FAIR and for making data accessible, in accordance with the principles and the articles of the project's Grant Agreement.

The main responsible for the data management of the project will be the coordinator (FHG), which shall keep a copy of all documentation during and beyond the project's lifetime in the Basic Support for Cooperative Work (BSCW) server that facilitates also document management for the project. Nonetheless, all partners will be responsible for data processed by their private servers. In addition, all partners will ensure that such data is protected, and any necessary data security controls have been implemented to minimize the risk of information leak and damage.

Long term preservation of the data will be performed both by all partners using their own repositories and them covering these costs, and by storing the data in well-respected public repositories (such as Zenodo), which are provided as free of charge. Moreover, all partners are expected to use their own repositories and to store copies of the data after the project execution, maintaining at the same time the confidentiality and not disclosing any information.

The BSCW will be accessible by all partners for at least one year after the end of the project. After that period, Fraunhofer will store the data in a cold storage environment for the next 4 years, and the access will be provided if needed prior request.

6 DATA SECURITY

PISTIS will pay particular attention to secure all collected and generated data, guaranteeing the findability, interoperability, and reusability both during the project lifespan and after its completion. This will be achieved by utilizing state-of-the-art software frameworks, libraries and protocols that enhance the data security, while the provisions for data security and recovery should be ensured by the partners responsible for the ICT implementations in the project (clearly specifying the data to be handled per ICT implementation). The security measurements will be constantly reviewed and updated during the course of the project.

In the field of technical security of the ICT developments delivered in the project, the appropriate measurements will be considered to ensure that the PISTIS solution is handling data security aspects. The security services to be involved in the PISTIS platform system involve end-to-end encryption services for data assets, and novel techniques and methods that have to do with Data Sharing Units (DSUs), attribute-based access control, encrypted ledger channels, etc.

Moreover, and in order to cover also the data security at business level (information exchange among the key demonstrator partners in the different hubs), PISTIS will review the relevant standardization as well as the guidelines from ongoing initiatives (supported by European institutions such as ENISA, ENCS, and EEISAC) in order to ensure that information will be shared in a secure way following widely used standards, whenever possible, and different models of disclosure, including limited, full and coordinated. All servers where data will reside shall use server-side encryption.

In case of personal data, the appropriate methods will be used so that these are made accessible only to authorised parties, as far as data will be stored in independent repositories with restricted access.

Apart from the core ICT outputs of the project, all other tools and software to be used by the consortium and that will be used as part of the data management operations for the PISTIS project, will be selected taking into account the security and privacy guarantees it offers and their compliance with the needs of the project.

7 ETHICS

There is a number of legal and ethical issues that could have an impact on data sharing and thus the project data management and these issues will be analysed and discussed under several project deliverables. More specifically, the legal and ethical aspects of the data sharing and trading activities on PISTIS Platform and the relevant legal requirements as well as ethical principles will be regularly assessed and updated in the deliverables D1.1 and D1.3 as a part of Task T1.3 - GDPR, Cross-Border Legal Aspects and Contracts Definition for Data Sharing and Trading throughout the project in accordance with the policy and regulatory developments in EU. Within these deliverables, the project identifies potential legal and ethical issues related to data protection and privacy and then set forth specific requirements to support the legal and ethical compliance of the project. Furthermore, Task T8.4 aims to ensure that the project adheres to the applicable legal requirements as per EU law. The work under this task will be reported and a complete ethics self-evaluation will be performed in the deliverable D9.1 - OEI - Requirement No. 1 in M12. Also, ethical dimension of the project, ethics self-assessment methodology and the envisioned safeguards are described in Section 4 Ethics & Security of the PISTIS Description of the Action Part B. To complement the aforementioned deliverables, this chapter gives an overview of potential ethical issues related to data and Artificial Intelligence.

7.1 DATA ETHICS

It is evident that there is much to be gained from data sharing. However, care must be taken to ensure that data sharing is conducted in an ethical manner meaning that the conflicting interests of all stakeholders must be observed and balanced. The goal is to enable the sharing of data while ensuring that all stakeholders are enabled and empowered to enjoy and exercise their rights. Therefore, data sharing and access to data require the creation of a trusted environment with sufficient levels of transparency, accountability and protection of personal data.

7.1.1 Transparency

In addition to transparency being a guiding data protection principle stipulated in the GDPR, the notion of transparency is also a crucial ethical principle ensuring that stakeholders are offered the appropriate level of control regarding sharing of their data. The project seeks to implement a platform for federated peer-to-peer (p2p) data sharing and trading among a wide range of users. Therefore, it is necessary to ensure that appropriate levels of transparency are achieved with regard to the types of data used to drive innovation and determine the sharing potential across the data life cycle.

Transparency is about clarity, openness and honesty between data providers and data holders including the data subjects whose personal data is processed. These concern who receives the data, for which purpose the data is shared, who the data controller is in case of personal data processing, and how and why personal data is used. Transparency is always important, especially in situations where data holders have a choice about whether they wish to interact with data users. If data holders or providers know at the outset what the data user will use their data for, they will be able to make an informed decision about whether to exchange with them, or perhaps to try to renegotiate the terms of that relationship.

With respect to sharing of personal data, there must also be appropriate transparency as required by law. Under Articles 13 and 14 of the GDPR, the data controller is required to provide individuals with certain information about the identity of the controller; contact details of data protection officer, the legal basis for the processing, the purpose of the processing, the recipients or categories of recipients of the personal data; the retention period of the personal data; existence of the data subject rights and information on how to exercise these rights; details of competent supervisory data protection authority and the existence of automated decision-making. Furthermore, this set of information must be provided in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child. In practice, privacy policies are used by controllers to fulfil the transparency requirements. Further information on the transparency requirement in the GDPR is available in D1.1 PISTIS Operation Principles and Context Detailing.

7.1.2 Accountability

Accountability represents a key ethical principle; it entails that someone can be held responsible for their actions. The concept of accountability has a slightly stricter form of responsibility. While responsibility is task-focused concept requiring the one to be responsible for ensuring that a task is completed, accountability goes beyond being responsible for the

decisions and actions and expects the relevant actor to explain these decisions and actions as well as provide proper relevant justification.

In a data exchange market, the issue of accountability is a rather complex one. Within the PISTIS project, there must be appropriate accountability mechanisms to ensure that stakeholders participating in the project can effectively hold someone accountable when things go wrong. For example, who can be held accountable should the data from a stakeholder fall into the wrong hands or are misused?

Data providers and data recipients should have the ability to hold someone accountable for any harm caused and, where appropriate, must be able to ensure they can receive compensation for any damages suffered. The concept of accountability is about owning and co-owning roles and responsibilities, performing proper actions and offering assistance should anything go wrong. Therefore, accountability must not be an afterthought but should be engrained thorough the various technical features of the data exchange market and appropriately communicated to stakeholders.

To ensure accountability, mechanisms that facilitate the system's auditability will be established in PISTIS. For example, the traceability and logging of data transfers between different stakeholders is ensured via blockchain technology. Therefore, the parties involved in these transactions are able to follow the life cycle of data provided by them.

7.1.3 Protection of Personal Data

When setting up a federated peer-to-peer data sharing and trading platform such as PISTIS, privacy and data protection laws need to receive special attention. During the demonstrator hub workshops carried out in WP1, the end-users of PISTIS are initially identified as legal entities at this stage and it is concluded that the PISTIS platform will not be used to collect personal data directly from individuals. However, datasets to be provided to the platform by the end-users might still contain personal data and thus sharing of such datasets would constitute processing of personal data. In fact, during the workshops, it is acknowledged by some of the platform user partners that the datasets to be provided by them is likely to contain personal data. However, these partners note that informed consent of individuals concerned will be obtained before the relevant datasets are being shared on the platform.

Furthermore, with sharing of a vast amount of data and use of AI technologies for data processing, personal and non-personal data become increasingly intertwined. PISTIS understands that there is a risk of re-identification of individuals from datasets containing non-personal data due to use of advanced AI systems. Therefore, right from the beginning, PISTIS has been committed to high standards of data privacy and placed a high priority on protecting and managing personal data in accordance with ethical standards and data protection law, both at European and national level. The GDPR Checker and Anonymisation components of PISTIS Architecture are excellent examples of PISTIS's approach to data protection.

The deliverable D1.1 PISTIS Operation Principles and Context Detailing focuses on a certain data protection-related issues particularly relevant to PISTIS. It explains the definitions of certain key data protection terms and the role of PISTIS platform and discusses the data protection principles and the respective legal requirements. Finally, it sets out certain legal requirements under the GDPR that are most relevant for PISTIS.

7.2 ARTIFICIAL INTELLIGENCE ETHICS

PISTIS aims to develop and deploy ML-enabled data analytics and fair data valuation service. This entails machine learning models and various data analytics techniques will be tested and implemented within the project. On this basis it seems appropriate to devote a section on the ethics of artificial intelligence. Thus, this section first examines if the tools developed by the consortium are likely to fall under the definition of artificial intelligence and, second, provides an overview of the guiding ethical principle and standards concerning AI.

7.2.1 Ethics Principles for Trustworthy AI

The EU strategy on AI aims to create a standard for trustworthy AI that EU businesses could use to capture market opportunities within and outside the EU. The goal is to improve the EU economic competitiveness and turn the Union into a global leader on the matter. In 2021, as part of its strategy on AI, the European Commission published Guidelines on Ethics by Design and Ethics of Use Approaches for Artificial Intelligence[7] which is developed from the EU's High-Level Expert Group's Ethics Guidelines for Trustworthy AI published in 2019[8]. In addition, Organisation for Economic Co-operation and Development ("OECD") has also set forth ethical principle for trustworthy artificial intelligence in the Recommendation of the Council on Artificial Intelligence in 2019.[9] Although the Commission's Guidelines and the OECD's recommendation are not legally binding and do not offer pieces of advice on legal compliance for AI, they are still considered as two of the most prominent and perhaps, most influential documents in the EU.

The main objective of these documents is to guide AI system providers to adopt an ethically-focused approach while designing, developing, and deploying and/or using AI powered systems by incorporating ethical principles into the development process. As both the documents adopt human-centred values, they seem to overlap in a set of generic principles that could enable the development of ethical AI. Therefore, the PISTIS consortium considers the following six (6) high level principles of the ethical AI set forth under these documents.

Respect for Human Autonomy. In practice, respect for human autonomy means that humans interacting with AI systems should be able to keep full and effective self-determination and to partake in the democratic process without being hindered by AI. It encapsulates three more specific principles, which define fundamental human rights: autonomy, dignity and freedom. According to this principle, AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans in any part of their life. Instead, these systems should be designed to augment, complement and empower human cognitive, social and cultural skills. In that regard, the principle requires allocation of functions between humans and AI systems to follow human-centric design principles and leave meaningful opportunity for human choice.

Privacy & Data Governance. AI systems must guarantee privacy and data protection throughout a system's entire lifecycle. Therefore, AI systems must be built in a way that

---

[7] European Commission, Ethics by Design and Ethics of Use Approaches for Artificial Intelligence, Version 1.0, November 2021

[8] High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019

[9] Organisation for Economic Co-operation and Development, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 22 May 2019.

embeds the principles of data minimisation and data protection by design and by default as prescribed by the GDPR. This covers both the information provided by individuals, and the information generated about them over the course of their interaction with the system. For instance, outputs that the AI system generated for specific users or how users responded to particular recommendations should fall under the sphere of this protection. It must also be ensured data will not be used to discriminate against them. Quality and integrity of data must be ensured through appropriate data governance models that ensure data accuracy and representativeness and enable humans to actively manage their personal data and the way the system uses it.

Fairness. It implies a commitment to equal and just distribution of both benefits and costs among people and ensuring individuals and certain groups are free from unfair bias, discrimination and stigmatisation. Similar to the principle of respect for human autonomy, the use of AI systems should never lead to people being deceived or unjustifiably impaired in their freedom of choice in favour of certain groups. Fairness implies that AI system providers should respect the principle of proportionality between means and ends and consider carefully how to balance competing interests and objectives of different social groups. The procedural dimension of fairness also requires the ability for humans to contest and seek effective redress against decisions made by AI systems whereas substantive fairness entails that the AI systems should not foster discrimination patterns that unduly burden individuals and/or groups for their specific vulnerability.

Prevention of Harm, Safety and Well-being. This principle means that AI systems and the environments in which they operate must be safe and secure. This entails the protection of human dignity as well as mental and physical well-being of humans. AI systems must be technically robust and should not cause or exacerbate adverse impacts due to asymmetries of power or information, such as between businesses and consumers, or governments and citizens. Preventing harm also requires consideration of the nature and all living beings. AI systems should also strive to make a positive contribution to these forms of well-being. To realize this goal, possible research participants, end-users, affected individual and communities and relevant stakeholders should be identified at the very early stage, to allow for a realistic assessment of how the AI system could enhance or harm their well-being.

Transparency and Explicability. Transparency requires that the purpose, inputs, and operations of AI programs are knowable and understandable to its stakeholders. Transparency therefore includes all elements relevant to an AI system including data, system and processes by which it is designed and operates, as stakeholders must be able to understand the main concepts behind it such as how, and for what purpose these systems function and reach their decisions. Explicability means the transparency of AI processes and the capabilities, in other words opening the black box. It requires that the purpose of AI systems is openly communicated, and decisions explainable to those directly and indirectly affected.

Accountability by Design, Control and Oversight. Those who play an active role in the AI systems, including the development and the operations of these systems, should take responsibility for the way these systems function and for the consequences. It further requires these actors to be accountable for the proper functioning of AI systems and for the respect of the above ethical principles. To be held to account, the developers or operators of AI systems

must be able to explain how and why a system exhibits characteristics or results in certain outcomes. Oversight entails that humans can understand, supervise and control the design and operation of the AI system. Hence, to ensure accountability, developers must be able to explain how and why a system exhibits particular characteristics. As oversight enables the providers of AI systems to understand and control the functioning and outcomes of the system, it becomes the key pre-condition, together with the transparency and explicability, of accountability.

8 CONCLUSIONS

The deliverable presents the first version of the Data Management Plan according the information available at M6 of the project. It provides an overview of the data and describes the steps for data management which are to be followed during the execution of the PISTIS project. As this report is generated in the early stage of the project execution is considered as a living document which will be further supplemented.