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

D1.1
PISTIS Operation Principles and Context Detailing

Editor(s)	Kyriakos Stefanidis
Lead Beneficiary	ATHENA
Status	Final
Version	1.0
Due Date	31/06/2023
Delivery Date	17/07/2023
Dissemination Level	PU

Funded by the European Union under Grant Agreement 101093016. 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 Commission. Neither the European Union nor the granting authority can be held responsible for them.

Project	PISTIS – 101093016
Work Package	WP1 - PISTIS Trusted and Interoperable Data Trading and Management Framework
Deliverable	D1.1 - PISTIS Operation Principles and Context Detailing
Contributor(s)	FHG, ATOS, IDC, IMPERIAL, ATHENA, ASSENT, SUITE5, UMALTA, EURECAT, DBL, UBITECH, ALEGAL, SPH, ICCS, UBIMET, OAG, AIA, GOLDAIR, DAEM, OASA, CUERVA, BAMBOO, OMIE, CARTIF, VIF, CARUSO, TRAF
Reviewer(s)	Gianluigi Viscusi (IMPERIAL) Sotiris Koussouris (SUITE5) Yury Glikman (FHG)
Abstract	This deliverable presents an overview of the operational context of the PISTIS platform including the use case driven end-to-end scenarios, data lifecycle, business requirements, legal framework and requirements, data management and trading processes, and technology radar.

Executive Summary

Deliverable D1.1 – "PISTIS Operation Principles and Context Detailing" is the first deliverable of WP1 and reports on the activities in the first six months of Tasks 1.1-1.4.

First, a summary of the PISTIS platform is given as envisioned in its initial concept along with the respective demonstration hubs and their relevant use cases. The eleven use cases of the demonstrator hubs of PISTIS project are described along with a preliminary list of the data available within the project.

Furthermore, the main operation principles of the PISTIS platform and the context in which the demonstrator hubs will operate within the platform are defined in detail, following a methodology that includes desktop study, questionnaires, and workshops among the PISTIS end-users. The results of this analysis include three of the main outcomes of this deliverable. 1) A set of end-to-end usage scenarios in the context of PISTIS demonstrators that detail the goals of each scenario, the role of PISTIS platform and the interactions of the end-users for each of the eleven use cases. 2) The full data lifecycle as seen from the point of view of the data trading organisations using the PISTIS platform 3) A set of business requirements that outline the expectations of end-users on how the PISTIS platform should operate. Those three results will be used as the basis of the definition of functional and non-functional requirements of the PISTIS platform and its core functionalities.

Additionally, the ethical and legal considerations for data sharing are provided, including key considerations for demonstrator partners and other to read and work on, for purposes of developing arrangements and related documents to enable and facilitate data sharing when the pilots, data flows, involved stakeholders and value models become clearer at a later stage during the project. In addition, a discussion of clusters of EU laws across four perspectives (data-centric, people-centric, market-centric, and system-centric) is conducted and key takeaways on the legal requirements are produced that are highly relevant to PISTIS. The result is an initial set of recommendations, for the consortium partners, on regulatory compliance with the identified EU laws. Moreover, the legal status of smart contracts in EU law is discussed and the essential requirements for smart contracts envisaged in the Data Act proposal are briefly explained.

Finally, D1.1 gives the background knowledge in the form of an overview and initial analysis of the as-is and to-be processes around data management and data trading as well as a technology radar that includes a comprehensive list of external initiatives, projects, and frameworks that are relevant to the core functionalities of the PISTIS platform. Those will be used as an initial knowledge base for the detailed state of the art analysis that will be conducted in WP2 and WP3 as well as the definition of the core functionalities of the PISTIS platform.

Since D1.1 is one of the first deliverables of the PISTIS project that deals with the fundamental principles of the platform does not depend on other deliverables. However, it is expected that many deliverables that follow will depend on D1.1 and refer to it.

Table of Contents

List of Figures

List of Tables

Terms and Abbreviations

ABAC	Attribute-based access control
API	Application programming interface
BR	Business requirements
CFREU	Charter of Fundamental Rights of the EU
CIM	Common information model
CSIRT	Computer security incident report
CSV	Comma separated values
DAC	Discretionary access control
DER	Distributed energy resources
DGA	Data governance act
DID	Decentralised identifier
DL	Deep learning
DPIA	Data protection impact assessment
DSO	Distribution service operator
EaaS	Energy as a service
EDPS	European Data Protection Supervisory
ETL	Extract, transform, and load
EU	European Union
FFoD	Free Flow of Non-Personal Data
GDPR	General Data Protection Regulation
GIS	Geographic information system
GSE	Ground support equipment
JSON	JavaScript object notation
KPI	Key performance indicator
MAC	Mandatory access control
MCP	Multi country project
ML	Machine learning
MO	Market operator
NFT	Non-fungible token
NWP	Numerical weather prediction
ODRL	Open Digital Rights Language
PoA	Proof of authority
PoW	Proof of work
RBAC	Role-based access control
RDF	Rich document format
REST	Representational state transfer
SLA	Service level agreement
SPoF	Single point of failure
SSI	Self-Sovereign Identity
TOBT	Target off block time
VC	Verifiable credentials
XML	Extensible markup language

1 INTRODUCTION

This deliverable provides an overview of the operational context of the PISTIS platform. Its primary purpose is to outline the platform's operation's high-level principles, along with any legal considerations. It includes the business analysis and use cases, business requirements, legal requirements, data management and trading processes and technology radar.

Among others, the deliverable presents the business analysis and use cases, exploring the possible application of PISTIS in various business scenarios. This section aims to show the platform's potential and relevance to different industries. Then, the document discusses business requirements, focusing on the specific needs that PISTIS must address to fulfil its objectives effectively. It also highlights the legal requirements the platform must comply with, ensuring its operations align with relevant regulations and standards.

The document also discusses data management and trading processes. It digs into data processing, emphasizing the importance of interoperability, quality, and data lineage. Furthermore, it explores data exchange and transfer, security and privacy, and the intricate processes of data trading and value exchange. This section sheds light on data markets, valuation methods, and the mechanisms involved in trading data.

The technology radar section provides insights into the technological landscape surrounding PISTIS. It covers initiatives related to data management, such as those concerning extract, transform, and load (ETL) processes, privacy, security, and trust. It also explores initiatives linked to data trading, including distributed ledgers and monetization. Lastly, it emphasizes the ethical considerations that should guide the operation of the PISTIS platform. It points out the importance of conforming with ethical standards, promoting transparency, and ensuring the responsible use of data.

Overall, the document provides an overview of the PISTIS platform's operational scope and offers valuable insights into its business analysis, requirements, legal aspects, data management, trading processes, technology landscape, and ethical considerations. By covering these essential areas, the document sets the basis for the development and implementation of the platform.

1.1 OVERVIEW OF PISTIS PLATFORM

PISTIS is a holistic framework and reference platform designed to unlock the capabilities of the data economy. It integrates various domains like data management, analytics, finance, crypto, and security to revolutionize how data is managed, shared, and monetized. By leveraging existing and emerging data space infrastructures, PISTIS aims to facilitate efficient and trustworthy data sharing and monetization in line with the concept of European Data Spaces. The platform introduces innovative approaches, enhancing data quality and interoperability while incorporating market principles. It incorporates cutting-edge technologies such as blockchain, fintech, and distributed ledgers to support data monetization and multi-stakeholder investment. PISTIS operates through locally deployed services within each data space, governed by a cloud-based control plane that ensures data discovery, transaction execution, auditing, and other horizontal aspects. It follows a federated approach, prioritizing data ownership and control, and enabling peer-to-peer data transactions. The platform empowers stakeholders by improving data quality, tracking lineage, transforming data through analytics, and selecting revenue models. PISTIS aligns with initiatives like GAIA-X and IDSA, focusing on data curation, interoperability, and security, while driving digital transformation and providing strategic insights in the EU data market.

The initial conceptual architecture of PISTIS (The main architectural design will be conducted within WP4 at a later stage of the project) follows a federated infrastructure approach, divided into two core blocks: the PISTIS Data Space Factory Environment and the PISTIS Data Trading & Value Exchange/Monetization platform.

The Data Space Factory Environment is a crucial component of the PISTIS platform. It is deployed on the premises or on private infrastructure owned by supply-side stakeholders. It is responsible for data transformation, enrichment, and execution of data sharing transactions by communicating directly with other transaction parties. The Data Space Factory Environment is divided into four bundles:

1. *Data Ingestion and Transformation bundle*: This bundle facilitates data ingestion from various sources, such as APIs, streaming data, and batch file uploads. Data enrichment is performed using standard semantic models, and metadata is extracted and stored in repositories. An Analytics/Insights Engine enables the execution of ML/DL analytics pipelines, generating valuable insights from primary data artifacts. Data lineage tracking, along with quality assessment and classification, ensures comprehensive monitoring of data operations.
2. *Data Security and Trust bundle*: This bundle focuses on the security and privacy aspects of data sharing. Access policies are defined using the Open Digital Rights Language (ODRL) standard to control visibility and permissions for data assets. Anonymization services transform data into anonymous or aggregated forms when required. GDPR provisions are verified for datasets containing personal data. Data cryptography services, including encryption and decryption, protect data integrity and confidentiality during exchange.
3. *Data Exchange Preparation bundle*: This bundle involves configuring monetization schemes for data assets and executing and monitoring sharing/trading contracts. The Trading Pre-Processor allows data owners/suppliers to determine the value of their data assets. A Smart Contract Execution Engine executes transactions recorded in the blockchain ledger, making the designated data artifacts available to the demand side. The Data Space Factory Clearing House manages the value attribution in digital wallets of data owners/providers.
4. *Data Peer-to-Peer Transfer Gateway bundle*: This bundle facilitates peer-to-peer data exchange between trading parties. A Contract Checker component validates data trading contracts and enforces their execution during the transfer. It provides interfaces for data provisioning to consumers and reception of data assets acquired through smart contracts. Zero-Knowledge Attestation Agents ensure secure and verified requests from authorized entities without revealing their identities.

Overall, the Data Space Factory Environment forms the core of sensitive data operations within the PISTIS ecosystem. It empowers stakeholders to improve data quality, ensure security and privacy, prepare data for exchange, and facilitate peer-to-peer transactions while maintaining the flexibility to integrate with existing services.

The Data Market Exchange environment is a cloud-based system that governs and orchestrates operations related to data monetization and data market exchange that need to be propagated to the Data Space Factory environments. It acts as a central intelligence hub and access point for the demand side, storing metadata rather than actual data to ensure secure and trusted data sharing. The environment consists of several components:

1. *PISTIS Data Explorer*: A distributed data discovery service that builds a catalog of metadata. It enables stakeholders to search for data assets using a distributed query engine, which combines information from the catalog, distributed ledgers, and Data Space Factory environments. An Access Policy Engine filters query results based on the requesting party's attributes to ensure authorized access.

2. *Monetization XAI Engine*: Utilizes eXplainable AI to analyze the data market, assess data value, and recommend optimal strategies to the supply side. It includes services like Marker Insights for revealing market trends and dynamics, Data Usage and Intentions Analytics for assessing data asset value, and FAIR Data Valuation Services for recommending pricing, monetization schemes, and usage terms.

3. *Data Value Contract Composer*: Drafts contracts that describe how data assets are traded within the PISTIS ecosystem. It includes modules like Asset Description Bundler for packaging asset information, Purchase/Subscription Planner for defining access methods, NFT Generator for trading NFTs and managing copyrights, and Data Investment Planner for acquiring funding and sharing equity.

4. *PISTIS Data Exchange Market*: Acts as the primary monetary trading infrastructure for data trading. It uses a stablecoin to link assets to real monetary values, providing stability for trading. Real money is inserted into the platform via an Open Banking Interface, converted into stablecoins, and distributed to custodial wallets. The PISTIS Market Monitor oversees market activity and facilitates stablecoin transfers.

5. *Data Exchange Governance*: Manages the distributed ledger network within PISTIS using a low-energy and permissioned blockchain solution. It includes services like Smart Contract Template Composer for managing contract lifecycles, Transactions Auditor for monitoring and auditing transactions, and On/Off Platform Contract Inspector for enforcing contract terms and detecting breaches.

6. *PISTIS Identity Manager*: Authenticates stakeholders within the ecosystem using an Identity Provider and validating their identities and attributes. It includes components like an Identity Validator for online or offline identity validation and a Secure Public Keys Store for encrypting data using recipient keys.

7. *Data Models and Platform Services Configurator*: Manages the platform's operation at a central level. It includes services like Model Manager for lifecycle management of sector-specific data models, a Global Semantics repository for transforming data and

metadata, an AI Model Designer with pre-trained AI models for sharing insights, and a Data Space Factory Connectors Manager for managing deployed Data Space Factory environments.

Overall, the Data Market Exchange environment serves as the central point of intelligence for the data market exchange and monetization services. It stores metadata about the data assets, handles data discovery, and provides monetization insights and recommendations within the PISTIS platform.

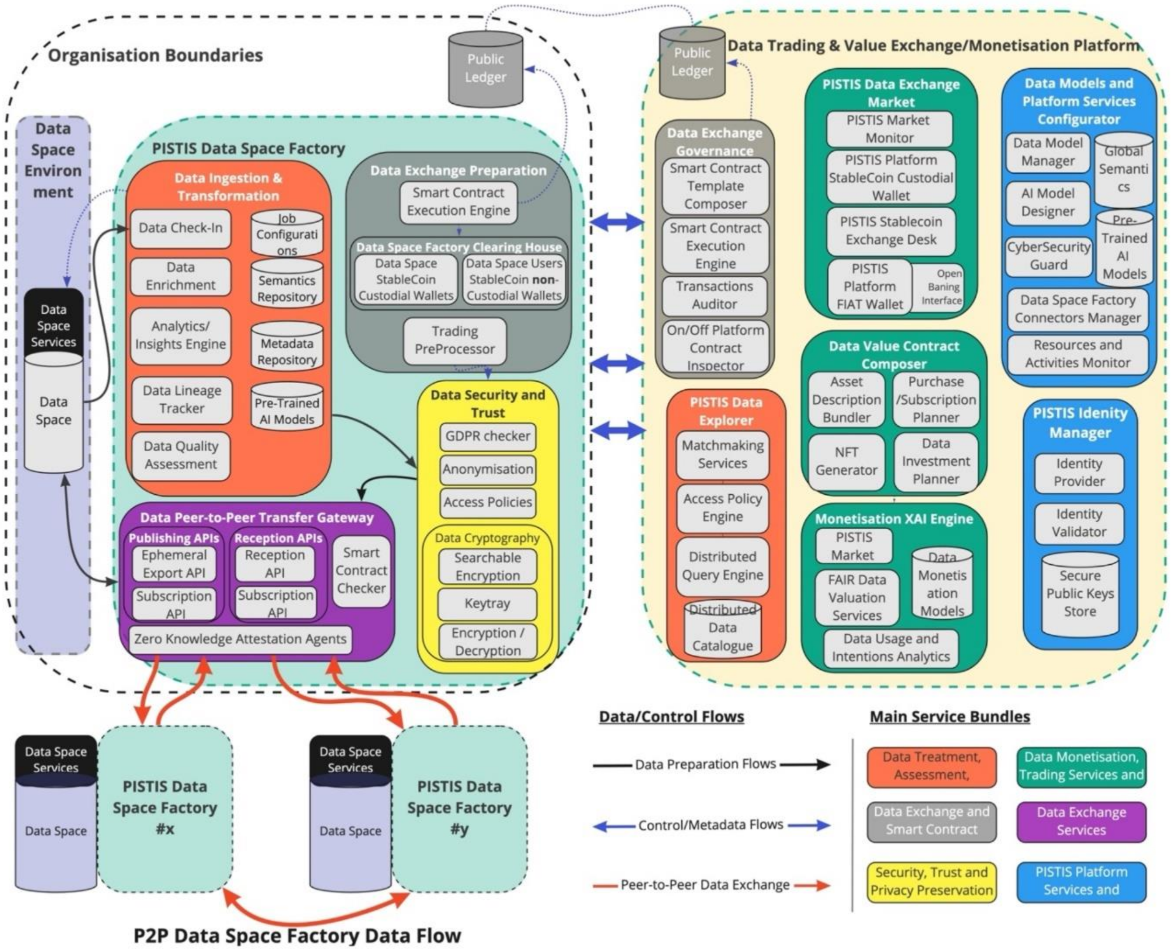

Organisation Boundaries
Public Ledger
Data Space Environ ment
PISTIS Data Space Factory
Data Ingestion & Transformation
Data Check-In
Job Configurati ons
Data Enrichment
Semantics Repository
Analytics/ Insights Engine
Metadata Repository
Data Lineage Tracker
Pre-Trained AI Models
Data Quality Assessment
Data Exchange Preparation
Smart Contract Execution Engine
Data Space Factory Clearing House
Data Space StableCoin Custodial Wallets
Data Space Users StableCoin non-Custodial Wallets
Trading PreProcessor
Data Security and Trust
GDPR checker
Anonymisation
Access Policies
Data Cryptography
Searchable Encryption
Keytray
Encryption / Decryption
Data Space Services
Data Space
Data Peer-to-Peer Transfer Gateway
Publishing APIs
Ephemeral Export API
Subscription API
Reception APIs
Reception API
Subscription API
Smart Contract Checker
Zero Knowledge Attestation Agents
Data Trading & Value Exchange/Monetisation Platform
Public Ledger
Data Exchange Governance
Smart Contract Template Composer
Smart Contract Execution Engine
Transactions Auditor
On/Off Platform Contract Inspector
PISTIS Data Exchange Market
PISTIS Market Monitor
PISTIS Platform StableCoin Custodial Wallet
PISTIS Stablecoin Exchange Desk
PISTIS Platform FIAT Wallet
Open Baning Interface
Data Models and Platform Services Configurator
Data Model Manager
Global Semanti cs
AI Model Designer
CyberSecurity Guard
Pre-Trained AI Models
Data Space Factory Connectors Manager
Resources and Activities Monitor
PISTIS Data Explorer
Matchmaking Services
Access Policy Engine
Distributed Query Engine
Distributed Data Catalogue
Data Value Contract Composer
Asset Description Bundler
Purchase /Subscription Planner
NFT Generator
Data Investment Planner
Monetisation XAI Engine
PISTIS Market
Data Monetis ation Models
FAIR Data Valuation Services
Data Usage and Intentions Analytics
PISTIS Idenity Manager
Identity Provider
Identity Validator
Secure Public Keys Store
Data Space Services
Data Space
PISTIS Data Space Factory #x
Data Space Services
Data Space
PISTIS Data Space Factory #y
P2P Data Space Factory Data Flow
Data/Control Flows
Data Preparation Flows
Control/Metadata Flows
Peer-to-Peer Data Exchange
Main Service Bundles
Data Treatment, Assessment,
Data Monetisation, Trading Services and
Data Exchange and Smart Contract
Data Exchange Services
Security, Trust and Privacy Preservation
PISTIS Platform Services and

Figure 1: PISTIS DoA High-Level Architecture

1.2 Demonstration Hubs of PISTIS

To assess the degree to which the proposed framework can be successfully introduced to the different data spaces, PISTIS will undergo real-life testing and validation in three demonstration hubs in Greece, Spain, and Austria & Germany. Each hub focuses on a specific domain, i.e., mobility and urban planning, energy, and automotive. The goal is to demonstrate the practicality of the PISTIS solution in both vertical (within a specific sector) and cross-domain/sector scenarios. The PISTIS demonstrator hubs encompass complete value chains in cross-domain settings, enabling the validation of data sharing, exchanges, and transactions among them. The three demonstration hubs regard:

- The **Mobility and Urban Planning Demonstrator Hub** (AIA, OAG, GOLDAIR, DAEM, OASA, UBIMET) aims to enable data trading and sharing among stakeholders from various sectors, such as aviation, public transportation, and public administration.
- The **Energy Demonstrator Hub** (CUERVA, BAMBOO, OMIE, CARTIF, UBIMET) aims to ensure the resilient operation of the distribution grid by leveraging the flexibility capacity provided by local prosumers.
- The **Automotive Demonstrator Hub** (VIF, CARUSO, TRAF, UBIMET) aims to promote environmentally friendly, safe, and efficient mobility by utilizing data from various sources.

More details on the demonstration hubs and the related use cases are given in the next section of this deliverable (Section 2.1).

2 PISTIS DEMONSTRATORS’ BUSINESS ANALYSIS & USE CASES

2.1 OVERVIEW OF THE DEMONSTRATORS’ USE CASES

In this Section, the eleven Use Cases (UCs) of the demonstrator hubs PISTIS project are briefly described together with a preliminary list of the data available for the study. The purpose of these UCs is to test and validate the applicability of the PISTIS solution in the settings of realistic scenarios. They are grouped in three “demonstration hubs” focusing on different industrial and geographical domains in Mobility and Urban Planning (Greece), Energy (Spain) and Automotive (Austria and Germany).

2.1.1 Demonstration Hub #1 – Mobility and Urban Planning

Demonstrator Hub #1 will focus on facilitating data trading and sharing amongst stakeholders in aviation, public transport, and public administration, which are key actors of a value chain that can be built around mobility and urban planning data. Five Ucs are foreseen.

2.1.1.1 UC.1.1 Baggage handling management.

Ground handlers receive baggage-related timestamps from the airport that enable proper baggage management and make it possible early identification of irregularities in the baggage delivery process to the passengers. Post-operation analysis of such data is very important to investigate further incidents or non-optimal performance of the baggage delivery process. The airport would benefit from information regarding the baggage loading and unloading process from the handler, the status, the number, and the availability of the ground support equipment (GSE) equipment, the scheduling and availability of the required personnel for baggage management processes.

Information from the handler’s staffing and rostering system, the daily flight schedule on a time horizon of 6 months, and weather predictions can be combined to produce the probability that the flights will be serviced with no delays. Also, the data can be combined with flight updates/changes in real time and signify the latest and most updated status and the probability of achieving the aircraft turnaround and flight departure times.

2.1.1.2 UC.1.2 Transfer Passenger Management.

The impact of delayed transfer passengers who transit at the Athens International Airport (AIA) will be investigated. Information on the number of passengers’ transfers at AIA and on the delay of arriving flights is useful to departing flights which are waiting for transfer passengers to calculate the impact of such delay on their schedules. Sharing information on scheduled and actual transfer passengers between airlines, ground handlers and the airport could also help the optimal allocation of aircraft stands to specific flights to minimize passengers’ connection times between arriving and departing aircrafts.

2.1.1.3 UC.1.3 Aircraft Turnaround process.

Information about Target Off Block Time and turnaround times of the aircraft’s servicing underlying process such as catering, fuelling, cleaning etc. is very important to be exchanged between involved stakeholders. Additionally, information on the aircraft turnaround process enables to estimate whether the daily schedule will be performed as planned. It is very important for the airport to know as soon as possible any issues regarding the turnaround of the aircraft and exchange real-time information with the handlers and airlines regarding issues or incidents in a secure manner. The sooner the airport or handler become aware of any irregularities the shorter the reaction time for mitigation measures will be, thus minimizing impact.

2.1.1.4 UC.1.4: Public Transportation Planning Support.

Data exchange between the airport and the municipality can help improve the overall planning of public transport, allowing the development of services that are able to predict the accuracy of the load within the day to improve route scheduling, deployment of vehicles and efficient the utilization of its fleet, thus offering to commuters better services and lowering operational costs.

The outputs of the analysis might also be traded back to parties such as the airport and the city, for the latter to be able to improve their own offer and services towards the commuters, or to relevant third parties (e.g., duty-free shops, local city businesses, etc.). This can be achieved by combining data for incoming passenger traffic to the airport from public transport (e.g., bus occupancy), data from the airport service handler (e.g., expected queuing, check-in counters availability, security checkpoint staffing), and information such as airport routing, flight, and weather information, etc.

2.1.1.5 UC.1.5 Insights for city commercial businesses.

Added value services will be offered across specific areas of interest in the city of Athens based on the analysis and prediction of the load expected within the city, informing local businesses on how people are expected to move within the commercial zones of the city aiming at improving local entrepreneurship and boost businesses turnover, and delivering services that can improve the mobility experience of the citizens and visitors of the city, which will lead to an optimization of mobility services In specific areas (e.g. dynamic PT timetables, less queuing on touristic sites) and could help the Public Transport Operator to adapt in the longer term its routes and timetables with respect to touristic/heavy load destinations.

2.1.2 Demonstration Hub #2— Energy

The energy demonstrator hub will focus 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. Four uCs are foreseen.

2.1.2.1 UC2.1: Increase the hosting capacity of the grid.

The increase in the Distributed energy resources (DERs) can cause problems to the grid because installations below 15kW can be connected without any licence and hence it is measures to ensure grid balance (also at infrastructural level) cannot be easily planned. Thanks to the data coordination the Distribution Service Operator (DSO), working closely with the Market Operator (MO) and the aggregators, can anticipate and manage these problems coming up with solutions related with the participation in global markets, introduction of the DERs and aggregators in the short-term flexibility market, also empowering the DSO to limit some of the DERs to protect the grid. Hence, there is a need of data interchange between

DSO, MO, and DERs, resulting from the negotiation of these short new term flexibility products, and one of the biggest advantages is that it can be automated.

2.1.2.2 UC2.2: Investment Deferral.

This UC focuses on investment reduction thanks to the data and the flexibility of the system. The most benefited actor is the DSO that can plan the grid operation, including the use of flexibility agents, and create a long-term local flexibility market. In this case, the market is planned to create commitments with the DSO for the following months or years, thus effectively enabling a new management tooling for the DSO. These data are interchanged over time thanks to the platform, being thus registered, and if it's necessary apply tradability or penalties, in validation process if necessary.

2.1.2.3 UC2.3: P2P trading between users or Energy communities.

Currently, in view of the promotion of self-consumption, shared self- consumption and the growing number of new energy communities, users are increasingly willing to be aware of the origin of the energy purchased from the grid and the destination of the energy fed into the grid. At present this information is unknown, beyond the guarantees of origin offered by some traders. A P2P trading is proposed to allow the user to choose which grid user sell to or buy from, thus enabling the creation of these peer-to-peer exchanges and, therefore, empowering the users to know exactly what they are consuming and where it comes from. This could be enabled by new platforms operated by the MO, where different resources trade energy in local level, but without the need to adjust to the grid ́s requirements coming from the DSO.

2.1.2.4 UC2.4: Monetisation of data

This UC will explore the possibility to monetise the data owned by the different actors participating in the energy exchange over the grid to third parties for use energy as a service (EaaS). The potential consumers are aggregators, installers, energy service companies, retailers, consultancy or advisory companies, private research groups and universities, EV charging companies, software companies, that would be able to optimise their existing services and develop new ones thanks to the additional information obtained through the platform.

2.1.3 Demonstration Hub #3— Automotive

The automotive demonstrator hub will support environmentally friendly, safe, and efficient mobility and transport. Specifically, the mobility hub will use various data sources, e.g. connected vehicle data from car manufacturers (via CARUSO, data provider operating a data marketplace), vehicle trip data (via VIF, Europ''s largest RTO for virtual vehicle technology), weather data (via UBIMET), map data (OpenStreetMap), and air quality data (open government data) to focus on traffic quality assessment in urban areas and driving style and driving risk assessment. The two use cases in this Hub will use the data and provide concrete data-driven services to individual drivers (driver warning and coaching), businesses (corporate mobility management for green driving), and public administrations (urban emission modelling, risk hotspot analysis).

2.1.3.1 UC3.1: Traffic quality assessment.

Driving patterns of vehicles will be analysed in their temporal, spatial and situational context using actual Trip Data (logged with a smartphone app owned by VIF) and vehicle sensor data (connected car data via CARUSO). From this data, specific information is calculated, e.g., acceleration, speed, and fuel consumption patterns along a specific route at certain times and their correlation with weather and air quality. These insights are used as input parameters for urban analytics applications facilitating planning processes, such as emission models or the hot-spot detection of stop-and-go patterns. Together with the driving style and risk assessment analytics, the generated information is also used for incentivising green driving styles and mobility decisions (such as reducing the share of motorised individual transport in the modal split of a commute) and optimising potentials for greener transport in the context of a corporate mobility management. Novel urban analytics components will be prototyped, including input parameters for emission models and traffic quality parameters, as well as corporate mobility management services and mechanisms for companies to promote more sustainable driving styles.

2.1.3.2 UC3.2: Driving Style & Risk Assessment.

The assessment of driving style and driving risk by using multiple data sources is a major topic. The driving risk is computed by analysing data from connected vehicles, data about past accident hotspots, past braking behaviours of car drivers, and weather data, to name a few examples. Thereby driving-risk relevant events are extracted from the data and form the input for a driving risk assessment pipeline. The individual driving style serves as another input to the risk assessment. Thereby drivers are categorised according to their driving styles, based on safety-relevant events such as speeding, phone use, use of assistance systems or reckless driving derived from the vehicle's sensor data while driving is weighted in terms of severity using contextual data (such as weather data). In a second step, this data is used for driver coaching, applying gamification approaches such as awarding points or ranking drivers according to their personal driving style to create competition for the most environmentally friendly and safest drivers.

2.2 METHODOLOGY

The approach to define the main operation principles of the PISTIS solution and to detail the context in which the Demonstrators Hubs will operate consisted in four phases, as follows.

1. A desktop study to identify and organise the actions data providers and data consumers typically need to perform.

2. A questionnaire for the participants in the Demonstrator Hubs to start identifying their needs, pain points and desirable solutions.

3. A set of workshops with the participants in the Demonstrator Hubs to investigate more in detail some of the aspects mentioned in their answers to the questionnaire.

4. A wrap-up round of discussion with all PISTIS partners to summarise the main information collected so far.

In the following, the discussion refers to the exchange of data between providers and consumers. However, it should be noted that in the PISTIS Demonstrator Hubs some participants might be interested in the possibility of making available, or exploiting, also services/applications. Therefore, while the terminology used in this context focuses on data for the sake of brevity (e.g., “data providers” and “data consumers” and analogous terms), the term should be interpreted in a broader sense.

Desktop study. In the first phase, the available information was collected about the actions data providers and data consumers typically need to perform in general when they want to conduct a transaction. The purpose of this study was to formally define the general process through which the different actors undergo when they wish to complete the transaction and identify those that could be performed through the PISTIS platform or facilitated by specific functions implemented in such platform. These actions describe the “Data Lifecycle” from the point of view of the potential users of the PISTIS platform. Table 1 presents the identified actions performed by data providers/data owners and data consumers, respectively, each time they want to conduct a transaction. A description of the process through which we iteratively defined the various stages of the Data Lifecycle is discussed in some detail in Section 2.4.1.

Questionnaire. The second phase consisted in identifying the information required to put the general outline of the Data Lifecycle into the context of the real problems that need to be addressed by the UCs in the Demonstrator Hubs. In particular, the activities focused in formulating a set of questions to be asked to the participants in the UCs to understand the characteristics of the problem they want to address and of their desired solution. Ideally, the answers to these questions are expected to provide the necessary information to identify the most appropriate technical solutions to be implemented on the PISTIS platform.

Table 1 List of actions performed by the potential data owners/providers and data consumers, i.e., the actors potentially interested in a transaction to sell and buy data, respectively.

	Actions of data providers	Actions of data consumers
Actions performed before a contract is drafted	Data Ingestion, Transformation and Treatment: • Data Check-In: The collection of data from the PISTIS system through various options (e.g., APIs, Pub/Sub, etc.) • Data Enrichment: the cleaning of data from errors and/or inconsistencies and the matching of ingested data to a common model for interoperability purposes • Analytics/Insights Engine: The application of some ready-made analytics on the data, to extract some information. • Data Lineage Tracking: The application of tracking tags on the data for allowing tracking of the subsequent actions. • GDPR Checker: The evaluation of whether the data contains GDPR relevant information and the suggestion to strip (if wanted) such information from the dataset (or to change it if needed) prior to exchanging it with other stakeholders. • Data Anonymization: Application of anonymization techniques	Data Exploration • Data Navigation/Querying: The search functionality of PISTIS system for search of data in the federated repositories according to specific parameters • Data Matchmaking Services: The recommendation of data based on various aspects such as previous searches, related terms, etc.

	• Data Quality Assessment: The assessment of the data for extracting indexes that can describe the quality in different dimensions. • Data Storage: The storage already treated data back to the original data storage facilities, keeping “pointers” at the PISTIS facility	
	Preparation for data publication • Access Policies Definition: Mechanisms for the application of policies for the access level on stored data of PISTIS system • Searchable Encryption: Application of Searchable Encryption for specific data samples to be published online. • Data and Metadata Publication: The publishing of metadata of the treated datasets in the federated PISTIS repositories for the allowing their querying and the publication of a small set of Data online as an example to be displayed to interested stakeholders	
	Data Monetary Value Estimation/Tagging: • Value Suggestion based on the XAI Engine • Data Asset Description (for enabling Sharing) • Definition of Monetisation Scheme, e.g., one or a combination of: • Subscription or One-Off Purchase Design Scheme • Data “Equity” Offering • NFT Generation	
	Data Contract Preparation • Contract Drafting: The initiation of a data contract between a data provider and a data consumer for data trading. • Contract Notification: The notification regarding the contract’s fate	
Actions executed after a contract is drafted	Data Contract Execution • Contract Negotiation: The negotiation between a data provider and a data consumer for the contract terms of data trading • Contract Signing: The finalization of a contract between a data provider and a data consumer	
	Data Encryption • Full Dataset Encryption: The full encryption of the Dataset prior to sending it (if required)	Data Acquisition: • Data Transfer: The acquisition of the dataset directly from the data owner • Data Decryption: The decryption of the acquired dataset, using the keys provided
	Peer-to-Peer Data Exchange • Data Export via different methods: The selection of how data shall be sent to the recipient (also based on the contract terms)	
Other actions	Transactions Monitoring • Auditing of Transactions: An interface where a user can have a log of his transactions.	

	• Auditing of On/Off Platform Usage: An interface where the user can witness how the data he has traded is used	

The first round of brainstorming resulted in a list of 55 questions addressing all relevant aspects to identify the data landscape, and the functional, non-functional, technical and user requirements about the data that needs to be shared across the participants in the uCs and the mechanisms to enable the transactions. These aspects include for example the type and format of data that are available to be exchanged, the actions that need to be performed on such data (e.g. pre-processing, encryption, etc.), the necessary mechanisms to protect and anonymise sensitive information, the characteristics of the users that will have access to the data, the timing at which the data needs to be available, the trading schemes, and so on.

The preliminary list of questions was subsequently revised with the aim of clustering them around the following fundamental topics (as discussed later in Section 2.4.1, these corresponded to the preliminary components of the simplified Data Lifecycle):

- Problem definition. It includes the description of the need(s) which the Use Case participants want to address, and which drives their motivation to use the PISTIS platform.
- Data identification and collection. It includes the characterisation of the type of data necessary to address the user needs and the process to collect such data. The purpose is to investigate whether this data is available, whether someone within the organisation of the Use Case participant is already collecting it and, if not, whether there are technical limitations (e.g. additional sensors have to be installed that currently do not exist, a process has to be put in place to enable the recording of the data, etc.) or regulatory constraints preventing the collection of such data.
- Data analysis. It includes information about the manipulation, processing, and analysis that must be performed on the data to extract the necessary information to address the user needs and solve the original problem.
- Data sharing. This covers the mechanisms and processes that need to be put in place to transfer information across UC participants in the case one user needs access to some specific data which could in principle be made available by another user of a different Organisation.
- Data monetisation. This assesses the value attributable to the data both by the data owner (whose goal is to capitalise on the data they produced) and consumers (who can estimate the extent to which the additional data can potentially improve their business).
- Data acquisition. This covers all aspects relating to possible security and privacy issues that might arise from the acquisition, use, and exchange of the data of interest.

To clarify these aspects, a questionnaire was prepared and shared with the participants in the three Demonstrator Hubs. Table 2 shows the questions presented to the participants for each UC. The rationale is that different participants will provide their point of view of data providers or data consumers, and the combination of their feedback provides a picture of the problems and desirable technical and procedural solutions from different perspectives.

Table 2 Questionnaire presented to the Use Case participants.

	Questions
INTRODUCTION	• Please specify for which Use Case you are filling the questionnaire. • Please provide the name of your organisation and email contact
PROBLEM	• Describe what is the problem you would like to solve within the use case
DATA COLLECTION	• What type of data do you need? • Is the data you need available? If the data you need is available, from whom would you get it? If the data is not available, could you explain why (e.g., technical/regulatory/timing/etc. issues)? • Do you require the data to comply with a specific format/standard? • Does the data you need contain any information associated with an identified or identifiable natural person, e.g., personal data? (Personal data could be, for example, a name, location, identification number, or physical, psychological, or social attributes of a person.) • How do you evaluate the quality of the collected data?
DATA ANALYSIS	• Once the data has been collected, how would you use that (e.g., combining it with other data source? analyse the data?) • If data analysis would be required, what kind of analysis? • Is pre-processing required (e.g., transformations)? • If so, what kind of pre-processing would it be required? • Is combination with other sources required? (Owned or external?)
DATA SHARING	• Are you able to retrieve data from other organizations? • If you can retrieve data from other organisations, how often do you need to retrieve them? • If you can retrieve data from other organisations, how do you retrieve them (e.g., is there any standards for data modelling)? • If you are not able to retrieve data from other organisations, what is preventing the data sharing? (e.g., technical/ regulatory/ etc. issues) • Considering the data that you directly collect; do you think they would be useful for anyone else? If so, to whom? • What requests you have currently for sharing your own data? • What is the type of data you are willing to share (structured, binary?)
DATA MONETISATION	• What would you need to know to capitalise on the data/knowledge produced/co-created? • If you could have the data you need, would you be able to estimate how much is worth for your business? If so, on what basis? • How would estimate the value of data you share? • What is going to be the pricing plan for your service?

DATA ACQUISITION	• Do you handle data with GDPR restrictions / personal data of humans? • Do you have access policies for your data? If so, how do you expect these policies will apply to the PISTIS access? • What data anonymization techniques would you need to be applied? • Are your data searchable? • Do you need to transfer to or share the data with any third party from outside the EU/EEA? (E.g., the USA, the UK) • What are the technical and organisational measures in place or to be implemented to ensure the security of the data processing/sharing activities? • Is any law or regulation preventing you from sharing data with partner companies, or preventing other organisations from sharing their data you need? • What kind of security and privacy issue do you see from the use of such data? • What kind of measure shall we put into place to mitigate the risk?

Workshops. The rationale of the questionnaire was to generate a preliminary, high-level description of the potential users of the PISTIS platform, their needs, and motivations to adopt it, and their expectations about the kind of interactions they can have with it and the information they would like to extract from it. The subsequent phase was dedicated to discussing in more detail these aspects to generate a set of end-to-end scenarios that describe the process through which each Use Case participant will be able to interact with the platform. This discussion took place in the form of a series of interactive workshops in which participants were asked to provide their insight on their motivations to use the PISTIS platform, their expectations on the information and services the platform should be providing, and the process they envisage to before, during and after the interaction with the platform. Specifically, four workshops were organised, two for the Demonstration Hub #1 – Mobility and Urban Planning (that has five Use Cases), and one for each of the Automotive and Energy Demonstration Hubs.

The workshops activities were structured as follows. First, a brief introduction of the Use Cases under discussion was provided. Subsequently, the participants accessed a virtual whiteboard[1] and, individually, provided their insight about the Use Cases by answering the questions shown in Table 3. In the first workshop of the series, which focused on the two Automotive Use Cases, the questions were structured and formulated slightly differently, as they were clustered around the main stages of the simplified Data Lifecyle rather than on the different phases of the interaction with the PISTIS platform, as in Table 3. However, the goal of the exercise was the same: gather the information about the interaction process with the PISTIS platform from the point of view of different stakeholders through a list of question prompts. In the final segment of the workshops, the participants' answers were briefly reviewed and discussed. Because of time constraints, participants were not able to fill in their feedback on all Use Cases and were therefore encouraged to do so independently in the following days. The collected information collected was in any case sufficient to generate a high-level description of the end-to-end scenarios of use for each Use Case, especially because the envisaged process and foreseen challenges are often similar, with the main differences between Use Cases of the same Demonstration Hub coming from the specific type of data exchanged for different purposes. The end-to-end scenarios of all Use Cases are presented in Sect. 2.3. The Miro boards filled in at the workshops are shown for reference in Appendix A.

Table 3 Question prompts and rationale presented on the Miro virtual whiteboards during the workshops.

Question prompt	Rationale
What motivated you to use the platform?	Captures what has motivated the user to come to the platform, what drives the user to seek the product.
What must the PISTIS platform provide to you?	Articulates what the user wants from the platform and therefore what the platform must provide to the leave the user feeling satisfied.
What must you do before you make an interaction with the PISTIS platform?	Defines the event which occurred before the user contacts the platform
How do you expect to achieve your goal when interacting with the PISTIS platform? (As a sequence of events)	Defines the event(s) which occurs in the interaction between the user and the platform at the user- interface
What outputs do you expect to receive from the PISTIS platform?	Defines the output that the platform provides as a response to the user input
What information do you require from the PISTIS platform?	Is there an information requirement for this event to occur?
Is there a decision point during your interaction with the PISTIS platform?	Is there a decision point that needs to be made for the event to occur?
What will you do after you have finished making an interaction with the PISTIS platform?	Define the event which occurs after the user contacts the platform
What do you think the challenges or complexities will be when using the PISTIS platform?	Captures the problems that the integration of the platform might lessen or worsen or remedy. Details points in the system that need to be carefully managed/integrated during user experience with the new platform.
What do you think should be done to guarantee that data are effectively exchanged by users within the PISTIS platform?	Captures the user' suggestions for solutions to identified challenges/ complexities.

Wrap-up discussion. The last phase of the process consisted in a round of discussion with the entire PISTIS consortium devoted to summarising the main information collected through the questionnaire and the workshops. In this discussion, the participants in the Demonstrator Hubs provided additional details about the data landscape and availability of their Use Cases, the actors involved, focusing on identifying who plays the roles of data providers and data consumers, and the way the different actors are expected to interact through the PISTIS platform.

The results of this four-phases process are the end-to-end scenarios presented in Section 2.3.

2.3 End to End Scenarios in the Context of PISTIS Demonstrators

Table 4 UC 1.1 - Baggage handling management

Mobility and Urban Planning Hub – Use Case 1.1: Baggage handling management			
Goal	To improve operational performance of baggage handling, reduce irregularities in the process and predict the probability of flights delays due to baggage handling. Efficient baggage handling reduces the risk of delays and coordination between the airport and ground handling would help mitigate operational anomalies.		
Problem	Lack of exchange of operational information across airport stakeholders slows down the baggage management process, with potential delays as cascading effect.		
PISTIS role	Provide a method for standardised communication and effortless exchange of data between stakeholders, in particular the Airport and the Ground Handler who are the main actors but with additional data coming from other providers of information, within agreed key performance indicators (KPI) and Service Level Agreements (SLA).		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	Define relevant data contracts, details, terms of use, pricing policies.	All	All
	Examining relevant registry available data sources and identifying the data that will be needed.	AIA, OAG	Goldair
	Agreement on conditions for data sharing	All	Data consumers
	Data quality assessment, transformation & analytics	Data providers	Data consumers
	Define licensing and policies to use the data	AIA, OAG	Data consumers
Step 1 - Input provided to PISTIS	Decision to share data (entire dataset or preliminarily a sample, as deemed necessary) through PISTIS	Data providers	Data consumers
	Arrival/Transfer/Departing Bag timestamps (AIA BHS system) data	AIA	Goldair
	Minimum connecting times data	AIA, Goldair	Goldair
	Flight schedule and day of operations updates data	AIA	Goldair
	Force majeure and operational irregularities data	AIA	Goldair
	Public Transport Scheduling Data	OASA	Goldair, AIA
	Transmit weather data through PISTIS	UBIMET	AIA
Step 2 - Output retrieved from PISTIS	Check example data for data format, data characteristics, compliance with requirements and standards, match with user needs, etc.	Goldair, AIA	Data providers
	Retrieve data	Goldair, AIA	Data providers
Step 3 - Data quality assessment	Data quality assessment, data decryption and - if necessary - preprocessing	Goldair, AIA	-

and generation of new analytics (through PISTIS or offline)	Analysis and combination of traffic data to derive predictive analytics and optimise baggage transfers	Goldair	Goldair, AIA
	Analysis and combination of traffic and weather data to compute predictive analytics of potential delays	Goldair, AIA	Goldair, AIA
	Data providers monitor that data is correctly transferred and that it is used as agreed in the licences	Data providers	All
	The new analytics are made available through PISTIS if requested	Goldair, AIA	All
Step 4 - After the interaction with PISTIS	Fine tuning data usage, evaluate the businesses value of the data received, identify additional data to increase business value.	AIA	-
	Develop algorithms and business logic that will use the input to provide useful and exploitable results, feeding internal systems with data.	Goldair	
	Improving internal processes, Exploring data sharing with third parties.	AIA, Goldair	Data consumers

Table 5 UC1.2 - Transfer Passenger Management

Mobility and Urban Planning Hub – Use Case 1.2: Transfer Passenger Management			
Goal	To optimise allocation of aircraft stands for specific flights and minimise passengers’ connection times.		
Problem	The delayed transfer passengers have an impact to overall airport operations and the respective stakeholders (airline, ground handler) performance.		
PISTIS role	To provide real-time data to enable better staff allocation and process improvement and to improve passenger experience.		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	Define relevant data contracts, details, terms of use, pricing policies.	All	All
	Examining relevant registry available data sources and identifying the data that will be needed.	AIA, OAG	Goldair
	Agreement on conditions for data sharing	All	Data consumers
	Data quality assessment, transformation & analytics	Data providers	Data consumers
	Define licensing and policies to use the data	Data providers	Data consumers
Step 1 - Input provided to PISTIS	Decision to share data (entire dataset or preliminarily a sample, as deemed necessary) through PISTIS	Data providers	Data consumers
	Flight status data, handler data-stamps, arrival/departure/transfer Bag number	AIA, OAG	Goldair
	transfer passenger number and destination, PTM Message, type of connection	Goldair	AIA

	PRM Passengers, Immigration – Customs Clearance numbers and waiting times		AIA, Goldair
	Minimum connecting times data	AIA, Goldair	Goldair
	Flight schedule and day of operations updates data	AIA	Goldair
	Weather data	UBIMET	AIA
Step 2 - Output retrieved from PISTIS	Check example data for data format, data characteristics, compliance with requirements and standards, match with user needs, etc.	Goldair, AIA	Data providers
	Retrieve data	Goldair, AIA	Data providers
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Data quality assessment, data decryption and - if necessary - preprocessing	Goldair, AIA	-
	Analysis and combination of traffic data, minimum transfer data, Customs Clearance processing times, etc. to derive predictive analytics and optimise passenger transfers	Goldair	Goldair, AIA
	Analysis and combination of traffic data, minimum transfer data, Customs Clearance processing times, and weather data to compute predictive analytics of potential delays	Goldair, AIA	Goldair, AIA
	Data providers monitor that data is correctly transferred and that it is used as agreed in the licences	Data providers	All
	The new analytics are made available through PISTIS if requested	Goldair, AIA	All
Step 4 - After the interaction with PISTIS	Fine tuning data usage, evaluate the businesses value of the data received, identify additional data to increase business value.	AIA	-
	Develop algorithms and business logic that will use the input to provide useful and exploitable results, feeding internal systems with data.	Goldair	
	Improving internal processes, Exploring data sharing with third parties.	AIA, Goldair	

Table 6 UC 1.3 - Aircraft turnaround process

Mobility and Urban Planning Hub – Use Case 1.3: Aircraft turnaround process	
Goal	Optimise processes to achieve an increasingly efficient turnaround process and avoid delays compared to target off-block time (TOBT)
Problem	Limited shared information, including about weather, across airport and ground-handling stakeholders might result in an inefficient process of aircraft servicing with potential delays compared to TOBT and thus consequences for the regularity of the flight
PISTIS role	To provide a platform for the exchange of real-time data that enables better coordination during the aircraft turnaround process.

Actions	Description	The action is performed by...	The action impacts...
Actions before interacting with PISTIS	Define relevant data contracts, details, terms of use, pricing policies.	All	All
	Examining relevant registry available data sources and identifying the data that will be needed.	AIA, OAG, UBIMET	Goldair
	Agreement on conditions for data sharing	All	Data consumers
	Data quality assessment, transformation & analytics	Data providers	Data consumers
	Define licensing and policies to use the data	Data providers	Data consumers
Step 1 - Input provided to PISTIS	Decision to share data (entire dataset or preliminarily a sample, as deemed necessary) through PISTIS	Data providers	Data consumers
	Flight status data, flight schedule, type of aircraft	OAG	AIA, Goldair
	Handler data stamps	Goldair	AIA
	Aircraft registration, Flight Type, Aircraft parking stand, Boarding gate, Landing, and taxi –intime, In-block and off-block time, Taxi-out and take off time, Turn round times	AIA	AIA, Goldair
	De-icing information, Aircraft movement data, TOBT updates, MTB messages, Operation planning information, Passengers numbers	AIA	AIA, Goldair
	Force majeure and operational irregularities	AIA	Goldair
	Weather data	UBIMET	AIA
Step 2 - Output retrieved from PISTIS	Check example data for data format, data characteristics, compliance with requirements and standards, match with user needs, etc.	Goldair, AIA	Data providers
	Retrieve data	Goldair, AIA	Data providers
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Data quality assessment, data decryption and - if necessary – preprocessing	Goldair, AIA	-
	Analysis and combination of all sources of data to compute predictive analytics, identify bottlenecks and optimise the aircraft turn around processes	AIA	Goldair, AIA
	Data providers monitor that data is correctly transferred and that it is used as agreed in the licences	Data providers	All
	Provided improved and analytics concerning weather conditions and forecasts	UBIMET	AIA, Goldair
	The new analytics are made available through PISTIS if requested.	Goldair, AIA	All

Step 4 - After the interaction with PISTIS	Fine tuning data usage, evaluate the businesses value of the data received, identify additional data to increase business value.	AIA	-
	Use individual live flight tracking to anticipate delays.	OAG	-
	Develop algorithms and business logic that will use the input to provide useful and exploitable results, feeding internal systems with data.	Goldair	-
	Improving internal processes, Exploring data sharing with third parties.	AIA, Goldair	-

Table 7 UC 1.4 - Public transportation planning support

Mobility and Urban Planning Hub – Use Case 1.4: Public Transportation Planning Support			
Goal	To design a better service for passengers from and to the airport and optimisation of the public transport service for the airport.		
Problem	There is a need for better scheduling of public transport routes, deployment of vehicles and the optimization of fleet utilization, to meet the need for improved commuter services, especially during peak season, and the reduction of operational costs.		
PISTIS role	To facilitate data trading and sharing between various sources such as the airport (AIA) and the city (DAEM) to improve the overall planning of the transportation of OASA, allowing the development of services that are able to predict the accuracy of the load within the day in order to better schedule routes, deploy vehicles and maximise the utilization of its fleet, offering to commuters better services and lowering operational costs.		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	Define relevant data contracts, details, terms of use, pricing policies.	All	All
	Examining relevant registry available data sources and identifying the data that will be needed.	AIA	DAEM, OASA
	Ensure that PISTIS use has interoperability with airport internal systems	AIA	All
	Data anonymisation and removal of business sensitive information	AIA, OASA	All
	Semantic enrichment of the data to make it searchable, also using keywords	AIA, OASA	Data consumers
	Agreement on conditions for data sharing	All	Data consumers
	Data quality assessment, transformation & analytics, licensing, and policies to use the data	Data providers	Data consumers
Step 1 - Input provided to PISTIS	Decision to share data (entire dataset or preliminarily a sample, as deemed necessary) through PISTIS	Data providers	Data consumers
	Public transport timetables (buses and metro) data	OASA	AIA

	Public transport vehicle data and occupancy (static) data	OASA	AIA
	Metro station incoming / outgoing passengers' data and bus geolocation data (currently available for previous day)	OASA	AIA
	Historic data (e.g., flight schedules, gate/terminal usage), inbound/outbound passenger per hour	AIA	OASA, DAEM
	Timestamps, Aircraft load factors	GOLDAIR	DAEM, OASA, AIA
	Transport modal split	OASA	OASA, AIA
	Weather	UBIMET	DAEM, OASA, AIA
Step 2 - Output retrieved from PISTIS	Check example data for data format, data characteristics, compliance with requirements and standards, match with user needs, etc.	AIA, OASA, DAEM	Data providers
	Retrieve data on public transport movements and occupancy	AIA, OASA, DAEM, ICCS	Data providers
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Data quality assessment, data decryption and - if necessary - preprocessing	AIA, OASA, DAEM	-
	Compute predictive analytics to optimize vehicle availability from/to the airport	OASA, ICCS	AIA
	Data providers monitor that data is correctly transferred and that it is used as agreed in the licences	Data providers	All
	Set up algorithms to receive specific data sets automatically on regular intervals (e.g., daily update of the datasets/operation plan).	ICCS	-
	The new analytics / services are made available through PISTIS if requested	Data providers	All
Step 4 - Output retrieved from PISTIS	Receive data on expected visitors flows based on destination locality	DAEM	ICCS, OASA
Step 5 - After the interaction with PISTIS	Improve planning and decision making of public transport	OASA	AIA, OASA, DAEM

Table 8 UC1.5 - Insights for city commercial business

Mobility and Urban Planning Hub – Use Case 1.5: Insights for City Commercial Businesses	
Goal	Enable data exchange between the airport and city to offer better services to citizens, to improve local entrepreneurship in Athens according to expected loads in specific areas and to leverage the mobility/touristic experience of the city according to foreseen mobility flows.
Problem	Lack of exchange of (open) data between stakeholders affected by mobility at the airport and in different parts of the city prevents the development of added value services across specific areas of interest in the city and improvements to local entrepreneurship, mobility experience and mobility services.

PISTIS role	Facilitate access to data from the partners of the Greek cluster and other external data sources, e.g., potentially touristic sources, GIS data and businesses data.		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	Define relevant data contracts, details, terms of use, pricing policies.	All	All
	Data anonymisation and removal of business sensitive information	Data providers	All
	Semantic enrichment of the data to make it searchable, also using keywords	Data providers	Data consumers
	Anonymisation of the city businesses registry and provide it for export	DAEM	Data consumers
	Agreement on conditions for data sharing	All	Data consumers
	Data quality assessment, transformation & analytics, licensing, and policies to use the data	Data providers	Data consumers
Step 1 - Input provided to PISTIS	Decision to share data (entire dataset or preliminarily a sample, as deemed necessary) through PISTIS	Data providers	Data consumers
	Timestamps	OAG	AIA
	Mobility/touristic/business registry	AIA, OASA, ICCS, DAEM	DAEM
	Number of expected visitors	AIA, OASA, ICCS	DAEM
	Destination of visitors	AIA, OASA, ICCS	DAEM
	GIS info on location, district, address, floor	DAEM	DAEM
Step 2 - Output retrieved from PISTIS	Check example data for data format, data characteristics, compliance with requirements and standards, match with user needs, etc.	DAEM, OASA, ICCS, DAEM	Data providers
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Data quality assessment, data decryption and - if necessary - preprocessing	AIA, OASA, DAEM	-
	Compute analytics to predict expected visitors flows based on destination locality	OASA, ICCS	AIA
	Data providers monitor that data is correctly transferred and that it is used as agreed in the licences	Data providers	All
	The new analytics / services are made available through PISTIS if requested	OASA, ICCS	DAEM
Step 4 - Output retrieved from PISTIS	Receive data on expected visitors flows based on destination locality	DAEM	ICCS, OASA
Step 5 - After the interaction with PISTIS	Notify local businesses.	DAEM	
	Create an open public call for local businesses to participate, to provide feedback.	DAEM	

Table 9 UC 2.1 - Increase the hosting capacity of the grid

Energy Hub – Use Case 2.1: Increase the hosting capacity of the grid			
Goal	Increase hosting capacity of the grid by accommodating DERs smoothly, also negotiating new short-term flexibility products		
Problem	Distributed grid with multiple DER "prosumers" creates challenges: for energy offer to meet demand, coordination is necessary between DSO, aggregators, and MO, to enable technical and market mechanisms that prevent grid congestions and protect the grid.		
PISTIS role	Provide an environment where data, services, and analytics are exchanged between the DSO, the MO, the aggregators, and service providers about the static topological properties of the network and the dynamic use of energy (demand and offer), so that optimised technical and short-term flexible market mechanisms can be identified to guarantee an efficient use and distribution of the resources across the network		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	Agree on a Common Information Model (CIM), data format and protocols	All	All
	Prepare data on grid topology, DERs location, DERs generation, historical data on grid events and user consumption, preprocessing to produce forecasts	CUERVA	-
	Semantic enrichment of the data to make it searchable	CUERVA, UBIMET	-
	Data quality assessment, transformation & analytics	CUERVA, UBIMET	-
	Price signals (before the flexibility market celebration) and flexibility market results (after the flexibility market celebration)	OMIE	CUERVA, BAMBOO, CARTIF
	Define licensing and policies to use the data	All	All
	Implement APIs/mechanisms to automatically transmit data, also in real time if necessary	Data providers	-
	Implement APIs/mechanisms to automatically retrieve data, also in real time if necessary	Data consumers	-
	Check and eliminate unnecessary GDPR-relevant information in data	Data providers	Data consumers
	Anonymise study-relevant personal data (e.g., energy consumption, DER production and location)	Data providers	Data consumers
	Data publication preparation	Data providers	Data consumers
	Data value estimation and monetisation scheme definition	Data providers	Data consumers
	Identify flexible assets in the grid	CUERVA	-
Step 1 - Input provided to PISTIS	Decision to share data (entire dataset or preliminarily a sample, as deemed necessary) through PISTIS	Data providers	Data consumers

	Transmit grid data through PISTIS	CUERVA	BAMBOO, CARTIF
	Transmit weather data through PISTIS	UBIMET	CUERVA, BAMBOO
	Provide price signals (before the flexibility market celebration)	OMIE	Data consumers
	Provide flexibility market results (after the flexibility market celebration)	OMIE	Data consumers
	Transmit data from field sensors and IoT		CUERVA
Step 2 - Output retrieved from PISTIS	Check example data for data format, data characteristics, compliance with requirements and standards, appropriate match with user needs, etc.	BAMBOO, CARTIF	CUERVA, UBIMET
	Retrieve grid data (including both network topology, prosumers data, energy consumption, etc)	BAMBOO, CARTIF	CUERVA
	Retrieve weather data	BAMBOO, CUERVA	UBIMET
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Data quality assessment, data decryption and - if necessary - preprocessing	BAMBOO, CARTIF, CUERVA	
	BAMBOO uses grid and weather data and uses it to compute analytics and predictions on flexibility	BAMBOO	OMIE
	CARTIF uses grid data to analyse hosting capacity	CARTIF	CUERVA
	CUERVA uses weather data to improve predictions of energy generation	CUERVA	CUERVA
	Data providers monitor that data is correctly transferred and that it is used as agreed in the licences	Data providers	All
	The new analytics are made available through PISTIS	BAMBOO, CARTIF	All
Step 4 - Output retrieved from PISTIS and generation of new analytics (through PISTIS or offline)	Data quality assessment, data decryption and - if necessary - preprocessing	OMIE	
	OMIE joins flexibility necessities (from DSO, to manage the grid) with flexibility providers (gathered by an aggregator)	OMIE	CARTIF, CUERVA, BAMBOO
	OMIE creates short-term flexibility markets (auctions, under the CUERVA call) from assets information (location, among other) and their bids	OMIE	CARTIF, CUERVA, BAMBOO
	CARTIF combines and analyses grid data, prosumers data, flexibility validation data to generate new insights and predictions of network hosting capacity	CARTIF	CUERVA
	Data providers monitor that data is correctly transferred and that it is used as agreed in the licences	Data providers	All
	The new analytics are made available through PISTIS	OMIE, CARTIF	All

Step 5 - After the interaction with PISTIS	CUERVA uses the analytics and predictions generated by BAMBOO, CARTIF to create action plans for upgrading the network, leveraging flexibility to reduce network events in the context of increased DERs connections	CUERVA	CUERVA, BAMBOO, CARTIF, OMIE
	OMIE shares the results of the flexibility markets with all the participants (CUERVA as the requester and providers who have resulted assigned in the auction), so that CUERVA could use them to optimise the use of the grid.	OMIE	All

Table 10 UC 2.2 - Investment deferral

Energy Hub – Use Case 2.2: Investment Deferral			
Goal	The focus is on studying the impact of utilizing data and flexibility in grid development as an alternative to traditional approaches that address issues like overvoltage and congestions. There is the need to facilitate the exchange of valuable information and resources, leading to more efficient grid management and cost savings in the energy sector.		
Problem	Challenges emerged on the data format compatibility among parties, communication protocols, interoperability issues, lack of a standardized Common Information Model (CIM). Ensuring data availability from the Distribution System Operator (DSO) is also a problem to be analysed		
PISTIS role	PISTIS platform will provide support to the DSO in the process of planning grid operations, and thus allowing the creation of a long-term local flexibility market and make commitments with flexibility agents.		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	To find agreement on the type of information, e.g., data format, that will be exchanged, communication protocols and other technical requirements.	OMIE	-
	Data Enrichment	CUERVA	-
	Data Anonymization	CUERVA	-
	Data Quality Assessment	All	-
Step 1 - Input provided to PISTIS	Transmit grid topology data	CUERVA	CARTIFF
	Transmit grid investment data	CUERVA	CARTIFF
	Transmit grid events (historical) data	CUERVA	CARTIFF
	Transmit user consumption (historical) data	CUERVA	CARTIFF
	Transmit DERs generation data	CUERVA	BAMBOO
	Provide historical data (results, prices) of previous auctions celebrated on these congested zones (before the flexibility market celebration)	OMIE	All

	Provide flexibility market results (after the flexibility market celebration)	OMIE	All
	Transmit flexibility aggregated data	BAMBOO	OMIE/CARTIFF
	Data Ingestion (tool + several connectors to important data sources/ data spaces)	All	All
Step 2 - Output retrieved from PISTIS	Data Ingestion (tool + several connectors to important data sources/ data spaces)	-	-
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Analytics/Insights Engine (the tool and scripts for data quality analysis)	-	-
	UC specific data Analytic for investment economic data	-	-
Step 4 - Output retrieved from PISTIS and generation of new analytics (through PISTIS or offline)	Data and Metadata Publication	-	-
	Data Trading and acquisition from/to another PISTIS user	-	-
Step 5 - After the interaction with PISTIS	Digital Twin creation	CARTIFF/ CUERVA	-
	Flexibility market results	OMIE	All
	Report system for the results	CARTIFF/ CUERVA	-
	Long-term Flexibility forecasting	CARTIFF/ CUERVA	-
	UC specific data Analytic for investment economic data	CARTIFF/ CUERVA	-
	Data scaling	CARTIFF/ CUERVA	-
	App / interface for the DSO	CARTIFF/ CUERVA	-

Table 11 UC 2.3 - P2P Trading between users or energy communities

Energy Hub – Use Case 2.3: P2P Trading between users or Energy communities			
Goal	The aim is to create peer-to-peer exchanges with new platform operated by the MO to facilitate free negotiation among energy resources/assets (i.e., without any DSO call)		
Problem	Challenges emerged on the data format compatibility among parties, communication protocols, interoperability issues, lack of a standardized Common Information Model (CIM). Ensuring data availability and communication among parties would empower the user and increase P2P exchange frequency		
PISTIS role	PISTIS platform will provide support to the partners in the enhancing and facilitation of data trading and communication among users		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	To find agreement on the type of information, e.g., data format, that will be exchanged, communication protocols and other technical requirements.	-	-

	Data Enrichment	-	-
	Data Anonymization	-	-
	Data Quality Assessment	-	-
Step 1 - Input provided to PISTIS	Transmit grid topology data	CUERVA	CARTIFF
	Transmit user consumption data	CUERVA	CARTIFF
	Transmit DERs generation data	CUERVA	BAMBOO
	peer to peer trading results	OMIE	all
	Data Ingestion (tool + several connectors to important data sources/ data spaces)	All	All
Step 2 - Output retrieved from PISTIS	Data Ingestion (tool + several connectors to important data sources/ data spaces)	-	-
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Analytics/Insights Engine (the tool and scripts for data quality analysis)	-	-
	UC specific data Analytic for investment economic data	-	-
Step 4 - Output retrieved from PISTIS and generation of new analytics (through PISTIS or offline)	Data and Metadata Publication	-	-
	Transmit flexibility aggregated data	BAMBOO	CARTIF
	Data Trading and acquisition from/to another PISTIS user	-	-
Step 5 - After the interaction with PISTIS	Digital Twin creation	-	-
	Report system for the validation of the transaction		-
	Power flow calculation	-	-
	Smart contract creation between users	-	-
	Use data to feed algorithms that will help the grid distributor to improve the management of the grid. To use data generated by the grid to solve problems not directly related to electricity	CARTIF	
	Storage of the contracts and validation result	-	-

Table 12 UC 2.4 - Monetisation of data

Energy Hub – Use Case 2.4: Monetisation of data	
Goal	The aim is to define the process of data valorisation to target potential consumers of energy assets who may use them for Energy as a Service (EaaS) including aggregators, installers, energy service companies, retailers, consultancy firms, research groups, universities, EV charging companies, and software companies

Problem	The challenge lies in acquiring adequate information to accurately assess the economic value of data, as it involves a complex analysis of not only data acquisition but also estimating future savings, to create a complete and adequate model.		
PISTIS role	PISTIS platform will provide support to the partners in the enhancing and facilitation of data trading and communication among users to inform the economic value of data		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	To find agreement on the type of information, e.g., data format, that will be exchanged, communication protocols and other technical requirements.	-	-
	Data Enrichment	-	-
	Data Anonymization	-	-
	Data Quality Assessment	-	-
Step 1 - Input provided to PISTIS	Transmit digitalization cost data	CUERVA	-
	Transmit grid O&M operation data	CUERVA	-
	Transmit data infrastructure cost data	CUERVA	-
	Data Ingestion (tool + several connectors to important data sources/ data spaces)	All	All
Step 2 - Output retrieved from PISTIS	Data Ingestion (tool + several connectors to important data sources/ data spaces)	-	-
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Analytics/Insights Engine (the tool and scripts for data quality analysis)	-	-
	UC specific data Analytic for investment economic data	-	-
Step 4 - Output retrieved from PISTIS and generation of new analytics (through PISTIS or offline)	Data and Metadata Publication	-	-
	Data Trading and acquisition from/to another PISTIS user	-	-
Step 5 - After the interaction with PISTIS	Establishing the value of data that will be monetised or traded in any of the use cases	Eurecat	-
	Definition of processes for the valorisation of data	Eurecat	-
	Studies dedicated to data valorisation	Eurecat	-
	Smart contract creation between users	-	-
	Use data to feed algorithms that will help the grid distributor to improve the management of the grid. To use data generated by the grid to solve problems not directly related to electricity.	-	

	Storage of the contracts and validation result	-	-

Table 13 UC 3.1 - Traffic quality assessment

Automotive Hub – Use Case 3.1: Traffic Quality Assessment			
Goal	Development of two demonstrators: urban analytics and corporate mobility management. Explore new datasets, including previously unknown sources, and establish mechanisms for selecting only the right sources that are truly relevant and available on a stable basis. A) Urban analytics Urban car traffic quality assessment. Combination of geospatial time series data to generate insights about traffic flows and potential correlation between data sources. Localize safety-critical zones, forecast safety risks, and plan congruent actions on both short and long-term horizons. B) Corporate mobility management Real-time transportation mode recommendations for commuters. Combination of individualized commuting data with available modes of transportation.		
Problem	Insufficient real time driving data (acceleration, speed, fuel consumption patterns) and their interrelation with weather poses challenges for urban traffic planning. Incentivizing green driving styles and optimizing greener transport options within corporate mobility management by using real time weather data.		
PISTIS role	Provide support to the integration and combination of diverse data sources, enabling informed decision and real-time analysis and prediction. Support to the integration of driving data, driving style and risk assessment, street graph data, and live weather data.		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	Pre-processing to fit the data into the data model of the software products, e.g., uniform time series.	Trafficon, CARUSO	-
	Agreement on conditions for data sharing	All	All
	Define licensing and policies to use the data	Data providers	Data consumers
	Data quality assessment, transformation & analytics	Data providers	Data consumers
	Integrating a road graph (e.g., OSM).	Trafficon	-
	Data harmonisation.	Trafficon	-
Step 1 - Input provided to PISTIS	Decision to share data (entire dataset or preliminarily a sample, as deemed necessary) through PISTIS	Data providers	Data consumers
	Processed data made available for querying through API.	Trafficon	-
	The Integration of floating car data (if available)	Trafficon	-
	Provision of driving/car and accidents data.	CARUSO	-
	Provision of safety data	VIF	Trafficon
	Provision of driving data	VIF	Trafficon

	Provision of live weather data	UBIMET	Trafficon
	Transmit vehicle sensor data/connected Car Data (fuel consumption, risk, driving styles)	CARUSO	Trafficon
	Transmit street graph data	OSM	Trafficon
	Transmit public transport data	GRAZ Holding	Trafficon
	Data visualisation dashboard for urban analytics and (Corporate) mobility management.	Trafficon	-
Step 2 - Output retrieved from PISTIS	Retrieve the data (sample or entire data set)	Trafficon, CARUSO	Data providers
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Implementation of data resolution, calculation of hotspots and calculation of correlations/recommendations.	Trafficon	-
	Data quality assessment, data decryption and - if necessary - preprocessing	Trafficon	-
	Combine spatial data and live traffic data	Trafficon	-
	Interact through the database through a web-app (frontend)	Trafficon	-
	Interact through a Service oriented Architecture (e.g., APIs)	Trafficon	-
	Monitoring system checking that data is correctly transferred and that it is used as agreed in the licences	Data providers	All
	The new analytics are made available through PISTIS	All	All
Step 4 - After the interaction with PISTIS	Data will be combined (e.g., weather data and traffic flows) and analysed, including statistical and geospatial analysis.	Trafficon	-
	Data will be modelled for user interaction through GUI.	Trafficon	-
	The results will be made available for platform users.	Trafficon	-
	Consider options on how to display risks or make risks visible to drivers.	CARUSO	-

Table 14 UC 3.2 - Driving style & risk assessment

Automotive Hub – Use Case 3.2: Driving Style & Risk Assessment	
Goal	Create a data-driven risk prediction and warning system for vehicle drivers that enhances driver awareness, reduces risky behaviours, and improves overall driving and road safety integrating different data sources from different data providers. Establish data trading partnerships or agreements with data providers to ensure a continuous supply of current and historical data for data-driven risk prediction. Obtain data from relevant sources to develop algorithms and models for a useful prediction of risky driving behaviour and the development of driver

	warning systems. Have access to large-scale car data alongside current and historic weather data, to make better driving risk predictions.		
Problem	Difficulty in ensuring effective data management for a scalable and reliable service/product. Insufficient relevant data availability and well-defined data integration process, as well as data quality and frequency, and time synchronization. Difficulty in setting up the relevance of alerts to consider in the model.		
PISTIS role	Provide support for efficient data management to ensure scalability and reliability. Facilitate the data integration process and simplify the process of acquiring and integrating data from different providers. Support the implementation of mechanisms to assess and ensure data quality.		
Actions	**Description**	**The action is performed by...**	**The action impacts...**
Actions before interacting with PISTIS	Pre-processing to fit the data into the data model of the software products, e.g., uniform time series.	Data providers	Data consumers
	Agreement on conditions for data sharing	All	All
	Define relevant data contracts, details, terms of use, pricing policies.	All	All
	Development of semantic enrichment for the shared data sets using a common ontology.	UBIMET	VIF
Step 1 - Input provided to PISTIS	Decision to share data (entire dataset or preliminarily a sample, as deemed necessary) through PISTIS	Data providers	Data consumers
	Automation of event detection, driving style analysis and risk computation	VIF	VIF/ TRA
	Mobile app including a kit that is connecting to the service platform to query the relevant information	VIF	
	Provision of driving style & risk data	VIF	VIF/ TRA
	Implementation of APIs/mechanisms to automatically transmit weather data, also in real time if necessary	UBIMET	VIF
	Real-time information, raw weather data via continuous data stream.	UBIMET	VIF
	Integration of weather data via VIF app/dashboard	UBIMET	VIF
	The provision of weather parameters and associated risks such as road icing conditions and low visibility	UBIMET	VIF
	Access use of smart contracts and crypto currencies	ALL	ALL
	Street graph data	OSM	VIF
	Accident hotspot data	STATISTIC AUSTRIA	VIF
	Live Weather Data	UBIMET	VIF
	Historical Weather Data for Styria	UBIMET	VIF
	Anonymized Connected Car Data	CARUSO	VIF

	Personalized Connected Car Data (using synthetic data)	CARUSO	VIF
Step 3 - Data quality assessment and generation of new analytics (through PISTIS or offline)	Data quality assessment, data decryption and - if necessary - preprocessing	Data providers	-
	Visual exploration of driving risk data to develop automated data quality analysis and improvement	VIF	Dashboard /VIF
Step 4 - Output retrieved from PISTIS and generation of new analytics (through PISTIS or offline)	Calculation of events (relevant to the calculation of driving risk) from raw data (e.g., trip data or anonymised trip data) to feed into risk prediction models	All	Platform/ All
	Queries events with a geo-spatial key computed on the driver's app	VIF	Platform/ Drivers app
	Computation of a cone for risk calculation	VIF	Drivers' app
	Event filtering and risk calculation	VIF	Platform/ Drivers app
	Driver warnings computation and visualisation	VIF	Drivers' app
	The new analytics are made available through PISTIS	All	All
	Data providers monitor that data is correctly transferred and that it is used as agreed in the licences	Data providers	All
Step 5 - After the interaction with PISTIS	Calculation of events (relevant to the calculation of driving risk) from raw data (e.g., trip data or anonymised trip data) to feed into risk prediction models	All	All
	Compute: risk score for drivers considering their driving style, geo-location and time and driving style out of vehicle movements	All	All
	Sell service: Alert drivers using a mobile app based on the risk events in their driving corridor and show risk-events to risk managers on a map (event dashboard) and let them explore.	All	All

2.4 LIFECYCLE AND DATA LANDSCAPE

This section provides a comprehensive overview of the Data Lifecycle and the Data Landscape of the PISTIS demonstrators. By examining the Data Lifecycle, as seen from the point of view of the potential data providers and consumers, and the data landscape, partners within the PISTIS project can gain valuable insights into the complexities and challenges associated with managing data effectively. This knowledge serves as a foundation for developing robust data management strategies and ensuring data quality, security, and compliance throughout the entire Data Lifecycle.

2.4.1 Data Lifecycle in PISTIS

During the initial phase of the project, discussions on defining the main high-level stages of the data lifecycle through the use of questionnaires and workshops were initiated. The purpose was to gain a better understanding of how the data lifecycle could be practically applied and guide PISTIS partners in defining its specific components. To facilitate this process, project partners were encouraged to envision a scenario in which data progressed through different stages. Suppose an organization encounters a data-related problem and their initial available data is not adequate to provide a solution, they will need to find additional data from external sources. They will need to collect, purchase, access, and analyse new data, assessing its suitability and effectiveness in resolving their issue. If the newly acquired data proves to be a viable solution, they may decide to purchase it from the respective provider through the platform. This transaction allows them to supplement their existing datasets and further advance their problem-solving efforts.

In this context, PISTIS will facilitate the query, access, and exchange of data across organisations. The purpose of the PISTIS platform is to enable the trading and exchange of data among organisations. Therefore, the PISTIS platform provides access to data and not analytics solutions. Envisioning a complete data trading scenario helped in defining the different roles involved in the data exchange, that is, Data Seller/Provider and Data Buyer/Consumer, and the different actions they need to perform in order to achieve their goals. Among the actions that describe the entire journey of the data, the relevant ones for PISTIS have then been identified and built the main stages of the PISTIS Data Lifecycle. These stages, and the corresponding high-level actions, are shown in Figure 2. As Data Seller/Provider, access to the PISTIS platform will allow it to transform and ingest data and publish them. Furthermore, the platform will provide Data Monetary Evaluation and supporting tools to Contract Execution for the data exchange with interested parties. Finally, Data Encryption will be supported to enable the Data Seller/Provider to exchange/sell their data in a secure way via Peer-to-Peer transfer. On the other hand, the Data Buyer/Consumer will be allowed to explore data available within the PISTIS platform and acquire the data after the Data Contract has been provided through the platform.

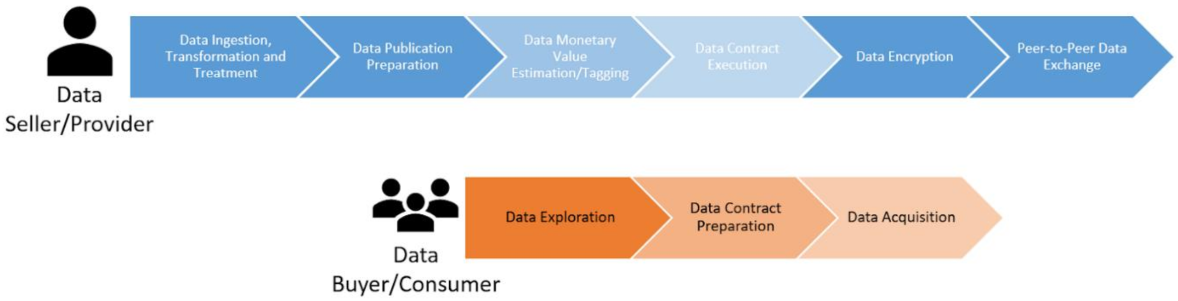

Data Seller/Provider
Data Ingestion, Transformation and Treatment
Data Publication Preparation
Data Monetary Value Estimation/Tagging
Data Contract Execution
Data Encryption
Peer-to-Peer Data Exchange
Data Buyer/Consumer
Data Exploration
Data Contract Preparation
Data Acquisition

Figure 2: Data lifecycle from the involved actors' perspective

The information collected during the questionnaire administration and co-creation workshops was used to further describe the actions performed within each individual stage, as presented in Figure 3.

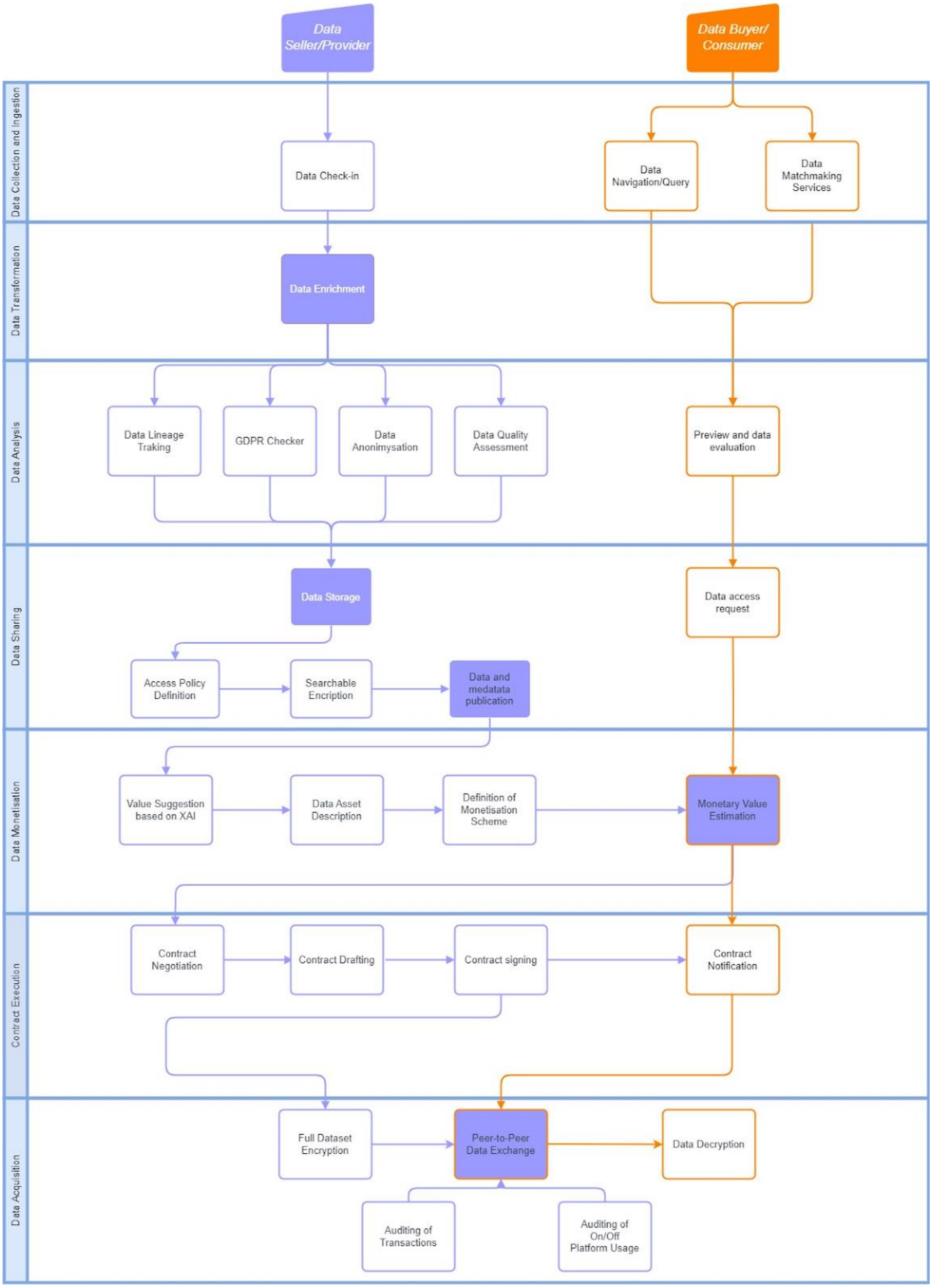


Figure 3: The final version of the Data Lifecycle, comprehensive of all main actions to be performed by the data providers and consumers every time they have a transaction.

For the data seller/provider, the data lifecycle is structured as follows:

- Data Collection and Ingestion: During the data collection and ingestion stage, the process of Data Check-In plays a crucial role. It entails gathering data from the PISTIS system utilizing a variety of options, such APIs, Pub/Sub mechanisms, and other relevant methods.
- Data transformation: Data transformation involves data enrichment. During data enrichment, the collected data undergoes a thorough cleaning process to identify and rectify errors, inconsistencies, and inaccuracies. Additionally, data enrichment involves aligning the ingested data with a standardized and common model to ensure interoperability across different systems and applications.
- Data Analysis and Preparation:
    - Data Analysis encompasses the utilization of an Analytics/Insights Engine to apply a variety of ready-made analytics on the data, facilitating the extraction of valuable information and insights. Through this process, patterns, trends, correlations, and other relevant findings are identified, contributing to a deeper understanding of the data.
    - Data Lineage Tracking involves the implementation of tracking tags on the data, enabling the seamless tracing of its journey and activities throughout the data lifecycle. This allows for comprehensive visibility and accountability, as well as facilitating the identification of dependencies, transformations, and the impact of subsequent actions on the data.
    - The GDPR Checker plays a crucial role in ensuring compliance with data protection regulations. It involves the evaluation of the data to determine if it contains any personally identifiable information or other GDPR-relevant data. If such information is present, the suggestion is made to either remove it entirely or modify it appropriately, based on the preferences and requirements of the stakeholders involved. This step ensures the responsible handling and safeguarding of sensitive data.
    - Data Anonymization involves the application of various techniques to protect the privacy and confidentiality of individuals whose data is being analyzed. By anonymizing the data, any personally identifiable information is removed or transformed in such a way that it cannot be linked back to specific individuals. This ensures that the privacy rights of individuals are respected while allowing for valuable data analysis and insights.
    - Data Quality Assessment involves the evaluation and measurement of data quality in various dimensions. This assessment seeks to identify and address any errors, inconsistencies, or issues with the data that could impact its reliability and usefulness. By extracting relevant indexes and metrics, data quality can be quantitatively and qualitatively assessed, leading to enhanced decision-making and more accurate insights.
- Data Sharing:
    - Data Storage: storing already treated data back to the original data storage facilities while maintaining "pointers" at the PISTIS facility ensures that the treated data is securely stored and easily accessible when needed.

  - Access Policies Definition plays a crucial role in governing the access levels and permissions associated with the stored data in the PISTIS system. It establishes mechanisms and guidelines for managing and enforcing access control policies to safeguard the data's confidentiality, integrity, and availability.
  - Searchable Encryption is a technique applied to specific data samples before publishing them online. This approach allows for efficient and secure searching of encrypted data without revealing the actual content to unauthorized individuals. By employing searchable encryption, sensitive data can be made searchable while maintaining privacy and confidentiality.
  - Data and Metadata Publication is a vital step in sharing data within the PISTIS ecosystem. It involves the dissemination of metadata associated with the treated datasets in federated PISTIS repositories. Metadata provides essential information about the datasets, such as their structure, attributes, and provenance. This enables efficient querying and discovery of relevant data by interested stakeholders, promoting collaboration, and facilitating informed decision-making.
- Data Monetization:
  - Value Suggestion based on the XAI Engine: Leveraging the power of the XAI Engine to propose an estimated value for the data, considering various factors and insights.
  - Data Asset Description (for enabling Sharing): Creating a comprehensive description of the data asset, including its characteristics, relevance, and potential applications. This description facilitates effective sharing and understanding of the data by potential users.
  - Definition of Monetization Scheme: Developing a well-defined strategy for monetizing the data, which can include:
    - Subscription or One-Off Purchase Design Scheme: Offering options for users to access the data through subscription-based models or one-time purchases, depending on their needs and preferences.
    - Data "Equity" Offering: Introducing the concept of data equity, enabling users to acquire a stake in the data asset and potentially benefit from its value appreciation.
    - NFT Generation: Exploring the generation of Non-Fungible Tokens (NFTs) as a means to represent and trade unique instances of the data, providing ownership and value recognition.
  - Monetary Value Estimation: Determination of the financial worth or value of the dataset or data asset
- Contract Execution:
  - Contract Negotiation: Engaging in negotiations between the data provider and consumer to establish the contract terms.
  - Contract Signing: Concluding the contract between the data provider and consumer.
  - Contract Notification: Notifying relevant parties about the finalized contract.
- Data Acquisition:

    - Full Dataset Encryption to ensure the security and confidentiality of the information during transmission.
    - Peer-to-peer Data Exchange: Data will be exported through a variety of methods, providing flexibility in selecting the most suitable method for exporting the data to the intended recipient. The choice of the data export method is influenced by various factors, including the terms specified in the contract between the data provider and consumer.
        - Auditing of Transaction: Users can access an interface that provides them with a detailed log of their transactions, enabling them to review and track the history of their data exchanges. This log serves as a comprehensive record, facilitating data management and facilitating compliance with regulatory requirements or internal governance policies.
        - Auditing of On/Off Platform Usage: Additional auditing feature that allows users to observe and monitor the usage of the data they have traded, to gain insights into how their data is being utilized by other parties, promoting transparency and enabling them to assess the value and impact of their data contributions.

For the data buyer/consumer, the data lifecycle is structured as follow:

- Data Collection and Ingestion: this stage involves the process of gathering data from the PISTIS platform. Data Navigation/Querying is a core functionality of the PISTIS system, enabling users to search and navigate through the federated repositories of data. Users can specify specific parameters, such as keywords, filters, or criteria, to narrow down their search and find relevant datasets. Additionally, PISTIS offers Data Matchmaking Services, which utilize advanced algorithms and techniques to recommend datasets to users. These recommendations are based on a variety of factors, including the user's previous searches, related terms, similar datasets, and user preferences. By leveraging these matchmaking services, users can discover datasets that align with their specific needs and interests.
- Preview and Data Evaluation. This step involves examining the dataset to gain an initial understanding of its contents and structure. The purpose is to assess the quality, relevance, and suitability of the data for the intended analysis. This evaluation helps identify any potential issues, such as missing values, outliers, or inconsistencies, that may require further preprocessing or data cleaning.
- Monetary Value Estimation: Determination of the financial worth or value of the dataset or data asset.
- Contract Drafting: The initiation of a data contract between a data provider and a data consumer for data trading.
- Contract Notification: Notifying relevant parties about the finalised contract.
- Peer-to-peer Data Exchange: Data will be exported through a variety of methods, providing flexibility in selecting the most suitable method for exporting the data to the intended recipient. The choice of the data export method is influenced by various

factors, including the terms specified in the contract between the data provider and consumer.

  - Auditing of Transaction: Users can access an interface that provides them with a detailed log of their transactions, enabling them to review and track the history of their data exchanges. This log serves as a comprehensive record, facilitating data management and facilitating compliance with regulatory requirements or internal governance policies.
  - Auditing of On/Off Platform Usage: Additional auditing feature that allows users to observe and monitor the usage of the data they have traded, to gain insights into how their data is being utilized by other parties, promoting transparency and enabling them to assess the value and impact of their data contributions.
- Data Decryption: the dataset undergoes decryption using provided keys, allowing access to the original, readable data.

2.4.2 Demonstrator Hubs Data Landscape

Data landscape refers to the comprehensive view of data assets within a specific context, including the types of data, data providers, data consumers, data availability, and data formats. In the context of the PISTIS project, it is crucial to provide an overview of the data available for each use case and demonstration hub, to ensure transparency and provide a more detailed overview of the use cases and their current needs and expectations from the PISTIS platform. By providing this comprehensive data landscape for each Demonstration Hub and UC, the PISTIS project aims to offer a clear understanding of the available data assets. This overview enables data traders to identify suitable data sources, understand the data's characteristics and restrictions, and make informed decisions regarding data acquisition and utilization.

The information provided in this section have been collected through a series of different activities, which are described in some detail in Section 2.2 and are summarised as follows. First, a preliminary questionnaire on the Data Lifecycle has been distributed to partners. This questionnaire served also to gather initial insights into the available data assets and their characteristics. Building upon the questionnaire responses, consolidation workshops were organised, bringing together key partners. These workshops provided a platform for in-depth discussions, knowledge sharing, and validation of the collected data landscape information. Through collaborative efforts, the participants refined and enriched their understanding of the data assets, ensuring accuracy and relevance. Finally, the findings from the questionnaires and consolidation workshops. were further examined, refined, and confirmed during discussions and presentations at plenary meetings. These meetings involved project partners who brought their expertise and perspectives to the table, offering valuable insights and input on the data landscape for each use case and demonstration hub.

The data landscape for all UCs in each Demonstration Hub are presented in the following Tables Table 15-to Table 25, which show the following key information:

1. Type of Data: this column specifies the nature of the data available for trading.

2. Data Provider: here, the partners responsible for supplying the data are listed.
3. Data Consumer: this column outlines the intended recipients or users of the data within the consortium. Data consumers would require the data for analysis, research, or decision-making purposes within the Use Case. In the case of large organisations with multiple departments, it is possible that the same organisation plays both roles of Data Provider and Consumer. The PISTIS platform could in principle facilitate data exchange also within the same organisation.
4. Data Availability: this aspect describes the accessibility and availability of the data for the purposes of the PISTIS project. It includes information on whether the data is already accessible or requires specific permissions or agreements to access within the project Demonstration Hubs.

Table 15 Data Landscape for UC.1.1 - Baggage handling management

Data Type	Data Provider	Data Consumer	Data Availability
Historic Traffic Data	OAG	AIA, GOLDAIR	Readily Available
Arriving/Transfer/Departing timestamps	AIA	GOLDAIR	Readily Available
Minimum Connecting Times for Transfer Bags	GOLDAIR	AIA	Readily Available
Flight Schedule and day of operation updates (Delays, Scheduled/Estimated/Actual Flight Timings)	AIA	GOLDAIR	Readily Available
Force majeure and operational irregularities	AIA	GOLDAIR	Readily Available
Weather data (adverse weather conditions)	UBIMET	AIA, GOLDAIR	Readily Available
Public Transport Scheduling data (staff arrival issues)	OASA	AIA	Readily Available

Table 16 Data Landscape for UC.1.2 - Transfer passengers management

Data Type	Data Provider	Data Consumer	Data Availability
Flight status	OAG	AIA	Readily Available
Handler data stamps	GOLDAIR	AIA, GOLDAIR	Readily Available

Arrival/Transfer/Departing Bag Number	AIA	AIA	Readily Available
Transfer passenger numbers and destinations	GOLDAIR	AIA	Readily Available
PTM Messages	GOLDAIR	AIA	Readily Available
Transfer Baggage information	GOLDAIR, AIA	AIA	Readily Available
Type of connections		AIA	Not Readily Available
Baggage Irregularities e.g., fragile, unclaimed, owner unknown, rush, etc.		AIA	Not Readily Available
PRM Passengers	GOLDAIR, AIA	AIA	Readily Available
Immigration – Customs Clearance		AIA	Not Readily Available
Minimum Connecting Times	OAG	AIA	Not Readily Available
Flight Schedule and day off operation updates	GOLDAIR, AIA	AIA	Readily Available
Force majeure and operational irregularities	GOLDAIR, AIA	AIA	Readily Available
Weather data	UBIMET	AIA	Readily Available

Table 17 Data Landscape for UC.1.3 - Aircraft turnaround process

Data Type	Data Provider	Data Consumer	Data Availability
Flight schedules	OAG	AIA	Readily Available
Type of Aircraft	OAG	AIA	Readily Available
Handler data stamps	GOLDAIR, OAG	AIA, GOLDAIR, OASA	Readily Available
Weather	UBIMET		Readily Available
Aircraft registration, Flight Type, Aircraft Parking Stand, Boarding Gate, Landing and Taxi Intime, In-block and Off-block Time, Taxi-out and Take off Time, Turn round	AIA	AIA[1]	Not Readily Available

Times, Deicing Information, Aircraft movement Data, TOBT Updates, MTB Messages, Operation Planning Information			
Passengers’ numbers, Force Majeure and Operational irregularities	AIA	AIA[1]	Readily Available

Table 18 Data Landscape for UC.1.4 - Public transportation planning

Data Type	Data Provider	Data Consumer	Data Availability
Throughput time	OAG	OASA	Readily Available
Metro/Bus existing schedules, capacity of fleet, available fleet/drivers, Origin-Destination matrices (i.e., demand), limitations of the network/fleet	OASA	OASA	Readily Available
Historic data (e.g., flight schedules, gate/terminal usage)	AIA	OASA	Readily Available
Timestamps	GOLDAIR	DAEM, OASA, AIA	Readily Available
Aircraft load factors	GOLDAIR	DAEM, OASA, AIA	Readily Available
Public transport schedules	OASA	DAEM, OASA, AIA	Readily Available
Airport modal share	AIA	DAEM, OASA, AIA	Readily Available
Weather	UBIMET	DAEM, OASA, AIA	Readily Available
Inbound/outbound passengers per hour	AIA	DAEM, OASA	Not Readily Available
Transport modal split	OASA, AIA	OASA, AIA	Readily Available
No. of Busses dedicated to the Airport routes	OASA	AIA, DAEM	Readily Available

[1] AIA would require this data for internal use and could utilize the PISTIS platform for internal data exchange.

Flight Schedule and day off operation updates	AIA	OASA	Readily Available
Passenger numbers - provisional loads	AIA	OASA	Readily Available
Force majeure and operational irregularities	OASA, AIA	OASA, AIA	Not Readily Available

Table 19 Data Landscape for UC.1.5 - Insights for city commercial businesses

Data Type	Data Provider	Data Consumer	Data Availability
Timestamps	OAG	AIA	Readily Available
Mobility/touristic/business registry	AIA, OASA	DAEM	Readily Available
Number of expected visitors	AIA	DAEM	Readily Available
Destination of visitors	AIA	DAEM	Readily Available
GIS info on location, district, address, floor	DAEM	DAEM	Readily Available
Type of business and admin changes	DAEM	DAEM	Readily Available

Table 20 Data Landscape for UC.2.1 - Increase the hosting capacity of the grid

Data Type	Data Provider	Data Consumer	Data Availability
Grid Topology	CUERVA	CARTIF	Readily Available
DERs Location	CUERVA	CARTIF, OMIE	Readily Available
DERs Generation	CUERVA	BAMBOO	Readily Available
Grid Events (historical)	CUERVA	CARTIF	Readily Available
User Consumption (historical)	CUERVA	CARTIF	Readily Available
Flexibility Aggregated Data	BAMBOO	OMIE, CARTIF	Not Readily Available
Bids	BAMBOO	OMIE, CARTIF	Not Readily Available

Hosting Capacity Analytical Results	CARTIF	CUERVA	Not Readily Available
Topology of the network, measurement of energy use by customers, future actions on the network and the economic value that this entails.	CUERVA	OMIE, BAMBOO, CARTIF	Readily Available
Unit power, unit location, unit schedule, unit bids (hour contract, quantity and price)	BAMBOO, CUERVA	OMIE	Readily Available
Requirement (quantity, limit price, hour contract)	CUERVA	OMIE	Readily Available
Data generated by the grid (voltage, current, power, switches position, etc.), Data generated by the users (energy consumption)	CUERVA	OMIE, BAMBOO, CARTIF	Readily Available
Weather data, energy forecasting data for PV-, wind- and hydropower applications, as well as feed in management solutions	UBIMET		Readily Available
Electrical consumption and generation.	CUERVA	BAMBOO, CARTIF	Readily Available

Table 21 Data Landscape for UC.2.2 - Investment deferral

Data Type	Data Provider	Data Consumer	Data Availability
Grid Topology	CUERVA	CARTIF	Readily Available
Grid Investment	CUERVA	CARTIF	Readily Available
Grid Events (historical)	CUERVA	CARTIF	Readily Available
User Consumption (historical)	CUERVA	CARTIF	Readily Available
DERs Generation	CUERVA	BAMBOO	Readily Available
Flexibility Aggregated Data	BAMBOO	OMIE, CARTIF	Not Readily Available
Result of problem solved thanks to flexibility	CARTIF	CUERVA	Not Readily Available

Topology of the network, measurement of energy use by customers, future actions on the network and the economic value that this entails.	CUERVA	OMIE, BAMBOO, CARTIF	Readily Available
Unit power, unit location, unit schedule, unit bids (hour contract, quantity and price)	BAMBOO, CUERVA	OMIE	Readily Available
Requirement (quantity, limit price, hour contract)	CUERVA	OMIE	Readily Available
Data generated by the grid (voltage, current, power, switches position, etc.), Data generated by the users (energy consumption)	CUERVA	OMIE, BAMBOO, CARTIF	Readily Available
Weather data, energy forecasting data for PV-, wind- and hydropower applications, as well as feed in management solutions	UBIMET		Readily Available
Electrical consumption, electrical generation.	CUERVA	BAMBOO, CARTIF	Readily Available

Table 22 Data Landscape for UC.2.3 - P2P Trading Users or Energy Communities

Data Type	Data Provider	Data Consumer	Data Availability
Any kind of data collected by or from stakeholders	CUERVA	EURECAT	Readily Available
Grid Topology	CUERVA	CARTIF	Readily Available
User Consumption (historical)	CUERVA	CARTIF	Readily Available
DERs Generation	CUERVA	BAMBOO	Readily Available
Flexibility Aggregated Data	BAMBOO	OMIE, CARTIF	Readily Available
Topology of the network, measurement of energy use by customers, future actions on the network and the economic value that this entails.	CUERVA	OMIE, BAMBOO, CARTIF	Readily Available
Unit power, unit location, unit schedule, unit bids (hour contract, quantity and price)	BAMBOO, CUERVA	OMIE	Readily Available

Data generated by the grid (voltage, current, power, switches position, etc.), Data generated by the users (energy consumption)	CUERVA	OMIE, BAMBOO, CARTIF	Readily Available
Weather data, energy forecasting data for PV-, wind- and hydropower applications, as well as feed in management solutions	UBIMET		Readily Available
Electrical consumption, electrical generation.	CUERVA	BAMBOO, CARTIF	Readily Available

Table 23 Data Landscape for UC.2.4 - Monetisation of Data owned by the different Actors to Third Parties

Data Type	Data Provider	Data Consumer	Data Availability
Digitalisation Costs	CUERVA		Readily Available
Grid O&M operation	CUERVA		Readily Available
Data Infrastructure Costs	CUERVA		Readily Available
Topology of the network, measurement of energy use by customers, future actions on the network and the economic value that this entails.	CUERVA	BAMBOO, CARTIF	Readily Available
Unit power, unit location, unit schedule, unit bids (hour contract, quantity and price)	BAMBOO, CUERVA		Readily Available
Requirement (quantity, limit price, hour contract)	CUERVA		Readily Available
Data generated by the grid (voltage, current, power, switches position, etc.), Data generated by the users (energy consumption)	CUERVA	BAMBOO, CARTIF	Readily Available
Weather data, energy forecasting data for PV-, wind- and hydropower applications, as well as feed in management solutions	UBIMET		Readily Available
Electrical consumption, electrical generation.	CUERVA	BAMBOO, CARTIF	Readily Available

Table 24 Data Landscape for UC.3.1 - Traffic quality assessment

Data Type	Data Provider	Data Consumer	Data Availability
Vehicle sensor data	CARUSO	TRAFFICON	Readily Available
Vehicle data	CARUSO	TRAFFICON	Readily Available
Fuel consumption, risk, driving styles	CARUSO	TRAFFICON	Readily Available
Weather Data (live)	UBIMET	TRAFFICON	Readily Available
Connected Car Data	CARUSO	TRAFFICON	Readily Available
Driving Style and Risk Assessment	VIF	TRAFFICON	Readily Available

Table 25 Data Landscape for UC.3.2 - Driving style and risk assessment.

Data Type	Data Provider	Data Consumer	Data Availability
Past trip data	VIF	CARUSO, VIF	Readily Available
Vehicle trip data	CARUSO	CARUSO, VIF	Readily Available
Driving style data	VIF	VIF, TRAFFICON	Readily Available
Weather Data (historical)	UBIMET	VIF	Readily Available
Weather Data (live)	UBIMET	VIF	Readily Available
Anonymised Connected Car Data	CARUSO	VIF	Readily Available
Personalized Connected Car Data	CARUSO	VIF	Readily Available

3 BUSINESS REQUIREMENTS

Business requirements are specific statements or descriptions that outline the needs, objectives, and expectations of a business or organization. These requirements define what the business aims to achieve, how it should operate, and what outcomes it desires. Business requirements typically focus on the desired results rather than the technical implementation details.

In principle, business requirements are derived from various sources, including customer demands, market analysis, industry standards, regulatory requirements, and internal stakeholders. They serve as a foundation for guiding the development and implementation of solutions, projects, or systems to meet the identified needs of the business.

In essence, business requirements articulate the "what" of a business initiative, describing the desired outcomes, functionalities, capabilities, constraints, and performance criteria that must be fulfilled to address a specific business problem or opportunity. These requirements act as a bridge between business goals and the technical or operational solutions that are designed to fulfil those goals.

In the PISTIS context, the methodology described in Section 2.2 was followed to support the business analysis process and extract the business requirements. The focus was given in extracting and defining business requirements with respect to the design and implementation of the PISTIS platform to serve the objectives of the use cases. The business requirements for each UC per se is beyond the scope of this deliverable.

Each PISTIS business requirement is defined by a unique identifier (id), a title, a description, the actors that are affected and the prerequisites for satisfying each requirement. This information will be used at a later stage to extract the technical functional and non – functional requirements for the PISTIS platform that will be documented in D1.2. The interdependencies among the business requirements is a valuable tool that will help to prioritize the implementation of respective components that address the related requirements. The extracted business requirements are reported in tables below.

Id	BR01
Title	User registration and management
Description	A user is registered in PISTIS as part of an organization and gets access to the PISTIS functionalities and assets that the organization dictates. With respect to PISTIS use cases such organizations are AIA, Goldair, UBIMET, OASA, CUERVA, CARTIFF, ICCS, OMIE, BAMBOO, etc.
Impacts	Data consumer, Data provider
Depends on	-

Id	BR02
Title	New data addition
Description	Data provider decides to make new data available through PISTIS. Semantic enrichment of the data should be supported to make it searchable using keywords, enhancing data accessibility and usability. This enrichment should be performed with the respect to the adopted interoperability standards. Use case specific examples: UC1.4: OASA decides to provide public transport timetables through PISTIS UC2.1: UBIMET decides to provide weather data through PISTIS
Impacts	Data consumer
Depends on	BR11

Id	BR03
Title	Data Discovery
Description	PISTIS should provide a registry with available data sources and support the discovery of the required data. This ensures that the necessary data sources are identified and accessed. PISTIS should include the necessary metadata for the data sources to be easily and effectively discoverable. Use case specific examples: UC2.2: CARTIFF discovers the availability of grid topology and investment data. UC3.1: TRAFFICON discovers the availability of driving data provided by VIF and weather data provided from UBIMET
Impacts	Data consumer
Depends on	-

Id	BR04
Title	Data Query
Description	Data consumers should be able to gain insights on the available data in addition to the related metadata, if it’s allowed by their data provider. This

	query mechanism should support searching through all the available distributed PISTIS data sources. Use case specific examples: UC3.2: Drivers' app queries events with a geo spatial key computed on driver's app; Visual exploration of driving risk data to develop automated data quality analysis and improvement.
Impacts	Data consumer
Depends on	-

Id	BR05
Title	Data anonymization
Description	The platform should facilitate the anonymization of data and removal of business-sensitive information to protect data privacy. Use case specific examples: UC1.1: AIA removes sensitive information from passengers' bag arrival data. UC2.3: CUERVA anonymises users' energy consumption data, DER production and location
Impacts	Data provider
Depends on	-

Id	BR06
Title	Data Quality Assessment
Description	PISTIS should enable data providers to assess the quality of their data. This ensures that the data meets the required standards (including GDPR restrictions) and can be used effectively by data consumers. This requirement strongly depends on the definition of the metrics that assess data quality.
Impacts	Data consumer
Depends on	BR07, BR11

Id	BR07
Title	Licensing and Policies

Description	PISTIS should support the definition of licensing and policies for using the traded data. Data should be made available through PISTIS under specific rules and regulations. Moreover, access policies of data by the data consumers should be defined in the context PISTIS data exchange and trading services. With respect to these licensing and policy schemes, PISTIS should be able to monitor and regulate the usage of exchanged data within the platform.
Impacts	Data consumer, Data provider
Depends on	BR11

Id	BR08
Title	Data exchange preparation
Description	Before the actual data exchange between consumer and provider takes place, specific processes need to take place such as secure peer-to-peer channel establishment, contract establishment and verification as well as monetary exchange.
Impacts	Data consumer, Data provider
Depends on	BR11

Id	BR09
Title	Data Exchange of datasets and data streams
Description	The platform should facilitate the exchange of data between stakeholders, adhering to agreed-upon Service Level Agreements (SLAs) that support both the exchange of static datasets as well as the exchange of data streams that provide dynamic data. Use case specific examples: UC3.2: UBIMET provides the required APIs / mechanisms for real time weather data transmission to VIF
Impacts	Data consumer
Depends on	BR11

Id	BR10
Title	Automated Data Trading

Description	PISTIS should support the setup of mechanisms to automatically receive specific data sets at regular intervals. For example, daily updates of datasets and operation plans can be received to ensure the availability of up-to-date information for planning and decision-making. This synchronization follows the synchronous mode where data provider provides required data to data consumer at predefined intervals (batch operations). Indicative reference use case(s): UC1.4: OASA provides the incoming / outcoming passengers' data and bus geolocation data from previous day. UC1.4: AIA provides historic data of flight schedules, gate/terminal usage, inbound/outbound passengers per hour to OASA and DAEM
Impacts	Data consumer
Depends on	BR9

Id	BR11
Title	Interoperability and data format
Description	PISTIS should provide common information models, data formats, and protocols for data exchange. This ensures interoperability and consistency in data exchange between the involved parties (data consumers & data providers. Indicative reference use case(s): UC2.1: Agree on common information model (CIM), data format and protocols among all the involved parties (CUERVA, CARTIF, BAMBOO)
Impacts	Data consumer, Data provider
Depends on	-

Table 34 BR12 - Data usage and analytics

Id	BR12
Title	Data usage and analytics
Description	PISTIS should provide data usage and market insights to the data provider. Moreover, the involved parties will need to get information regarding the progress of their data exchange along with any related notifications. Indicative reference use case(s):

	UC2.1: OMIE creates short-term flexibility markets (auctions, under the CUERVA call) from assets information (location, among other) and their bids. UC2.1: CARTIF combines and analyses grid data, prosumers data, flexibility validation data to generate new insights and predictions of network hosting capacity.
Impacts	Data provider
Depends on	-

Id	BR13
Title	Peer-to-Peer Trade Auditing
Description	The platform should support the recording and reporting of peer-to-peer trading results, enabling users to track and validate their transactions. A report system for validating transactions will provide users with a clear overview of their trading activities. An auditing mechanism should be available to monitor the trading process (transactions and records). This auditing mechanism will enable PISTIS to verify the transactional accuracy, security and contractual compliance that the involved parties have agreed on within the PISTIS context in favour of PISTIS governance and transparency. UC2.1: CUERVA / UBIMET check retrieved data for data format, data characteristics and their compliance with contractual requirements.
Impacts	Data consumer, Data provider
Depends on	BR14

Id	BR14
Title	Data Contracts and Terms
Description	The platform should enable agreement on conditions for data sharing, ensuring that stakeholders have agreed upon how and when data will be shared, including data security and privacy issues, legal requirements, terms of use, and pricing policies. This ensures clarity and agreement among all parties involved in data sharing. PISTIS should provide this functionality through the creation and signing of smart contracts.
Impacts	Data consumer, Data provider
Depends on	BR11

Id	BR15
Title	Storage of Contracts and Contract execution reports
Description	PISTIS should provide a secure storage mechanism for storing smart contracts and validation results, ensuring transparency and traceability of peer-to-peer transactions.
Impacts	Data consumer, Data provider
Depends on	-

Id	BR16
Title	Secure Authentication and Authorization
Description	PISTIS should implement and support robust authentication and authorization mechanisms to verify the identity of users accessing the platform. This can include strong password policies, multi-factor authentication, and secure session management. Additionally, implement proper authorization controls to ensure users have appropriate access privileges based on their roles and the resource (data) attributes.
Impacts	All
Depends on	-

Id	BR17
Title	Platform Security
Description	PISTIS should implement network and system security measures, such as intrusion detection and prevention systems, automated system and network configuration, data encryption in transit, network segmentation and isolation to limit the risk of breaches and access to data and transaction ledgers.
Impacts	All
Depends on	-

4 LEGAL REQUIREMENTS AND ETHICAL PRINCIPLES

4.1 ETHICAL AND LEGAL CONSIDERATIONS FOR DATA SHARING

This section analyses the legal and ethical aspects of PISTIS and carries out initial plotting of the legal and ethical requirements that are highly relevant to the project within the scope of Task 1.3 - GDPR, Cross-Border Legal Aspects and Contracts Definition for Data Sharing and Trading. The legal and ethical analysis is largely based on the outcome of the workshops with the demonstrator partners while also considering the overall data sharing activities as described in the Grant Agreement envisaged in the PISTIS proposal. Particular attention is given to the key considerations for data sharing, protection of personal data, legal nature, and function of PISTIS platform, use of smart contracts and data sovereignty. Therefore, the legal and ethical analysis aims to identify potential challenges with respect to data sharing and to offer insights on business and end-user limitations that may affect the demonstration and validation of PISTIS. It is aimed that the key legal requirements and the ethical principles described in this section complement PISTIS Operation Principles to be defined in Work Package WP1 - PISTIS Trusted and Interoperable Data Trading and Management Framework. Overall, as the project progresses and the PISTIS architecture becomes more solid, the legal requirements and ethical principles will be refined. Taking into account, among other, the Model Contractual Terms to be prepared by the European Commission's Expert Group on B2B Data Sharing and Cloud Computing Contract s2 in accordance with the requirements set under the upcoming Data Act, overall, it is. intended the related instruments to be produced by PISTIS that they facilitate trustworthy data sharing. This follow-up progress will be documented in the final deliverable D1.3 - PISTIS Technical Requirements and MVP - v2 of WP1.

4.1.1 Introduction Digital Decade 2030 & Data Sharing, based on Trust.

4.1.1.1 Dynamics in the Digital Age

Technology changes the world at an ever-increasing pace. Whether we like it or not, and whether we are active in mobility, logistics, energy or other societal-relevant sectors, domains and dimensions the change is expedited by both non-digital global occurrences such as pandemics and geopolitical developments as well as by increased and ever-converging technical capabilities such as connected devices, platforms, available data, computing, artificial intelligence and the like. These enable connecting, inter-connecting and hyper-connecting billions of individuals, organizations, communities, societies, and data, with tens of billions of objects and entities. Furthermore, digital has become a must-have, for people, society, and our ecosystems, within the European Union as well as globally.

4.1.1.2 Data (b)locking or data sharing?

However, smart, and otherwise advanced one may want to market this Digital Age is, it is for sure not immune to evil, ignorance, build-fast-fix later business models and other breaches of norms and values. These threaten systems, services, lives of people, key networks and even entire nations, democracies, and societies. And malicious and other less-ethical actors do not work alone. They have joined forces, and they are increasingly winning because of that.

One may feel the urge to blame it all on the others. Or feel the urge to believe that the battle – and the war – is lost. However, there is no way one can point to the other and blame them for everything. Whatever and whoever is part of the Digital Age is part of the problem. However, whatever and whoever is part of this, is part of the solution as well. One of the solutions is to join forces, to partner up, to start and continue collaborating, orchestrating our knowledge, values, and other capabilities – and to organise ourselves for this dynamic Digital Age.

Stopping pointing fingers, starting to collaborate and contribute, and meanwhile demonstrating accountability and co-accountability are prerequisites. One key essential to do so, is to share. Share data, share information, share experience, share good practices, share lessons-learned, and share knowledge. It should not be that hard to do, right?

4.1.1.3 Data sharing is not an easy feature

Sharing any data, information, or knowledge with another has proven not to be an easy feature. These and other queries, potential concerns and considerations come up, even before sharing any information to anyone:

a. Who am I? Who am I representing here? What is my mandate?
b. Why do I feel the urgency or other need to consider sharing that information? Or am I obliged to share, based on new or existing regulations or industry standard practices?
c. What do I share, and what not? And what is the provenance, quality and relevance of the information that could be shared?
d. To what extent am I allowed to share it? And to whom?
e. Do I know the other party, or not?
f. What will the recipient do with it? What is in it for me?
g. And what is the risk? What if something is wrong or goes wrong with the information, with the sharing, with the use?
h. Do the potential, envisioned efforts and benefits outweigh the potential risks and consequences?
i. Why not just leave it, not engage, and not share information?

4.1.1.4 Data sharing, based on trust

These and other considerations all boil down to five letters: trust. One needs trust, before, during and after data sharing. Trust and related trustworthiness are always the main enablers, also in the PISTIS focus areas and related domains, dimensions, communities and other stakeholders, and the relevant ecosystems. Does one have the appropriate level of trust in the digital means and assets, trust in its own competences, trust in the organisations and community involved, trust in the technical systems and trust in the ecosystem at large? The right level of trust both brings the comfort, confidence and courage to engage, and share.

One needs many different stakeholders in any use case or related situation or scenario to come to sufficient levels of transparency, trust, engagement and accountability in order to come to sufficient amounts of relevant and interesting data sharing.

4.1.1.5 Data is the common denominator

Data is the common denominator. However, where the increasing use of data processing and related computing technologies such as cloud, edge, and far edge computing, are promoting and enabling data exchange, analysis, processing and storage, the current operations and implementations do not generally meet the minimal threshold for the envisioned European digital sovereignty [1] and are not yet regulated.

The European Union, both on Union level, member state, regional and local levels is enabling and facilitating various trust anchors on the numerous and obvious opportunities of data sharing and creating digital sovereign cyber-physical and other digital data ecosystems, as well as on unlocking data and addressing various ethical and legal challenges related thereto.

Digital data and related data processing for once make up a crucial cornerstone of the 2030 Digital Decade strategy [2]. This is called, the Digital Decade 2023, for once to make Europe fit for the Digital Age. This has also been reconfirmed by the European Declaration on Digital Rights and Principles and the related Digital Decade Policy Programme 2023 [3]. The latter endorses and implements common value-based objectives [4] and concrete targets [4] for Europe’s digital transformation with a robust governance monitoring and cooperation mechanism, as well as policy tools to accelerate and deepen multi-country projects (‘MCPs’) throughout the EU (and where applicable: its periphery).

Europe’s Digital Decade vision, strategy, declaration, objectives, targets, and programmes to the extent most notable for PISTIS are visualised in Figure 4.

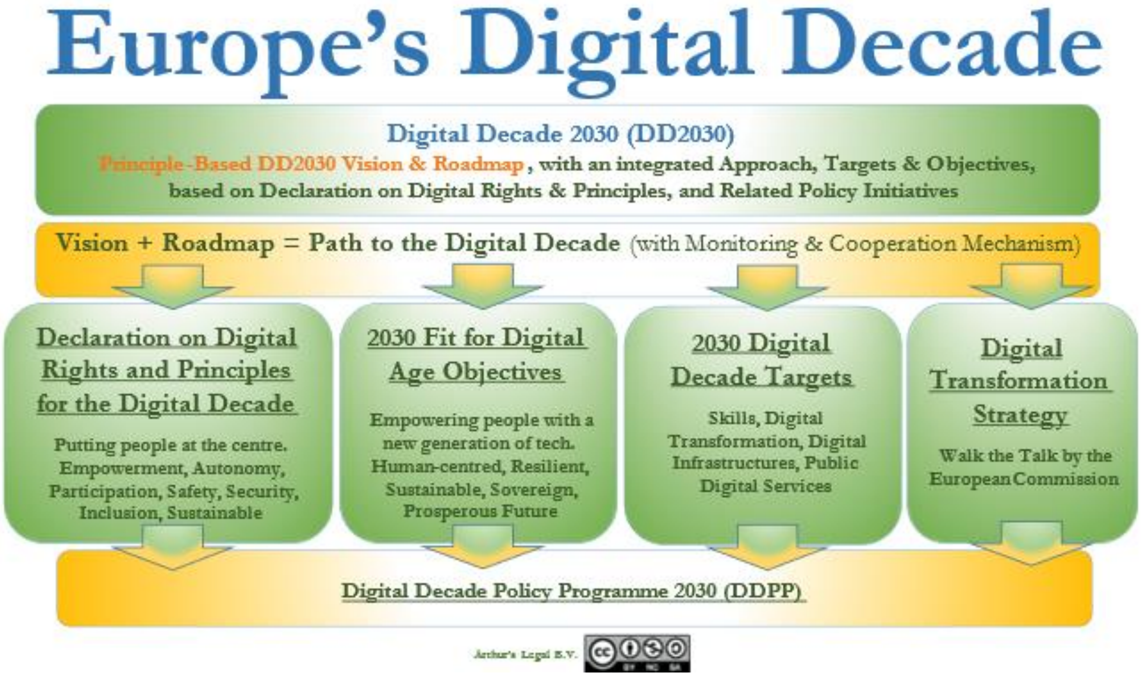

Europe’s Digital Decade
Digital Decade 2030 (DD2030)
Principle-Based DD2030 Vision & Roadmap, with an integrated Approach, Targets & Objectives, based on Declaration on Digital Rights & Principles, and Related Policy Initiatives
Vision + Roadmap = Path to the Digital Decade (with Monitoring & Cooperation Mechanism)
Declaration on Digital Rights and Principles for the Digital Decade
Putting people at the centre. Empowerment, Autonomy, Participation, Safety, Security, Inclusion, Sustainable
2030 Fit for Digital Age Objectives
Empowering people with a new generation of tech. Human-centred, Resilient, Sustainable, Sovereign, Prosperous Future
2030 Digital Decade Targets
Skills, Digital Transformation, Digital Infrastructures, Public Digital Services
Digital Transformation Strategy
Walk the Talk by the European Commission
Digital Decade Policy Programme 2030 (DDPP)
Arthur's Legal B.V.

Figure 4: Overview of the Digital Decade 2023 Vision and Roadmap

4.1.1.6 Data Strategy & Cybersecurity Strategy

Regarding PISTIS, the main underlying strategy that is part of the above, is the European strategy for data (Data Strategy) [5] and the cybersecurity strategy [6].

The latter is about building resilience to cyber threats and ensure citizens and businesses benefit from trustworthy digital technologies. It consists of regulatory, investment and policy initiatives, and address three areas of EU action: (i) resilience, technological sovereignty, and leadership; (ii) operational capacity to prevent, deter and respond; (iii) cooperation to advance a global and open cyberspace. In paragraph below about System-Centric Perspectives, some for PISTIS most notable regulations related to the Cybersecurity Strategy is elaborated on.

Regarding the Data Strategy, its motto is facilitating access to high-quality data for businesses and public authorities to boost growth and create value. Its main provisions are basically:

1. Adopt legislative measures on data governance, access, and re-use.
2. Opening high value publicly held datasets and allowing their re-use.
3. Enable access to secure, fair, and competitive data processing, and.
4. Enable the investment of billions of Euros in European High Impact Projects to develop data processing infrastructures, data sharing tools, architectures, and governance mechanisms for thriving data sharing and to federate energy-efficient and trustworthy cloud infrastructures and related services.

Same as the Cybersecurity Strategy, the Data Strategy demonstrates that policy making is not merely about regulations. It is about strategy, tactics, and operations, such as organising, investing, enabling, incentivising, facilitating and otherwise catering for new ecosystems, while mitigating risks and increasing and optimising positive, people-centric data- and other digital sovereignty, while addressing societal challenges and facilitating innovation of new digital value models, business, and financial models, in a trustworthy, secure, safe, viable, feasible, accountable, and sustainable manner.

Said otherwise, it helps build, achieve, and sustain the ability at European level to act and decide independently in a global environment on digital products, systems, and services. It benefits the development of strong competencies, research and innovation, the promotion of the European industrial, economical, and societal position on data and data processing, and the protection of European values and democracy. Some components are harder to tackle from the sovereignty point of view (such as microchips and raw material), while others are lower hanging fruits (e.g., software (open source), data architectures, data components and data governance).

Obviously, for the project about Promoting and Incentivising Sharing and Trading of Interoperable Data Assets, also known as PISTIS, these European ethical and legal frameworks are quite essential, while it helps to load these frameworks with its various pilots and related use cases. This Deliverable aims to bring guidance to make it work.

4.1.2 Data Sharing Key Considerations

When an organisation needs to or is aiming to or otherwise considering sharing data, either inbound or outbound, these are some key notions and recommendations to consider and consider:

Read the European Declaration on Digital Rights and Principles, as well as the Data Strategy. Both the Declaration as well as the Data Strategy, mentioned in the previous paragraph, are easy, non-legal reads. These provide guidance about where the digital decade strategies are aiming for, how to benefit from those, how to be part of it, and what main public values, objectives and some other ethical, behavioural, responsibility and accountability components and dimensions are pursued.

There is a lot to share. Data is not a four-letter word. It can be OT-, IT- or sensor-generated data coming out in a 'raw' manner. It can be structured data – being information or other attributes – either from registers or other root sources, combined with other data sets, or even without revealing the data source or data while providing appropriate level of trust about its provenance, quality, and integrity (such as, for instance, zero knowledge proof). It be used data and information, enriched, or otherwise combined and feedback-looped – being knowledge and experience –. There also numerous data classifications, such as open data, confidential data, classified, sensitive, personal, non-personal, derived, meta and other data.

Data should not be treated as a four-letter word. It can have many classifications and dimensions. Data classification, both on high level as well as in a fine-grained way is the initial step to take. Furthermore, it is important to note that data has quantum capabilities; it can and it generally will have several classifications at the same time as it can have different contexts simultaneously.

Data is both an asset and a means. This, for once, is also reconfirmed and reinstated in the Data Act. Data is an asset of diverse values and with certain integrity, and therefore in itself already valuable enough for any actors (friendly and malicious) to be interested in. And a means to do good, and a means to be used in a malicious way, either intentionally or unintentionally, and by anybody that has or is able to obtain access to the data.

Data needs to be addressed as a dimension, intertwined with other domains and dimensions, as it is relevant in all technical, organisational and operational layers, domains and dimensions, throughout the various ecosystems and life cycles; end-to-end. Data is that Each data record or data set will have its own data life cycle. Essentials to be considered are for instance: what data is collected, created or otherwise concerned; what are its classifications; can it be segmented, minimised and isolated; what if it has multiple classifications; what if the classification changes; who controls the data; for what purposes is one entitled to process the data; what meta data and derived data is generated during the data life cycle; what does true data deletion mean; what is other life cycles evolve or otherwise change – such as the digital system life cycle, stakeholders life cycle, contextual life cycle or legal life cycle?

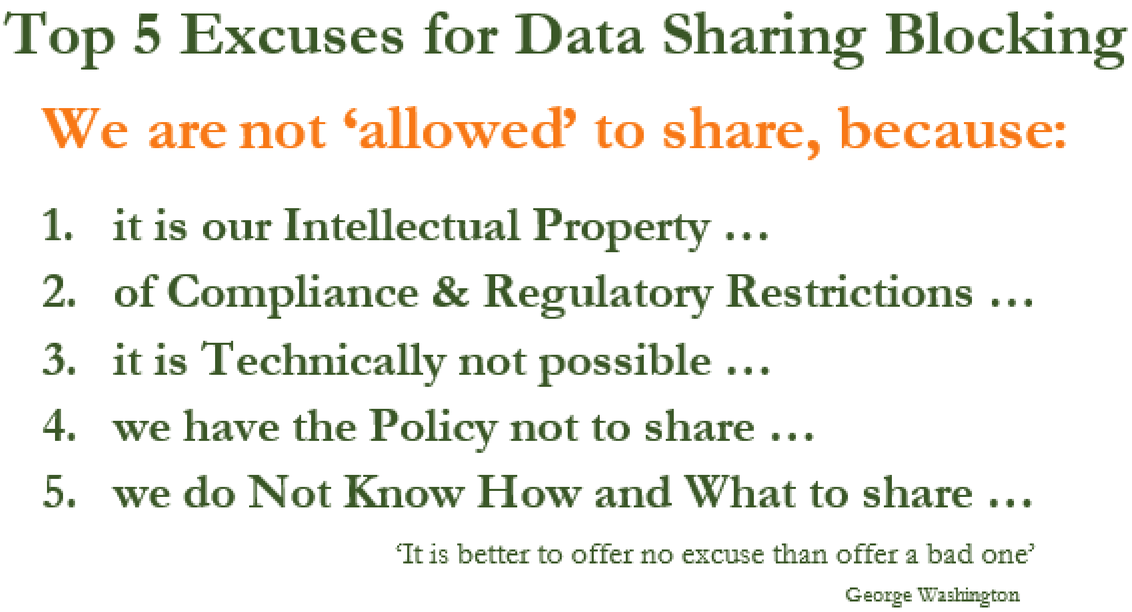
Start with strategic, tactical, and operational questions, and discuss those internally, and thereafter with your potential data sharing partner(s). These include the questions already mentioned in the previous paragraph, such as:

1. Who am I? Who am I representing here? What is the mandate of my organisation? What is my mandate? What is my persona as a professional within the organisation?
2. Why does my organisation feel the urgency or other need to consider sharing that information? Or is my organisation obliged to share, based on new or existing regulations or industry standard practices?
3. What should my organisation share, and what not? And what is the provenance, quality and relevance of the information that could be shared?
4. To what extent am I allowed to share it on behalf of my organisation? And to whom?
5. Does our organisation know the other party, or not?
6. What will the recipient do with it? What is in it for us?

7. And what is the risk? What if something is wrong or goes wrong with the information, with the sharing, with the use?
8. Do the potential, envisioned efforts and benefits outweigh the potential risks and consequences?
9. Why not just leave it, not engage, and not share information?

Data Sharing: no one can do it alone. To share, one needs to know who the data holders, data providers, data recipients and other stakeholders are. These include a proper stakeholder plotting and mapping of the stakeholders within one's own organisation, as well as obviously the external stakeholders.

Avoid Excuses. One should expect persons or organisations that will state that they are not allowed to share data. While the Digital Decade 2030 strategies, including policy instruments such as the Data Governance Act, Data Act and others, support and incentivize – and in various cases also make mandatory – and otherwise create trust frameworks to share data, trying to avoid excuses and help these persons or organisations with bringing comfort about data sharing, and building confidence by understanding what the benefits are of sharing while addressing potential risks, and facilitating the courage to act, share and engage. Some excuses that have been heard in the past 20 years are visualised in Figure 5.



Top 5 Excuses for Data Sharing Blocking

We are not 'allowed' to share, because:

1. it is our Intellectual Property …
2. of Compliance & Regulatory Restrictions …
3. it is Technically not possible …
4. we have the Policy not to share …
5. we do Not Know How and What to share …

'It is better to offer no excuse than offer a bad one'

George Washington

Figure 5: Top 5 Excuses for Data Sharing Blocking

Avoid unpleasant surprises. Nobody likes unpleasant surprises. While trust is the equation of consistency through time, data sharing and any other engagement, communication, interaction, transaction, relationship, or transformation need transparency, agility, resilience, accountability, and continuous improvement in order to work. The principle of no surprises is one to consider by design and before and during engagements and deployments. Data sharing should be based on trust, and when there is enough trust, it is easy to share.

4.2 The Ethical and Legal Requirements For PISTIS

4.2.1 Data-Centric Perspective

This section discusses legal principles that apply to PISTIS project from the data-centric perspective. This perspective focuses on the legislation concerning various types of data as well as various stakeholders processing this data in various ways. On the one hand, it discusses legal requirements relating to the sharing of non-personal data, whilst discussing provisions addressing the sharing of data held by public sector bodies, and the proposal for accessibility to and availability of data generated by connected devices, on the other. Data and the context it is used in remain the focus of legislation discussed here. Of course, the General Data Protection Regulation also concerns a specific type of data (namely personal data), however, as this regulation focuses on an individual and their rights from a personal perspective, the GDPR is discussed in a separate section below focusing on the people-centric perspective.

It is worth noting that this perspective is highly relevant considering the Commission's "European strategy for data" published in February 2020. In this communication, the Commission has acknowledged the growing data volumes and technological change and has called for measures to help the EU remain competitive in the global market. These include, amongst others, adopting legislative measures on data governance, access and re-use; opening up high value publicly held datasets and allowing their reuse; and enable access to secure, fair and competitive data processing.

4.2.1.1 Free Flow of Non-Personal Data Regulation

4.2.1.1.1 Scope

To retain users and prevent them from leaving to competitors, some digital service providers had adopted practices not allowing users to transfer their data to another provider. This 'vendor lock-in' has been seen as undesirable, as it prohibits users from choosing freely between providers and thereby stifles competition in the internal market. Where in 2018, the GDPR has provided for measures to avoid this from happening with regards to personal data, the Free Flow of Non-Personal Data Regulation [7] (FFoD) promotes portability of non-personal data.

FFoD has been in application since 28 May 2019 and aims at fostering the data economy by facilitating the exchange and storage of electronic, non-personal data across EU borders by removing certain obstacles to its free movement. The overarching purpose is to encourage businesses to share data, provide an underlying legal framework facilitating such sharing, and foster competition in the internal market.

FFoD applies to the processing of electronic data, other than personal data, within the EU, thereby complementing the GDPR. In addition to data processing within the EU, the FFoD is also applicable when the users of the service are in the EU, or when the services are carried out by an entity in the EU [7].

4.2.1.1.2 Key Legal Requirements

With the introduction of FFoD, data localisation requirements are prohibited, unless they serve public security objectives and are justified [7]. Another key implication of FFoD is that competent public authorities can retain access to the data stored and processed in another Member State [7]. Furthermore, FFoD facilitates businesses' ability to switch cloud service providers without losing their data. Under the framework, the Commission encourages providers to adopt codes of conduct setting out key requirements for transfers of data between cloud providers or local IT environments [7].

4.2.1.1.3 Key Takeaways for PISTIS

As PISTIS aims at creating a platform facilitating the monetization and secure transfer of data, the principles enshrined in FFoD are highly relevant for the project and its various demonstrator hubs.

By way of an example, partners participating in the automotive demonstrator hub propose to process and potentially offer for sale certain non-personal data including the records of a specific vehicle's geolocation, speed, acceleration, mileage, and seatbelt status. This dataset is highly reflective of a driver's behaviour on the road. The driver may therefore wish to retain this data or port it to another automotive manufacturer when purchasing a new vehicle, as this data could be assessed by an insurance company when working out the driver's insurance premium. By encouraging the development of self-regulatory codes of conduct for data porting, the FFoD can facilitate an efficient transfer of the driver's data to another automotive manufacturer, potentially achieving the lowest possible insurance premium for them.

Similarly, partners participating in the energy demonstrator hub propose to process and potentially offer for sale certain non-personal data including the records of grid topology, location, energy generation and consumption, and any unusual events. This data can play an important role when a user wishes to switch a utility provider (both in the case of the user consuming or generating electricity). Again, FFoD facilitates an efficient transfer of the user's data set to the desired utility provider. In addition, were the user able to make use of services offered by an entity in another Member State (this may be limited by certain sector-specific national legislation, an assessment of which is outside the scope of this study), then FFoD would prohibit any limits on the transfer of the relevant dataset simply on the ground that it originates from another Member State.

Finally, partners participating in the mobility and urban living hub propose to process and potentially offer for sale certain non-personal data including records concerning the baggage handling management, transfer passenger management, aircraft turnaround process, and public transportation planning support. The FFoD could facilitate the free movement of data between the various stakeholders also in this demonstrator hub. At this same time, however, this demonstrator hub may process certain data that may be deemed sensitive for national security, such as aircraft details, times, and details of any irregularities. It should be noted, that were justified on the grounds of public security and in compliance with the principle of proportionality, the FFoD allows a Member State to prevent free movement of data to another Member State.

4.2.1.2 Data Governance Act

4.2.1.2.1 Scope

Data Governance Act (DGA) [8] is an EU regulation that is part of the European Commission's aim to make Europe fit for the digital age and to facilitate Europe's digital transformation by 2030. This transformation focuses on skills, secure and sustainable digital infrastructures, digital transformation of businesses and digitalisation of public services.

In its *European Strategy for Data*, the Commission has set out the European strategy for creating a single data market to improve its competitiveness. In this strategy, the Commission has described its vision of a common European data space, meaning an internal market for data in which data could be used irrespective of its physical storage location in the Union. Like PISTIS, such data space would essentially consist of several domain-specific data spaces for data sharing and data pooling, including in the fields of health, mobility, manufacturing, financial services, energy or agriculture, or a combination of such areas. Common European data spaces should make data findable, accessible, interoperable, and re-usable (the 'FAIR data principles'), while ensuring a high level of cybersecurity [9].

The DGA plays a central role plays a central role in this strategy. The regulation has been approved by the European Parliament and the Council and will apply from 24 September 2023.

The DGA aims to improve the conditions for the sharing of data between organisations within the EU. The improvement of these conditions will increase the availability of data for research and innovative uses. Since public entities and organisations separately generate large amounts of data daily, the DGA seeks to effectively harness the combined value of this data to facilitate innovation and economic growth. To achieve this, the regulation lays down certain conditions for the re-use of certain categories of data held by public sector bodies. These 'public sector bodies' refer to the State, regional or local authorities, bodies governed by public law or associations formed by one or more such authorities, or one or more such bodies governed by public law [8].

Furthermore, the DGA introduces a framework for the provision of data intermediation services, a framework for the registration of data altruism organisations and a framework for the establishment of a European Data Innovation Board. With these measures, the goal of the DGA is to establish trust between organisations concerning data sharing, to make available public sector data for re-use under certain conditions and overall to generate wealth for society.

4.2.1.2.2 Key Legal Requirements

The DGA does not create actual obligations for public sector bodies to allow re-use of data. It is for every Member State to decide whether certain data is made accessible for re-use. Instead, the DGA delivers a trustworthy framework through which protected data, not regulated by the Open Data Directive, can be re-used and as such be of further benefit to society [10]. Public sectors bodies that are competent under national law to allow or refuse access for the re-use of one or more of the categories of their data are required to publicly make available the conditions for allowing such re-use [8]. The DGA clarifies that such conditions must be non-discriminatory, proportionate, and objectively justified with regard to categories of data and purposes of re-use and the nature of the data for which re-use is allowed. These conditions shall not be used to restrict competition. In the same context, public sector bodies that allow for such re-use may also charge a fee for allowing the re-use of such data provided that such fee shall not be non-discriminatory, proportionate, and objectively justified and shall not restrict competition.

Furthermore, the DGA underscores the key role of providers of data sharing services, which act as data intermediaries, in the data economy. Data intermediaries capable of offering services that can connect the different actors have the potential to contribute to the efficient pooling of data as well as to the facilitation of bilateral data sharing. Conditions for providing data sharing services are also laid out in the DGA which include putting in place procedures to prevent fraudulent practice in relation to access of data, taking adequate technical, organisational, and legal measures to prevent transfer or access to non-personal data that is unlawful and a high level of security for the storage and transmission of non-personal data.

4.2.1.2.3 Key Takeaways for PISTIS

Under the DGA, a data intermediary, or a "data intermediation service" is defined as "a service which aims to establish commercial relationships for the purposes of data sharing between an undetermined number of […] data holders on the one hand and data users on the other, through technical, legal or other means […]" [8]. PISTIS aims to create a platform for secure sharing and trading of data and facilitates the valuation and monetization of data. As it can be seen as a de-centralised service through which a data holder can transfer data to its recipient for value consideration (regardless of any specific demonstrator hub), PISTIS can fall within the scope of the definition of "data intermediation service".

As a result, PISTIS will have to submit a notification to the competent authority for data intermediation services in accordance with DGA [8].Further, PISTIS will have to comply with the conditions for providing data intermediation services set out in the DGA [8].This includes an obligation to ensure that access to its service is fair, transparent and non-discriminatory for both data holders as well as for data users, adopting procedures to prevent fraudulent or abusive practices in relation to its services, and maintaining a log record of the data intermediation activity.

4.2.1.2.4 Special Focus on High-Value Datasets

It has also been recognised that public sector organisations within the EU process high volumes of datasets whose reuse could offer significant benefits for society and the internal market. Hence, the EU has adopted the Open Data Directive [11] requiring Member States to ensure that publicly funded data are reusable for commercial or non-commercial purposes with fair, proportionate, and non-discriminatory conditions. In that regard, it has laid down the legal framework for the re-use of data, including any content or any parts thereof, regardless of whether it is on paper, in electronic form or a sound, visual or audio-visual recording, held by public sector bodies or public undertakings and of publicly funded research data. In its Annex I, the directive defines Therefore, the Directive defines six thematic categories of high-value datasets, namely: (1) geospatial data such as postcodes, national and local maps, (2) data on earth observation and environment, (3) meteorological data, (4) statistics, (5) data on companies and company ownership and (6) mobility data. The Implementing Regulation [12] has laid down the arrangements for publishing and reusing high-value datasets, in particular the applicable conditions for re-use and the minimum requirements for disseminating data via application programming interfaces.

PISTIS aims to provide a platform facilitating secure transfers of data. As such, PISTIS itself is not a public sector body or a public undertaking and is therefore unlikely to trigger the applicability of the Open Data Directive and the Implementing Regulation. However, it is apparent that certain PISTIS participants may themselves qualify as public sector bodies or public undertakings. As a result, certain datasets they may process may represent high-value datasets to which those acts apply. Therefore, general awareness of the requirements is desirable and the relevant PISTIS participants should ensure their compliance (further assessment in this regard falls outside the scope of this deliverable).

4.2.1.3 Proposal for Data Act

4.2.1.3.1 Scope

On 23 February 2022, the Commission proposed the Regulation on harmonised rules on fair access to and use of data (Data Act). Currently, the Council is negotiating with the European Parliament on the proposed regulation. The Data Act lays down rules for allowing access to data generated by connected products, by design and upon request. The Regulation would ensure more fairness in the digital environment, stimulate a competitive data market, open opportunities for data-driven innovation and make data more accessible for all [13]. Also, it aims to give businesses and individuals access and more control over their data through a reinforced portability right. This means that the regulation will facilitate the transferring of data across services, at least where data are generated through connected devices. In addition, customers of cloud data-processing services providers can now effectively switch between providers. Finally, public sector bodies could access and use private sector data in exceptional circumstances, most prominently during public emergencies.

4.2.1.3.2 Key Legal Requirements

Central to the legal requirements under the proposed Data Act are the accessibility to and availability of data generated using products. To that end, the current proposal mainly creates obligations for manufacturers of connected products and data holders of data generated by those products.

4.2.1.3.3 Key Takeaways for PISTIS

The preliminary analysis of the proposal for Data Act has indicated that its requirements will largely apply to manufacturers of connected devices and the data produced. PISTIS aims to provide a platform facilitating secure transfers of data; it is not envisaged that PISTIS will develop any connected devices collecting or generating any data. As a result, while the proposal for Data Act may be directly applicable to certain PISTIS participants, it is unlikely that it will be directly applicable to the way PISTIS is set up and operates.

4.2.1.3.4 Special Focus on B2B Data Sharing

It is envisaged that organisations transferring or exchanging data through PISTIS will fall within the scope of ‘data holders’ and ‘data recipients’ and, as such, they will have to comply with the principles on business-to-business data sharing set out in the (draft) Data Act.

These stipulate that any data transferred on behalf of the user/data subject has to be transferred without undue delay, free of charge to the user, and must be of the same quality as is available to the data holder [8]. This does not oblige PISTIS participants to share any datasets in their original form: while this obligation addresses data sharing following a user’s request, it does not place any conditions or restrictions on the sharing of anonymised or derived datasets (which are mostly the kind of transfers anticipated to take place in PISTIS).

It is important to point out that PISTIS participants won’t be able to, without prior consent, “use any non-personal data generated by the use of the product or related service to derive insights about the economic situation, assets and production methods of or use by the third party that could undermine the commercial position of the third party on the markets in which the third party is active.” In the context of the PISTIS energy demonstrator hub, this may, for example, present an obligation on an energy provider not to obtain any consumption- or generation-related data from a smaller supplier in order to exert any commercial effort or market force with the view of undermining the commercial position of the smaller supplier.

Finally, the Data Act does not make any exemptions to the applicable personal data protection legislation or trade secrets principles. In relation to personal data, it explicitly states that any personal data must only shared where there is a valid legal basis under Article 6 (1) GDPR. As far as trade secrets are concerned, those shall only be disclosed on a need-to-know basis and with necessary safeguards in place. Therefore, data holders within PISTIS should always consider whether they comply with the GDPR (the envisaged GDPR checker will likely assist here) have the necessary consents for disclosing any commercially sensitive information.

4.2.2 People-Centric Perspective

This section outlines the scope and the key principles set out in the General Data Protection Regulation [14] (“GDPR”) and discusses how GDPR applies to PISTIS. GDPR is presented here as the most noteworthy and highly relevant example of people-centric law, i.e., legislation primarily focusing on an individual and information and data about them considering their fundamental rights and freedoms.

The European Declaration on Digital Rights and Principles for the Digital Decade adopted on 26 January 2022 places considerable emphasis on people-centricity. Its Chapter I sets out the key principle of putting people at the centre of digital transformation. By ensuring individuals’ sovereignty and enabling them to stay in control of their personal information, the GDPR aligns with the people-centric approach envisaged by the Declaration.

4.2.2.1 General Data Protection Regulation

The Data Strategy commits to ensure that European fundamental rights and values, particularly the right to the protection of personal data provided under Article 8 of the Charter of Fundamental Rights of the EU (CFREU) and Article 16 of the Treaty on the Functioning of the European Union, underpin all aspects of the Data Strategy and its implementation. In his opinion [15] the European Data Protection Supervisory (EDPS) notes that the GDPR provides a solid basis, also by virtue of its technologically neutral approach, for the development and implementation of a fair and competitive Digital Single Market and it will enable more "data protection-compliant business models" to emerge in the EU. Furthermore, the EDPS acknowledges the crucial role of privacy preserving technologies for the future of European data economy as enablers of data sharing which is both privacy-friendly and socially beneficial. In that regard, as a data intermediary, PISTIS gives a special attention to adhering to data protection principles and compliance with the GDPR requirements.

4.2.2.1.1 Scope

GDPR lays down the rules for the protection of individuals' personal data. It is doing so, it aims to protect certain fundamental rights and freedoms regarding the protection of personal data, as laid down in the European Convention for Human Rights and the Charter of Fundamental Rights of the European Union. GDPR applies to any processing of personal data by a person or an entity within the EU, as well as any processing outside the EU that relates to personal data of individuals in the EU. The second condition only applies when either the processing activities relate to the offering of goods or services to individuals withing the EU, or when the monitoring of their behaviour takes place within the EU. GDPR does not apply to certain circumstances laid down in Article 2, such as the processing of personal data by a natural person during a personal or household activity.

According to Article 4 of the GDPR, "personal data" is defined as any information relating to an identified or identifiable natural person. This natural person is referred to in the GDPR as the 'data subject'. This definition of personal data is deliberately quite broad to cover a wide range of a person's information and, in doing so, provide an appropriate and proportionate guarantee of privacy. "Identifiable" is a key word within the concept of personal data, since it means that it is not necessary that the data subject is directly named in the data (set) for data to be considered "personal". The identifiers named in the paragraph are "*a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*".[32] Another important definition to note is 'processing', meaning any action performed on personal data, including but not limited to collection, storage, and alteration.

Many stakeholders are addressed in the GDPR, but in essence the most important factors to consider are the controllers, processors, data recipients and the data subjects. The Regulation designates certain obligations or rights to each role. A "controller" is a natural or legal person, public authority, agency, or other body which determines the purposes and means of the processing of personal data. He can do this alone or jointly with others, in which case the controllers are 'joint controllers. The 'processor' is a natural or legal person, public authority, agency, or other body which processes the personal data on behalf of the controller. The responsibilities of the controller and the processor are laid down in Chapter IV of the GDPR.

4.2.2.1.2 Data Protection Principles

Rapid technological developments and globalisation have increased the scale of collection and sharing of personal data significantly. Both companies and public entities can use these data on a unprecedently large scale. However, this increase of data processing requires a coherent data protection framework to create trust among EU citizens and ensure that the citizens, or data subjects, have control over their own data. The GDPR has therefore laid down requirements for controllers and processors. First, the GDPR stipulates that for processing to be lawful, there needs to be a lawful basis. A lawful basis can be consent of a data subject, performance of a contract, legitimate interest, or the protection of vital interests of data subjects or another person or in public interest.

Article 5 of the GDPR lays down six key principles that form a critical part of the data protection regime, namely:

a. Lawfulness, fairness, and transparency.

b. Purpose limitation.

c. Data minimisation.

d. Accuracy.

e. Storage limitation.

f. Integrity and confidentiality.

The data controller is responsible for these six principles regarding personal data and needs to be able to always demonstrate compliance with these. Simultaneously, the GDPR provides certain rights to individuals as well. These include the right to access of their personal data, the right to be forgotten, right to rectification and the right to object to processing of personal data for certain purposes. Again, these rights create obligations for the controllers of personal data. No specific measures are required, but Article 24 states that controllers shall implement appropriate technical and organisational measures to abide by this Regulation.

These measures are to be chosen based on the state of the art, costs of implementation and the characteristics and risks of the processing activities. Additionally, the GDPR establishes the concept of data protection by design and by default. This mandates organizations to incorporate privacy safeguards into the design of their business operations, processes, and services. Another important provision is the requirement for conducting data protection impact assessments in specific situations where processing is likely to pose a significant risk to the rights of data subjects. These assessments may require pre-consultation with the relevant supervisory authority.

4.2.2.1.3 The Function of PISTIS Solution and the Respective GDPR Roles

Before identifying the legal requirements and obligations in the GDPR, the multiple functions of PISTIS Solution in data processing activities and the GDPR roles of the stakeholders in the data lifecycle over PISTIS Platform should be examined. The European Data Protection Board clarifies the precise meaning of the concepts of data controller and processor, and the criteria for their correct interpretation under its Guidelines on the concepts of controller and processor in the GDPR numbered 07/2020 in July 2021 [16] .In line with the guidelines, this section interprets the envisaged functions of PISTIS and the use cases to the extent that they have been described by the demonstration partners during the workshops and discusses the potential GDPR roles in accordance with this interpretation.

During the demonstration hub workshops, it is noted that the PISTIS will be used only by businesses, i.e., the demonstration partners and individual data subjects do not interact with the platform. This means that the platform will not be used to collect personal data from individuals and the end user will use the platform to evaluate, share and trade their datasets containing personal data which have already been collected by them. In the context of the PISTIS platform deployment, the end-users will have both the roles of data provider and data user/receiver depending on whether they provide or acquire datasets and the platform will be used as means of data processing for the end users.

From the perspective of the platform operator, who oversees the governance of the platform and oversees provision of the services to the end users, the two core parts of PISTIS Solution consists, Data Space Factory and FAIR Data Trading & Value Exchange Monetisation Platform has distinct functions. As explained in the chapters above, PISTIS data space factory is a mere product that will be used by the end-users to store their datasets whereas the FAIR data trading and value exchange monetisation platform is data processing service provided to the end users by the platform operator. Therefore, considering the function of PISTIS FAIR Data Trading & Value Exchange Monetisation Platform, the platform operator could be seen as a service provider. These different functions of the PISTIS Solution in the data lifecycle are visualised in the figure below. The GDPR roles, e.g., data controller and data processor, will be attributed to the end-users and the platform operators considering the different functions of PISTIS Solution.

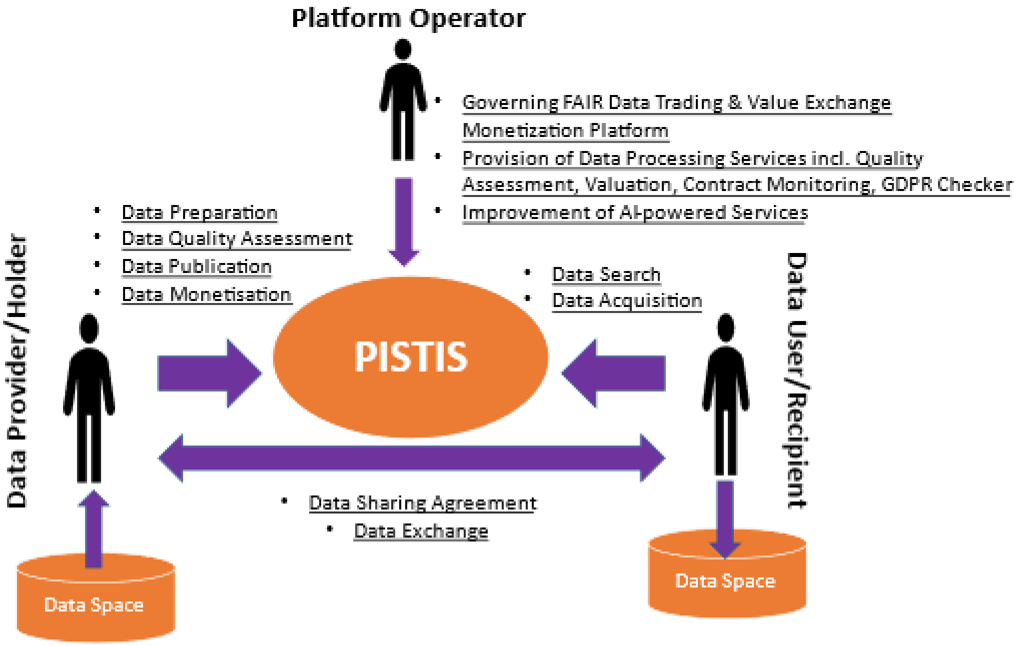


Figure 6: Overview of PISTIS Functions and Roles

The end-users including the demonstrator partners regardless of whether they act as data provider or data user, can be considered as a separate data controller in their respective data processing activities including data preparation, quality assessment and exchange of datasets containing personal data as they individually determine the purpose and the means of data processing. For the data controllers function as a mean of data processing.

The platform operator which provides PISTIS services to the data controller on PISTIS FAIR Data Trading & Value Exchange Monetisation Platform is likely to be qualified as the data processor of each end-user. Because the platform services will entail processing of personal data for the purpose determined by the end-users. On the other hand, if the data provided by the end-user is further used to improve AI-powered PISTIS services, for example as training dataset for AI, this set of processing may qualify the platform operator as data controller.

4.2.2.1.4 Key Legal Requirements for Data Protection and Privacy

The concepts of controller and processor play a crucial role in the application of the GDPR, since they determine who is responsible for compliance with different data protection rules, and how data subjects can exercise their rights in practice. According to the different role the partners have in the processing of personal data, the GDPR allocates responsibility for compliance and imposes specific rules.

Considering that the demonstrators are likely to be considered as a separate data controller, they have the primary responsibility to demonstrate compliance with the abovementioned data protection principles and the relevant GDPR requirements. For example, before making their datasets containing personal data available in PISTIS, data controller demonstrator should individually ensure that the processing of personal data is carried out in accordance with one of the legal bases set out in Article 6 (1) of the GDPR. However, PISTIS may assists the demonstrator partners to comply with the specific GDPR requirements concerning data disclosure or transfer of personal data. In that regard, this section covers the requirements that become relevant to the end-user due to their use of PISTIS.

First, pursuant to the transparency principle, data controllers must inform the data subjects concerned if and how their personal data will be transferred to a third party on PISTIS.This information on a disclosure to another recipient should be provided to data subject at the latest when the personal data are first disclosed. Therefore, it is recommended that certain privacy clauses should be inserted in data sharing contracts, data processing agreements or general terms and conditions to be used in PISTIS by the respective demonstrator partners to inform contracting partners of how they process personal received and transmitted to each other. This will support the data provider/holder to comply with this transparency requirement.

Secondly, as a part of accountability principle, whenever a data controller uses a processor to process personal data on their behalf, Article 28 of the GDPR requires that data controller to have a written data processing agreement in place with the processor. According to Recital 81 of the GDPR, data processing agreement should consider “the specific tasks and responsibilities of the processor in the context of the processing to be carried out and the risk to the rights and freedoms of the data subject” and include specific terms or clauses regarding the following issues:

- the processor will process personal data only on written instructions of the data controller.
- the processor uses all appropriate technical and organizational measures to ensure the security of the data.

- (If applicable) the processor will not subcontract to another processor unless instructed to do so in writing by data controller;
- the processor will assist the data controller to uphold their obligations under the GDPR, particularly concerning data subjects' rights.
- the processor agrees to delete all personal data upon the termination of services or return the data to the data controller.
- audits and inspections.

Furthermore, according to the EDPB's interpretation of the mandatory content requirement, data processing agreement needs to specify:

- the subject-matter of the processing (for instance, data sharing and trading, dataset valuation, etc) including the type of operations performed as part of the processing and purpose of the processing: this should be as comprehensive as possible, depending on the specific processing activity, so as to allow external parties (e.g., supervisory authorities) to understand the content and the risks of the processing entrusted to the processor.
- the duration of the processing on PISTIS
- the type of personal data: this should be specified in the most detailed manner as possible, for instance geolocation data of car drivers, smart meter readings.
- the categories of data subjects: this, too, should be indicated in a quite specific way for instance, car drivers, employees of Athens International Airport, etc.
- the obligations and rights of the controller: for example, the right of the controller to perform inspections and audits and as regards the obligations of the controller, examples include the controller's obligation to provide the processor with the data mentioned in the contract, to provide and document any instruction bearing on the processing of data by the processor, to ensure, before and throughout the processing, compliance with the obligations set out in the GDPR on the processor's part, to supervise the processing, including by conducting audits and inspections with the processor [17].

As explained in the previous section, in case PISTIS will be operated by a platform operator, it is likely to be considered as data processor and therefore a data processing agreement needs to be executed between the platform operator and the controller end-users. As regards assisting the data controller, it is important that PISTIS implements appropriate technical and organisational measures to ensure the security of personal data, including protecting against accidental or unlawful destruction or loss, alteration, unauthorised disclosure, or access. In that regard, any data breaches and potential data protection infringements occurred in PISTIS should be communicated to the end-users without undue delay.

Furthermore, a data processor is also required to comply with other GDPR accountability requirements, such as maintaining records of processing activities carried out on behalf of the end-users and appointing a data protection officer where applicable. It is recommended for PISTIS to have technical components that could automatically record all data processing activities carried out on the platform such as data valuation and make the records available to the respective end-users. As for the requirement to designate a data protection officer, the platform operator should first assess whether PISTIS is likely to be used to process special categories of personal data. If this is the case, the platform operator needs to designate a data protection officer who could oversee the compliance of the processing operations in PISTIS with the GDPR.

Article 25 of the GDPR requires that the data controller, considering all the elements of processing, puts in place adequate technical and organisational measures, which have also to be demonstrated, to prove that the processing has been carried out in compliance with the GDPR principles particularly data minimisation and data integrity and confidentiality. In that regard, PISTIS provides the end-users with the necessary technical tools such as data anonymisation, encryption components to assist the end-users in compliance with the GDPR. Lastly, as a data processor, the platform operator is also subject to the restriction of transfer of personal data to outside of the Union. Based on the input received during the workshops, it is not expected PISTIS to be used to transfer personal data provided by the end-users to a third country.

Finally, in case when the platform operator acts as data controller, for example processing personal data to improve the platform AI-driven services, it must comply with main GDPR obligations and execute its operations in compliance with the data protection principles described in the previous section. It means that this specific set of processing activities must be based on one of the lawful grounds listed in Article 6 and it has to be carried out following a specified, explicit and legitimate purpose. As for the accountability principle, the platform operator will be required to adopt and demonstrate that appropriate technical and organizational measures have been taken and implemented during the whole process.

Moreover, to comply with the transparency principle, it is recommended that the information stipulated in Article 14 of the GDPR should be made available to any PISTIS users. Following Article 32, the platform operators also must ensure a level of security that is adequate to the risk for the rights and freedoms of data subjects that can occur during processing activities in place. Specifically, the platform operator could adopt risk mitigation strategies and risk analysis should be developed at the technical and management level.

Data Protection Impact Assessment under the GDPR

A data protection impact assessment (DPIA) is a way for data controllers to analyse their data processing operations and help them identify and assess the likelihood and severity of risks to the rights and freedoms of individuals comprehensively and systematically. As per Article 35 of the GDPR, in principle, a DPIA is required where personal data processing "is likely to result in a high risk to the rights and freedoms of natural persons"[2]. The GDPR does not provide for

[2] It should be noted that there are exemptions to this requirement under Article 35 of the GDPR. For instance, the data controller is not required to carry out a DPIA if the nature, scope, context and purposes of the envisaged processing are very similar to the processing activity for which DPIA have already been carried out or, if a

any specific risk assessment model, instead it only states that an objective assessment should be carried out to assess whether processing operations involve a risk or a high risk [18]. In that regard, this requirement in particular applies to large-scale processing operations which aim to process a considerable amount of personal data at regional, national or supranational level and which could affect a large number of data subjects and which are likely to result in a high risk, for example, on account of their sensitivity, where in accordance with the achieved state of technological knowledge a new technology is used on a large scale as well as to other processing operations which result in a high risk to the rights and freedoms of data subjects, in particular where those operations render it more difficult for data subjects to exercise their rights [19].

To guide data controller in determination of whether data processing is “likely to result in a high risk”, the EDBP provides data controllers with a list of nine criteria together with some examples in its Guidelines on Data Protection Impact Assessment [20].According to the EDBP, personal data processing meeting two or more of these criteria are more likely to constitute a high risk to the rights and freedoms of individuals, and therefore the controllers are required to carry out a DPIA. However, the EDBP notes that in some cases where the risk is too serious, even only one of these criteria may obliges the data controller to carry out a DPIA. EDBP’s criteria together with brief descriptions and examples is made available in Appendix B.

Moreover, it should be noted that if the envisaged processing operation may correspond to one of the cases in Appendix B and thus such processing is considered “not likely to result in a high risk”, the data controller must be able to justify and document the reasons for not carrying out a DPIA. It is always recommended for data controller to obtain the view of the data protection officer, if any, prior to initiating data processing activities. Furthermore, given that national data protection authorities are required to establish, make public and communicate a list of the processing operations that require a DPIA in line with these nine criteria, as per Article 57 of the GDPR, data controllers should consider those lists published by the competent national data protection authority when assessing whether their envisaged data processing require a DPIA.

In context of PISTIS, DPIA could be considered as a useful tool for identifying and mitigating the risks arising from the data processing activities. It is thus suggested for the partners developing and operating the platform to evaluate whether their processing operations are likely to result in a high risk in accordance with the EDBP’s criteria provided in Appendix-B then to document their findings and the reasons for not carrying our DPIA. This will allow the partners to mitigate potential adverse effects of the processing activities on the participants and to implement proper safeguards for the rights and freedom of the participants. However, even if data processing operations do not trigger this legal requirement, carrying out a risk assessment prior to commencing data processing is always recommended. Nevertheless, it should be noted that DPIA is not only a key part of compliance with the GDPR where high risk data processing is planned but also allows organisations to proactively address to safeguard the rights and freedom of individuals.

processing activity has a legal basis in EU or Member State law, where the law regulated the specific processing operation.

In case the platform operator decides to carry out a DPIA, it must include at least: i) a description of the envisaged processing operations and the purposes of the processing; ii) an assessment of the necessity and proportionality of the processing; iii) an assessment of the risks to the rights and freedoms of data subjects and iv) the measures envisaged to mitigate the identified risks and to demonstrate compliance with the GDPR [21].DPIA should be completed as early as is practicable in the design of the processing operation and updated throughout the lifecycle of the processing operation. DPIA is an on-going process, not a one-time exercise and therefore it should be continuously reviewed, and the risk should regularly be re-assessed by the controller [22].

4.2.2.1.5 Key Takeaways for PISTIS

PISTIS envisages that participants active in its various demonstrator hubs will likely process certain personal data. For example, information about specific needs passengers is processed in the mobility and urban planning demonstrator hub, and a range of vehicle-specific data that may in aggregate lead to an identifiable individual is processed in the automotive demonstrator hub. As PISTIS aims to create a platform for secure transfer of data, it is likely that certain personal data may be transferred as well.

To ensure compliance with the GDPR an inclusion of a semi-automatic 'GDPR Checker' is considered under WP2 to review the type of data to be shared via PISTIS and the purpose of such sharing.

In general, where personal data will be shared via PISTIS, participants must ensure that such sharing is lawful, i.e., carried out in accordance with one of the legal bases set out in Article 6 (1) of the GDPR. Participants are mostly considering relying on data subject's consent with the processing and sharing of their personal data. Whilst consent presents a valid legal basis, it must be noted that several criteria must be met for a consent to be valid. Amongst others, consent should be clear, freely given, specific, informed, and withdrawable at any time. As these criteria may be arduous to meet, it is generally advisable to rely on other legal basis allowing a controller to process personal data without the risk that the consent may be found invalid or withdrawn.

Given the legal complexities in the processing of personal data, PISTIS participants should always carefully assess whether it is necessary to process an individual's personal data to reach the intended purpose, or whether the same purpose can be reached with anonymised data. A similar assessment should also be carried out when considering data transfers via PISTIS platform. Such approach is also required by the data minimisation principle set out in Article 5(1)(c) of the GDPR.

4.2.3 Market-Centric Perspective

When considering the offering and obtaining of services online, there are certain actors that play an ever more significant role in the digital ecosystem. This section focuses on the legal regulation of such significant market players, as their significance may rise with the progressing digital developments. This section discusses legislation aimed at the regulation of so-called gatekeepers, certain intermediary services, as well as entities offering crypto assets. Regard is also given to a brief analysis of the expected applicability of that legislation to PISTIS.

4.2.3.1 Digital Markets Act

4.2.3.1.1 Scope

Acknowledging the growing market power of certain large-scale providers of digital services, recognising the impact they and their services may have on users' lives, the Digital Markets Act [23] (DMA) aims to make the markets in the digital sector fairer and more contestable to ensure that the digital space continues to serve business users as well as end users. The DMA has become applicable on 2 May 2023.

To that end, DMA establishes and defines the criteria for "gatekeepers", i.e., large digital platforms providing so called core platform services, such as online search engines, app stores and messenger services. Gatekeepers are undertakings:

a. having a significant impact on the internal market (providing services in at least three Member States, with an annual turnover of 7.5 billion EUR or market capitalisation of 75 billion EUR),
b. providing a core platform service which is an important gateway for business users to reach end users (at least 45 million monthly active end users established or located in the EU), and
c. enjoying an entrenched and durable position in its operations. DMA then specifies certain conditions and restrictions that apply to gatekeepers.

Further, the DMA also grants certain investigative and monitoring powers to the European Commission. These include powers to examine whether an undertaking should be designated as a gatekeeper, to examine possible systematic non-compliance, as well as to investigate new services and practices which may be unfair, and which are not effectively addressed by the DMA [24].

Finally, the DMA has introduced significant and detrimental fines for non-compliance. Gatekeepers may face fines up to 10% of their annual worldwide turnover for failing to comply with the provisions of the DMA. In addition, if the Commission finds that an undertaking has committed the same of a similar undertaking it may issue an additional fine of up to 20% of the company's annual worldwide turnover.

4.2.3.1.2 Key Legal Requirements

Due to its objective criteria, the DMA may apply to a variety of digital services, including cloud storage services, social platforms, instant messaging services as well as mass media. Whist it sets out an extensive list of obligations with potential relevance for and applicability to each of the above, the following paragraphs focus on obligations potentially relevant to PISTIS as a data exchange platform.

First, gatekeepers are obliged to notify the Commission of their gatekeeper status and specify the criteria which their notification is based on. The Commission may also designate a gatekeeper based on its own investigation.

Secondly, a gatekeeper is not permitted to cross-use personal data from the its core platform service in other services provided separately by the gatekeeper. Similarly, the gatekeeper is not permitted to prevent business users from offering the same products or services to end users through third-party online intermediation services or through their own direct online sales channel at prices or conditions that are different from those offered through the online intermediation services of the gatekeeper. Further, no gatekeeper is permitted to require its users to only use a specific payment method of that service provider. Every gatekeeper must also make it easy for end users to switch to another (possibly competing) service that is accessed using the core platform of the gatekeeper. It can be seen that all the above requirements are aimed at encouraging a greater choice of services in the market and increasing competition.

Finally, the DMA requires gatekeepers to avoid any disproportionate termination clauses that would be difficult to exercise. Again, this provision is in line with the overarching theme of facilitating easier switching of competing services, providing greater choice to businesses as well as end users.

4.2.3.1.3 Key Takeaways for PISTIS

The DMA contains several requirements that could be highly relevant for a platform providing data sharing services, such as PISTIS. These include and go beyond the requirements discussed above.

However, the regulation only applies to entities that satisfy the three key criteria resulting in their classification as gatekeepers. Although PISTIS is an ambitious project involving leaders in their respective fields and leveraging and building some of the most up-to-date technologies, it is unlikely that it will operate with an annual turnover of any significant value. Similarly, given that it currently only has about 30 contributing partner organisations, it is unlikely that it will reach the benchmark 45 million monthly active users in the first few years of its operations. Without the need to assess the third requirement, it can be safely said that the project is highly unlikely to meet the required criteria to be considered a gatekeeper. As a result, the requirements set out in the DMA won't apply to PISTIS unless and until it meets the required criteria.

4.2.3.2 Digital Services Act

4.2.3.2.1 Scope

The Digital Services Act (DSA) [25] follows the Commission's aims to better protect consumers and their fundamental rights online, establish a powerful transparency and clear accountability framework for online platforms, and foster innovation, growth, and competitiveness within the single market.

Whilst the DMA focuses on the role of online gatekeepers, and their responsibilities and obligations, the Digital Services Act (DSA) complements it by focusing on intermediary services. These include 'mere conduit' services that facilitate access to or transmission of data in a communication network, as well as 'hosting' services providing data storage.

The DSA addresses three key legal areas in relation to intermediary services:

a. it puts forward a framework for the conditional exemption from liability of providers of intermediary services,

b. it introduces rules on specific due diligence obligations tailored to certain specific categories of providers of intermediary services,
c. it introduces rules on its implementation and enforcement [26].The DSA also establishes the European Board for Digital Services (the "Board") to supervise the providers of intermediary services, contribute to the consistent application of the regulation, and provide guidance on its application.

Finally, the DSA imposes an obligation on Member States to set out the rules on penalties applicable to infringements of the regulation. The maximum amount of fines a Member State can issue is 6% of the provider's annual worldwide turnover [27].Note that the Commission may itself issue fines to providers of very large online platforms (e.g., platforms with an average of more than 45 million monthly active users) not exceeding 6% of their annual worldwide turnover.

4.2.3.2.2 Key Legal Requirements

The DSA applies to a wide range of intermediary services and, as such, it sets out obligations covering the entire scope of services it applies to (including very large online platforms, 'caching' services, or advertising on online platforms). This sub-section focuses on the requirements and principles relevant for the PISTIS project and does not aim to provide an exhaustive overview of legal requirements put forward by the regulation.

The DSA exempts liability for services providing mere transmission of or access to data in a communication network, where they don't play any active part in such transmission (i.e., they don't initiate the transmission, change the information, or select its receiver) – such services are referred to as "mere conduit" services. Similarly, the regulation makes hosting services exempt from liability where they don't have actual knowledge of illegal activity or content, and, when notified, they act expeditiously to remove such content. While there is no general obligation to monitor the transmitted content, a service provider is obliged to act when notified, or when issued with an order by a relevant authority.

Further, the DSA sets out a number of due diligence obligations that intermediary services are expected to comply with to facilitate a transparent and safe online environment.

First, intermediary services are expected to designate and notify the Commission and the Board of their single point of contact for communicating with authorities as well as with recipients of the service.

Secondly, providers of hosting services should enable their users to notify them of any potentially illegal content. Any implemented notification mechanism should be easily accessible and user-friendly. Upon their investigation, the provider should inform any affected user about any service restrictions arising from their transmission of illegal content or non-compliance with the service terms and conditions. A provider of hosting services is also obliged to notify the relevant law enforcement authorities where they suspect that a transmission may give rise to a criminal offence threatening an individual's life or safety [28].

It is also worth noting that providers of online platforms are obliged to establish and operate an internal complaint-handling system, enabling users to lodge complaints electronically and free of charge, against any decision taken by the provider and they are expected to submit to out-of-court dispute settlement.

Finally, the DSA makes a provision for the Commission and the Board to remain active in the implementation of the regulation and providing ongoing guidance and monitoring. Specifically, the regulation expects the Commission and the Board to consult and draw up or facilitate various standards, codes of conduct, codes of conduct for accessibility, and crisis protocols.

4.2.3.2.3 Key Takeaways for PISTIS

PISTIS aims at creating a platform facilitating the monetization and secure transfer of data. It is envisaged that PISTIS will operate as a decentralised data exchange platform, merely enabling various participants to connect and exchange data directly, as opposed to creating a data repository. In other words, PISTIS will provide access to network where participants can transmit data (information).

As a result, PISTIS will likely fall within the definition of “intermediary service” and will certain principles set out in the DSA will apply to it. On the one hand, PISTIS will be able to rely on the exemption from liability mechanism (unless notified). On the other hand, PISTIS will have to properly comply with the various transparency requirements and adopts the prescribed complaints handling procedures. It is also advisable that PISTIS regularly monitors the Commission’s and the Board’s activities in relation to the implementation of the DSA and follows any guidance or adopted model codes of conduct.

Finally, it is worth pointing out that the principle of exemption from liability of intermediary service providers has already been introduced by E-Commerce Directive, setting out similar criteria [29]. Although E-Commerce Directive specifically refers to intermediary service providers, it does not provide further clarity on what entities “intermediary service providers” entails. By defining this term (as well as the ancillary terms of “mere conduit service” and ‘hosting service’) and re-iterating the principle of exemption from liability in relation to intermediary service providers, the DSA has removed much ambiguity around the term, and, as a result, it is now clear that the principle of exemption from liability will apply to PISTIS.

4.2.3.3 Proposal for a Markets in Crypto-assets Regulation

4.2.3.3.1 Scope

In recent years there has been a considerable surge in crypto assets, particularly 'stablecoins', which incorporate stabilising mechanisms against major currencies, including the euro and the US dollar. The EU has recognised that although the stablecoin market is still modest in size and is unlikely to pose systemic threats in its current state, it may expand substantially in the future. Whilst the currently applicable rules also apply to crypto-assets, they do not cover all crypto-assets and their application varies across the EU Member States.

The proposed regulation on markets in crypto-assets [30] (MiCA) would harmonise rules for crypto-assets at EU level, thereby providing legal certainty for crypto-assets not covered by existing EU legislation. Specifically, MiCA “lays down uniform requirements for the offer to the public and admission to trading on a trading platform of crypto-assets other than asset-referenced tokens and e-money tokens, of asset-referenced tokens and of e-money tokens, as well as requirements for crypto-asset service providers.” [31]. By enhancing the protection of consumers and investors as well as financial stability, the regulation would promote innovation and use of crypto assets. MiCA would apply to all parties engages in the trading of crypto assets within the EU.

The proposal has been voted in plenary and was adopted by Council on 16 May 2023 after Parliament’s first reading and is currently awaiting signature. When signed and published, MiCA is expected to enter into application 18 months after the date of its entry into force, i.e., no sooner than around January 2025.

4.2.3.3.2 Key Legal Requirements

MiCA lays down certain legal requirements on persons offering to public a crypto asset, defined as “a digital representation of a value or of a right that is able to be transferred and stored electronically using distributed ledger technology or similar technology.” Those requirements primarily concern the issuance of a white paper by the offeror of the crypto-asset and their honest and professional conduct.

MiCA also aims to reinforce certain safety measures for crypto-assets, including the allocation of key reserves, so that the assets are backed by an equal value in reference currencies (1:1 rule). The issuer of crypto-assets must also provide a redemption plan in case of distress, so that crypto-asset holders are guaranteed to receive equivalent currencies. Redemption is also set without delay, to avoid liquidity issues. The text reaffirms the EBA as the lead regulator and supervisor, while relevant information would be transmitted to the European Securities and Markets Authority (ESMA) [32].

4.2.3.3.3 Key Takeaways for PISTIS

It is envisaged that PISTIS platform will use a stablecoin (i.e., a crypto asset) as means of payment for data transfers.

It should be noted that MiCA will not apply to crypto-assets that are not fungible, or interchangeable, with other crypto-assets. In other words, if the stablecoin introduced and used in the context of PISTIS platform only serves as value consideration for data transfers on the platform and cannot be exchanged for another crypto-assets, then MiCA will not apply.

On the other hand, if the PISTIS stablecoin will be interchangeable or tradeable for other crypto assets, then MiCA will likely apply. This means, amongst others, that an operator of PISTIS platform will be required to obtain legal personality, it will have to produce a white paper in respect of its stablecoin, notify the stablecoin to a competent authority in the relevant Member State, comply with requirements relating to its marketing communications, and maintain all their systems and security access protocols in conformity with the appropriate EU standards.

4.2.4 System-Centric Perspective

The rapid acceptance and natural uptake of certain products of the ongoing digital developments have resulted in several undesirable systemic trends and presented several challenges for the digital ecosystem. This section considers EU legislation that has been proposed or adopted to address those systemic trends and challenges, ensuring safe and secure digital environment. This section discusses proposed and adopted legislation on cyber security across the EU, the protection of customers purchasing products with digital components, and the regulation of artificial intelligence based on its potential to cause harm. Regard is also given to a brief analysis of the expected applicability of that legislation to PISTIS.

4.2.4.1 NIS 2 Directive

4.2.4.1.1 Scope

Directive on measures for a high common level of cybersecurity across the EU, repealing Directive (EU) 2016/1148 (“NIS2 Directive”) [33] entered into force on 16 January 2023 and Member States now have until 17 October 2024 to transpose its measures into national law.

NIS2 Directive aims to extend and future-proof the scope of its predecessor, NIS Directive, to keep up with the unprecedented digitalisation that has occurred in the last few years. To achieve high common level of cybersecurity across the EU, the Directive lays down:

a. obligations requiring Member States to adopt national cybersecurity strategies and designate the relevant competent authorities,
b. cybersecurity risk-management measures and reporting obligations for certain entities,
c. rules on cybersecurity information sharing, and
d. supervisory and enforcement obligations on Member states.

NIS2 Directive requires Member States to adopt certain cybersecurity measures in relation to ‘essential’ and ‘important’ entities.

Essential entities include digital infrastructure providers (as a content delivery network provider, or provider of public electronic communications networks), as well as providers of public electronic communications networks or of publicly available electronic communications services, provided (in each case) they employ more than 250 persons and which have an annual turnover exceeding EUR 50 million, and/or an annual balance sheet total exceeding EUR 43 million [34]. A Member State may also identify an entity as essential entity.

Important entities include all entities in categories listed in Annexes I and II of the directives that do not meet the above thresholds. Categories relevant for PISTIS include digital infrastructure providers, ICT service management, digital providers, and research.

4.2.4.1.2 Key Legal Requirements for Cybersecurity

Under NIS2 Directive, Member States are required to establish a list of essential and important entities and update it at least once in every two years. Those shall therefore be required to notify their details to the competent authorities.

Member States are then required to ensure that essential and important entities take appropriate and proportionate technical, operational, and organisational measures to manage the risks posed to the security of network and information systems which they use for their operations, and to prevent or minimise the impact of incidents on recipients of their services and on other services. These measures include policies on information systems security, incident handling, supply chain security and the application of basic cyber hygiene practices, amongst others.

Further, Member States are required to ensure that essential and important entities notify, national computer security incident response team (CSIRT) or, where applicable, its competent authority of any incident that has a significant impact on the provision of their services. CSIRTs are in turn expected to share their findings with national competent authorities as well as with other CSIRTs.

4.2.4.1.3 Key Takeaways for PISTIS

Given the extended scope of essential and important entities that are subject to security requirements laid down in NIS2 Directive, the directive may be relevant for PISTIS platform itself, as well as its various participants.

PISTIS platform could qualify as digital infrastructure (if it is considered a content delivery network provider, or provider of public electronic communications networks), ICT service management (if it is considered a managed service provider), a digital provider (for example, a provider of an online marketplace), or a research organisation. Therefore, extra care should be taken when defining the platform, its purposes, as well as the way it operates. Given the fact that PISTIS will unlikely meet the employee figures and revenue threshold to make it automatically qualify as an essential service, it should be noted that a Member State’s competent authority may regard it as an essential entity, nonetheless. If regarded an essential entity, PISTIS will then be expected to comply with the requirements set out in Article 21 of NIS2 Directive, as transposed into the relevant national law.

NIS2 Directive also lists certain types of essential and important entities in a range of sectors which PISTIS participants are active in, and which are relevant for the individual demonstrator hubs. For example, participants in the mobility and urban living demonstrator hub should assess whether they fall into the transport category. Similarly, participants in the energy demonstrator hub should see whether they are an essential or important entity in the energy sector. Finally, there is a separate category for manufacturers of motor vehicles and electrical equipment, so participants in the automotive demonstrator hub should assess whether they fall into one of those categories. An assessment of the directive’s applicability to individual participants falls outside the scope of this deliverable.

4.2.4.2 The Proposal for a Cyber Resilience Act

4.2.4.2.1 Scope

The proposed Cyber Resilience Act (CRA) [35] aims to protect consumers and businesses buying or using products with digital components and does that by laying down rules for those products regarding the design, development, production, placing on the market and even creates obligations for manufacturers of digital products to ensure their cybersecurity during the entire product lifecycle (with a maximum of five years).

CRA will apply to all products with digital elements that do not have a sectoral exemption, such as medical devices and airplanes. The term ‘product with digital elements’ is defined as ‘*any software or hardware product and its remote data processing solutions, including software or hardware components to be placed on the market separately’,* thus it will likely encompass a wide range of devices. As the data offered for sale and transferred via PISTIS platform may be collected or generated by such devices engaged by various PISTIS participants (in the context of various demonstrator hubs), CRA should be considered considering its possible relevance and applicability to PISTIS.

The legislative proposal for CRA was presented by the Commission on 15 September 2022. In the Parliament, the file has been assigned to the Committee on Industry, Research and Energy (ITRE). The draft report was published on 31 March 2023. No further development has been noted.

4.2.4.2.2 Key Legal Requirements

CRA states that products with digital elements may only be made available on the market where both the products and the processes put in place by the manufacturer comply with the essential requirements respectively set out in Section 1 and 2 of Annex I [36]. These requirements partially depend on the outcomes of the risk assessment undertaken by manufacturers. Some security requirements include an appropriate level of cybersecurity by design, minimisation of data and protecting the availability of essential functions. CRA also lays down additional requirements for critical products with digital elements. The exhaustive list of products which can be designated as critical products with digital elements is set out in Annex II of CRA.

4.2.4.2.3 Key Takeaways for PISTIS

CRA is mainly aimed on the placing on the market of products with digital elements and the aftercare by manufacturers. As PISTIS platform mostly concerns transfers of certain specific existing datasets in the business-to-business context, rather than any software of hardware products, CRA may not be directly applicable to PISTIS as such.

However, it may be directly relevant and applicable to products placed on the market by participants in various demonstrator hubs which are used to collect or generate those datasets. For example, CRA will likely be relevant to manufacturers of connected vehicles or devices used withing them, as such devices will likely fall within the wide definition of ‘products with digital elements. However, any assessment of CRA’s applicability to individual PISTIS participants is outside the scope of this deliverable.

4.2.4.3 Proposal for AI Act

4.2.4.3.1 Scope

The proposal for harmonised rules on artificial intelligence [37] (the AI Act) aims to regulate artificial intelligence (AI) based on its potential to cause harm. More specifically, the AI Act proposes to lay down:

a. harmonised rules for the placing on the market and putting into service AI systems in the EU;
b. rules on prohibiting certain AI practices;
c. specific requirements for high-risk AI systems and obligations for operators of such systems;
d. harmonised transparency rules for certain AI systems;
e. rules on market monitoring, market surveillance, governance and enforcement;
f. measures to support innovation; and

g. rules for the establishment and functioning of the EU AI Office [38]. The AI Act is intended to apply to providers of AI systems in the EU irrespective of whether they are based in the EU or not.

EU lawmakers are aware of the need to adopt a technology-neutral definition of AI to ensure effective application of the rules. They have proposed to align the AI definition with that of the Organisation for Economic Development and Cooperation. As a result, AI is defined as *"a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments."* [38] *.*

On 11 May 2023, the European Parliament's Civil Liberties and Internal Market committees jointly adopted the proposed text by large majority. The plenary adoption is expected to take place on 14 June. Once adopted, trialogue negotiations between the Parliament, EU Council and the Commission will take place.

4.2.4.3.2 Key Legal Requirements

Recognising that the use of AI can affect a number of fundamental rights and users' safety, the current text proposes to lay down a classification for AI systems with different requirements and obligations tailored on a "risk-based approach". Some AI systems presenting 'unacceptable' risks would be prohibited; these include AI systems that deploy harmful manipulative techniques or exploit specific vulnerable groups. A wide range of "high-risk" AI systems would be authorised (such as systems for biometric identification or the management of critical infrastructure), but subject to a set of requirements and obligations to gain access to the EU market. Those AI systems presenting only "limited risk" (including chatbots or system that manipulate or generate image, audio, or video content) would be subject to very light transparency obligations. Finally, there would be no obligations for AI systems presenting low or minimal risk; however, providers of such AI systems would be encouraged to comply with the relevant codes of conduct [39].

4.2.4.3.3 Key Takeaways for PISTIS

Following recent PISTIS participants' discussions, it appears unlikely that PISTIS platform will deploy any AI systems in achieving the facilitating of participants' data transfers. Therefore, it is unlikely that the proposed AI Act would directly apply to the project. If AI system is implemented it would most likely support the data transfers; as such, it would unlikely pose any unacceptable or high risks. As a result, it is envisaged that in such an unlikely scenario, certain transparency obligations would have to be complied with. However, these considerations should not have any significant impact on the development of PISTIS platform at the current stage.

It is worth noting that the above assessment has been made solely in relation to PISTIS platform and its envisaged means of operation. In reaching and presenting the opinion, no consideration has been taken in relation to individual participants, or the possible application of AI Act to their AI systems – such further assessment falls outside the scope of this deliverable.

4.3 Overview of the Legal Issues Related to Blockchain and Smart Contracts

4.3.1 EU Policies Towards Blockchain Legal and Regulatory Framework

Although there are many blockchain-related some policy and regulatory initiatives at EU level, we have seen so far, a few legal instruments governing the use of blockchain technology in finance. The European Commission have adopted a comprehensive package of legislative proposals for the regulation of crypto assets in order to increase investments and ensure consumer and investor protection. The proposal on the Markets in Crypto-Assets Regulation, which is explained in detail in Chapter 1.2.3.3. is the latest example of the European regulatory development.

Another important initiative in EU is the launch of the European Regulatory Sandbox for Blockchain [40]. The objective of the sandbox is to provide legal certainty for different use cases of blockchain technology by identifying obstacles to their deployment from a legal and regulatory perspective and providing legal advice, regulatory experience, and guidance in a safe and confidential environment. The sandbox currently accepts applications from EU based companies with a validated proof of concept for a blockchain/DLT application. Although the first selection process has ended in April 2023, the sandbox will run until 2026 and the PISTIS consortium may consider vetting its technology on this sandbox initiative.

4.3.2 Legal Qualification and Role of Smart Contracts in Data Sharing

Firstly, it is important to consider the non-traditional use of the term ‘contract’. A smart contract is merely an ‘if/then’-system on a blockchain. Whether this indeed constitutes a legal contract is another matter. Since EU contract law is not harmonized, one must look at the most contentious rules on contract law across the member states. The two most important elements of contracts that must be assessed per Member State are the exchange of offer and acceptance and the form requirements. The former requires intention to create a engage in a contract, which can be troublesome if the code of the smart contract is not understood properly. In the context of smart contracts, the concepts of “offer” and “acceptance” can be considered as deployment of the code and the call. Once the smart contract is deployed on the blockchain by one end-user, it could be seen as offer. Once the other party sends an electronic message, signed with its own key, she would be accepting the offer and the contract would be considered as concluded [41].However, it is also argued that the deployment of the smart contract as an invitation to engage into a contract, and not as a formal offer that triggers the other party’s acceptance. Furthermore, due to its immutable character, smart contracts do not allow accepting party to negotiate the terms of a contract beforehand and thus preventing contracting party to meet on a common ground. Perhaps, the lack of “mutuality” in smart contracts deprives them of legal character [42].Therefore, the question exists whether deploying a smart contract equals an offer and calling one equals an acceptance which will be subject to the application and the interpretation of each Member State law.

As for data sharing purposes, smart contracts are seen as an interoperability tool which enables individuals and businesses to promote data sharing in EU in the Data Act. Data Act defines smart contracts as computer programs stored in an electronic ledger system wherein the outcome of the execution of the program is recorded on the electronic ledger which have the potential to provide data holders and data recipients with guarantees that conditions for sharing data are respected [43].

However, the Data Act requires smart contracts to be supported and covered by separate data sharing agreements.

Article 30 of the Data Act [44] lays down the essential requirements regarding smart contract to be used for data sharing. These requirements should be considered by the operator of data marketplace, and it will be required for the operator to perform a conformity assessment concerning these requirements stated below, after which it will be entitled to issue an EU declaration of conformity.

“The vendor of an application using smart contracts or, in the absence thereof, the person whose trade, business or profession involves the deployment of smart contracts for others in the context of an agreement to make data available shall comply with the following essential requirements:

a. robustness: ensure that the smart contract has been designed to offer a very high degree of robustness to avoid functional errors and to withstand manipulation by third parties.
b. safe termination and interruption: ensure that a mechanism exists to terminate the continued execution of transactions: the smart contract shall include internal functions which can reset or instruct the contract to stop or interrupt the operation to avoid future (accidental) executions.
c. data archiving and continuity: foresee, if a smart contract must be terminated or deactivated, a possibility to archive transactional data, the smart contract logic and code to keep the record of the operations performed on the data in the past (auditability); and
d. access control: a smart contract shall be protected through rigorous access control mechanisms at the governance and smart contract layers” [44]

4.4 The Guiding Ethical Principles for PISTIS

There are several principles and guidelines regarding “ethical” or “trustworthy” data governance that have been developed in the last five years by international and European authorities and independent institutions. Some of these frameworks have already been adopted by public and private organisation across the world, as well as formed the basis for additional and ongoing policy developments. In the context of data marketplace aiming to facilitate secure peer-to-peer data sharing in the future European data economy, the fundamental values enshrined in the European Declaration on Digital Rights and Principles for the Digital Decade, in the 2018 Report on Towards A Digital Ethics by Ethics Advisory Board of the EDPS and the principles set forth in Rulebook for a Fair Data Economy by SITRA [45], the EC’s Guidance on Ethics By Design and Ethics of Use Approaches for Artificial Intelligence [46] and the EC’s Guidance on Sharing Private Sector Data in the European Data Economy [47] are found to be highly relevant to PISTIS and have been carefully reviewed as part of the work carried out under Task T1.3 – GDPR, Cross-Border Legal Aspects and Contracts Definition for Data Sharing and Trading.

In light of the aforementioned documents, this chapter looks at certain fundamental values and define the following guiding ethical principles for PISTIS.

Accountability: Accountability is one of the key principles that flows from ethics. In data-driven ecosystems, accountability means that the party involved in the development or operation of data processing activities takes responsibility for the way that these processes function and for the resulting consequences. To be held to account, the party should be able to exercise effective oversight and control over data processing operations. Therefore, it is suggested for the project to establish a mechanism that facilitates the system's auditability and ensures traceability and logging of the automated decision-making processes and outcomes. Furthermore, carrying out a risk assessment which considers various stakeholders that are directly and indirectly affected by the project is always recommended. Furthermore, in line with the aforementioned frameworks, accountable data use in PISTIS should provide for that i) data users comply with all applicable policy, legislative, and regulatory requirements by design; ii) clear and common data management rules are in place to support the fair and trustworthy access, sharing and use of data; iii) data governance structures are available to provide advice, intervene or correct actions; and iv) training and education are provided to help develop accountability practices.

Privacy and Protection of Personal Data: The rights to privacy and data protection are fundamental rights which must be always respected, and they must be safeguarded by data governance models that ensure data accuracy and representativeness; personal data must be processed on a fair and lawful basis and enable humans to actively manage their personal data and the way the system uses it. Furthermore, a data marketplace must provide for an appropriate level of security of the data including protection against unauthorized or unlawful processing and against accidental loss, destruction, or damage, using appropriate technical or organizational measures including information security policies and data breach notification procedures.

Transparency: The data network is based on co-operation and respect for information sources. Transparency is important to develop trust. The data must be shared and traded in a transparent manner. Any information addressed to the end-users must be concise, easily accessible, and easy to understand, and clear and plain language and, additionally, where appropriate, visualisation is used. The participants of a data marketplace should know (when/if possible), what data are offered, who will have access to that data, what is the purpose for using such data and by what requirements to promote transparency of network. However, transparency does not mean that all information about data sharing and trading activities should be open to everybody without restriction.

In the context of AI systems, 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 operated, as stakeholders must be able to understand the main concepts behind it such as how, and for what purpose, these systems operate and come to their decisions.

Fairness, Justice, and Equality: All actors in the data network should promote fairness, justice, and equality among individuals. Fairness means that everyone is treated with respect regardless of their socio-economical background or status. Likewise, the benefits (economical and others) must be balanced between all stakeholders in such a manner that individuals that are the source of data are not seen as mere exploitable resources.

No Harm: All actors in the data network should avoid causing harm but instead focus on creating shared value (direct or indirect) for the whole data network and all the people that are affected by the actions of this data network. This means that PISTIS must consider whether data sharing and trading transactions could cause harm or increase the risk of harm. Preventing harm also requires consideration of the nature and all living beings. Data marketplace should make its operations more sustainable and circular, thus reducing its negative externalities on the environment, climate, and natural resources. To comply with this principle, possible end-users, affected individual and communities 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.

Respect for Each Other's Commercial Interest and Ensuring a Level Playing Field: The data sharing contract to be used on PISTIS should include appropriate clauses to protect both the commercial interests and secrets of data holders and data users. Furthermore, contracts should not distort competition by foreclosing access to data marketplace for a particular group and PISTIS should avoid creating a dominant player in the relevant sectors and markets.

Communication: Having appropriate communication mechanism is fundamental to put the above-mentioned ethical principles into play. The platform operator has a special responsibility to articulate, apply, and support the proper implementation of these ethical principles. It will further support PISTIS to promote and establish empathic, honest, and trustworthy relations with the stakeholders and ensure that the project's proposition, its purpose, and legitimate interest that justifies the reason why data collection, access, sharing, or use is needed are clearly communicated with them. Particularly, it is important to engage in social dialogue with relevant actors whose data is being used, or with their representatives, and secondary stakeholders who could be affected by the respective data use. In this engagement, the role of data (e.g., expected benefits and trade-offs), and the primary purpose of its processing is recommended to be communicated in a clear, plain, and understandable language.

It should be noted that the ethical principles explained above are only a start and aim to guide the consortium partners throughout the project development progress. However, as the project progress and in accordance with the policy developments in EU, additional ethical principles may be added.

4.4.1 Special Focus: Data Sovereignty and Monetisation of Personal Data

Monetization of data is one of the main motivations for data holders to share their datasets with third parties. Particularly, in the Business-to-Business data markets context, data monetisation is the first step before the participants interact with data sharing transaction.

However, determining the value of is not an easy task, since analogies with either tangible or intangible assets fail since rules governing pricing of traditional commodities are not quite fit for this new kind of digital resource. Even if it would seem logical to treat personal data as an intangible asset, it is not clear whether it can be measured and valued as a distinct resource. Although it is evident that data processing generates value, a proper understanding of what is the value of data still needs to be clarified. Therefore, challenges revolve around the valuation of data, e.g., who and how it is determined, by the producer, by the market demand, or by a broker or third party or whether the value for a specific data asset is universal or depends on the buyer-seller relationship. In fact, all stakeholders providing or receiving personal data may assign them a specific monetary value different from each other [48].

In a study published by the Global Partnership for Sustainable Development Data, it is noted that there are different methods attempting to conceptualize and measure the value of data however each method seems to be grappling in their own way with the implications of data as a new economic asset, and yet there appears to be little consensus on how best to measure its value [49]. Based on this conclusion, the study unpacks the following five diverse approaches that have been used to measure the value of data.

- Cost-based approach: The value of data is determined based on the full cost to produce that data and statistics either through pure cash recording annually or the full accruals method where costs are allocated as much as possible to the years in which they add to the production of statistics so that major expenditures (e.g., census) are smoothed over time.
- Market-based approach: In this approach, the value of data is determined based on the market price of equivalent products or data users 'willingness to pay for it.
- Income-based approach: According to the income-based approach, the value of data is defined by estimating the productivity improvements and future cash flows that can be derived from the data. This approach has been applied most to assess the fiscal benefits of open government data.
- Benefit monetization: The value of data is estimated by defining the benefits of particular data products, such as a census, and then monetizing the benefits.
- Impact-based approach: In this last approach, the value is determined by assessing the causal effect of data availability on economic and social outcomes, or the costs in terms of inefficiencies or poor policy decisions due to limited or poor-quality data. There are many case studies showing these impacts, from deliberate experiments such as randomized control trials to retrospective assessments of impact.

After reviewing these five approaches the study notes that although the impact-based approach seems to have the most potential to translate investment in datasets into meaningful outcomes for data users, its dependency on context, i.e. context specificity, is a serious drawback limiting its influence. On the other hand, the study notes that the income-based approach have the advantage of providing headline numbers backed up by a methodology. Perhaps, this could be convincing for PISTIS in relation to the initial implementation of data marketplace. Another approach introduced by Short and Todd argues that the value of data is the composite between the value of the asset itself, the value resulting from its use and its expected or future value [50]. Notwithstanding the necessity to understand the monetary value of data, the societal value of data for public good should also be considered.

From data protection perspective, data monetisation is a highly controversial topic for the European data protection regulatory authorities. The EDPB, in its Guidelines 2/2019 on the processing of personal data in the context of online services notes that “one of the main purposes of the GDPR is to provide data subjects with control over information relating to them, personal data cannot be considered as a tradeable commodity” because once control over personal data has been lost, it may not necessarily be regained unlike the other tangible and intangible assets [51].Later, in its statement on Data Governance Act, the EDPB further adds that “data controller (…) is not entitled to 'exchange' or 'trade' personal data (as a so-called “commodity”) in a way that would result as not being in accordance with all applicable data protection principles and rules” [52]. Lastly, in their joint opinion, the EDPB and EDPS followed the same reasoning and states that the prospect of “commoditization” of personal data would not only undermine the very concept of human dignity and the human-centric approach the EU wants to uphold in its Data Strategy, but it would also risk undermining the rights to privacy and data protection as fundamental rights [53]. In this context, it becomes evident that the absence of individuals from the B2B data marketplace where their data are exchanged between businesses may have detrimental impact on individuals and prevent them from having a better understanding about the economic valuation of their personal data.

In light of above, it is recommended for PISTIS first to take into account of alternative valuation approaches and the data protection concerns explained in this chapter when developing a methodology for data valuation assessment and dynamics pricing in WP3. Secondly, the end user of PISTIS should be encouraged to address the asymmetries in data valuation between them and their data subject whose personal data is being shared and take proactive measures. In that regard, PISTIS may ask the end-user to verify that data valuation methods used in PISTIS is explained to the data subject whose personal data is being shared and that the data subjects concerned receive benefit in whatever form, from the data trading activity.

5 DATA MANAGEMENT & TRADING PROCESSES

This section provides an overview and initial analysis of the as-is and to-be processes around data management and data trading. The broad topics that have been identified and are elaborated include data processing, data exchange and transfer, data security and privacy as well as data trading and value exchange. This section, together with the technology radar in section 6, will act as an initial knowledge base that can be used for the detailed state of the art analysis that will be conducted in WP2 and WP3 as well as the definition of the core functionalities of the PISTIS platform.

5.1 DATA PROCESSING

5.1.1 Data interoperability and quality

In the context of PISTIS project, both interoperability and quality are key factors to achieve any results in the project. Data interoperability is widely defined as the capability to transfer data among various systems or functional units without additional effort from any user. A dataset is interpreted as interoperable if it can be stored, retrieved, processed, and transferred among many systems. An interoperable dataset can be merged and used even though they are not in the same format. It has significant importance to organizations because it allows them a defined data access and governance over all their data sources which would help increase the value of those data assets [54] .

Data interoperability can be targeted to achieve in two ways, *planned* interoperability or *maximised* interoperability. *Planned* interoperability is achieved when data is made interoperable with a specific functional unit or system which makes it tightly coupled with those system. With the help of broad and all-purpose data formats and data standards (for example, Dublin Core or schema.org), data can be made interoperable by design with multiple systems, which can be termed as *maximised* interoperability. Current solutions aim at developing standardisation protocols/initiatives increasing data interoperability even for intrinsically heterogeneous unstructured data. For example, the Universal Dependency framework aims at unified and consistent manual annotation of grammar across languages [54] [55].

Data is interoperable at distinct levels. With the help of some high-level transmission protocols (for example, special REST APIs like OAI-PMH, SPARQL, and Linked Data API), systems achieve *foundational* interoperability where data can be transferred among information systems even though the receiving information system cannot interpret the data. *Structural* interoperability is achieved if the received data can be interpreted at the data field level with the help of file and data formats such as CSV, JSON, etc. It ensures structure and syntax to the transferred data. *Semantic* interoperability combines these two types of interoperability and the systems involved in this can transfer, interpret, and use the data. Along with specific data and file formats, semantic technologies embed machine-readable semantic elements from specific vocabularies in some data formats (for example, RDF/XML, Turtle) [54] [55].

Machines interpret data with the help of metadata. It is important for organizations to achieve interoperability in metadata along with achieving them for the actual data. In most cases, the data and metadata can be considered as a key-value pair where the key is a metadata element that provides meaning to the value which is the data. In such cases, interoperability of data can be achieved through the interoperability of metadata [54].

Interoperability is linked to quality. Data quality can be regarded as the ability of data to serve its purpose – seen as the needs of an organisation in terms of operations, planning and decision-making. "A Data Quality Dimension is a recognised term used by data management professionals to describe a property of data that can be measured or assessed against defined standards in order to determine the quality of data". Dimensions focus on measuring and communicating the quality of data, as opposed to describing what data represents. Data quality can be measured in different ways, including: accuracy of the data (how well does the data represent what it is intended to be represented), consistency of the data (lack of contradictions or error values), completeness (if the data provides all the information), exclusiveness (if the data is not widely available), biases (the lack of biases on the dataset and during the creation of the dataset provides quality to the data), etc.

In the context of PISTIS, some use cases are to be considered. Each use case might have their own interoperability standards and their own metrics to evaluate the data quality. For instance, data interoperability plays a critical role in meteorological applications, such as the assimilation of different data sources into Numerical Weather Prediction (NWP) models. NWP models are complex computer simulations that use mathematical equations to predict future weather conditions. Assimilation involves integrating real-time observational data into these models to improve their accuracy and reliability.

To do this, we collect data from a variety of sources, including weather stations, satellites, radar systems, lightning systems, and ground-based instruments. These sources produce data in different formats and at different spatial and temporal resolutions. Data interoperability ensures that data from these different sources can be collected, standardised and seamlessly.

5.1.2 Data lineage

Information passing through systems and subsystems of a company can be viewed as a fluid stream, which on its way continues to merge with other streams until its origin becomes unrecognizable. Throughout its lifespan, a company´s data is being continuously exposed to various sources of consumption and manipulation, ultimately leading to a blurry, if not incomprehensible lineage. Who introduced which changes and how the data´s values were prior to these changes may become untraceable [56].

Meanwhile, the growing importance of data security leads to a bigger number of increasingly complex regulatory *compliance* requirements to be fulfilled. Thus, related auditing procedures can cause time consuming and expensive overhead. In other words: Companies, especially data driven ones, are expected to have a clear overview of their data and the changes it is subjected to [57].

Apart from pure compliance and regulatory reasoning, uncovering the "black box" of data flows and visualizing data lineages increases *transparency* to business users, while improving *collaboration* and understanding between reporting specialists, business analysts and data specialists. This is especially relevant to financial institutions [58].

In addition, having more control over data transformation- and reporting processes increases *data quality* [58], reduces efforts related to error resolving and thereby saves time and money [56].

These advantages can be achieved by following a *Lineage Tracking Approach.* The foundation for this is a detailed documentation process regarding the following aspects:

- *Who* modified the data?
- *What* was modified?
- *When* was it modified? [56]

On top of collecting and managing the information flow related metadata, it can further be utilized to provide more high-level functionalities, such as:

- Visualizing the data flow
- Inferring reasoning for data changes
- Detecting data anomalies [56] [58],

As with many fields, there are also various use cases for Artificial Intelligence to support Lineage Tracking, e.g., by formulating the *Data Lineage* Problem as a matter of input, transformation, and output, teaching the algorithm to answer questions about and fill gaps within the data´s transformation history [57].

- All in all, data driven businesses, especially those, who work with sensitive data, should for their own benefit have control over their data flows and be able to provide transparency for them. It facilitates staying compliant in an increasingly complex data-driven world and provides clarity as well as accessibility for internal information streams.

5.2 Data Exchange & Transfer

With the evolution of distributed computing, the sharing of growing amount of data generated in heterogeneous sources became a necessity, driving the development of many data sharing approaches, both centralised and decentralised. Many organisations are using centralised data sharing platforms, resulting in poor data interaction between data owners and data consumers, also increasing the risk of data and privacy leakage during attacks due to adopting a single point-of-failure (SPoF) approach. Notably such risks and concerns also influence the enthusiasm of data owners to share their data, bringing more obstacles to the sharing of data resources and value extraction. More details on how PISTIS handles these risks (e.g., privacy risks and assurance of data integrity) will also be documented in D5.3.

Existing owners sharing models, utilise data security mechanisms (though attribute-based access control (ABAC) or role-based access control (RBAC) partly addressing on their own the data sharing risks, as these typically lack in terms of data transactions' transparency, data are managed/ stored in single third-party platforms and data analysis or usage tracking is out of the data owners control, thus still not addressing the needs for transparency, trust, control and value. Decentralised distributed methods rely on multiple connected platforms and utilise the blockchain technology (characterised by distribution, security, and reliability) along with smart contracts and are enabling user authentication/authorization through configurable access policies in a fully distributed environment, offering an immutable ledger technology that makes it undisputable against any reliability issues in a trustworthy way. This will assist on to support different sharing profiles aligned with privacy and authorization requirements. However, the transparency nature on most of the blockchain platforms (such as Ethereum) has limited the potential developing blockchain-based applications on permissioned and private environments [59].

The subsections below provide the state of the art on the core pillars focusing on the secure and privacy-preserving data sharing functionality which is related to identify management and access control mechanisms; the Blockchain in tandem with the user wallet supporting the data transactions; and finally, the secure data trading and the requirements of empowering the user to have control of his/her data.

5.2.1 Identity Management and Access Control

Identity management covers all aspects from authentication and authorization (following the concept of Self-Sovereign Identity (SSI) with the issuance and use of Verifiable Credentials (VC) as part of a user's wallet) to Role-based Access Control (RBAC) and Attribute-based (ABAC).

In a nutshell, the SSI approaches are focused on the user’s control over their data. The most prominent SSI approaches only store non-sensitive data on public blockchains and build on components such as Decentralized Identifiers (DIDs), Verifiable Credentials (VCs) with Zero Knowledge Proofs that allow for highly privacy-friendly solutions [60]. More specifically, DIDs are JSON files, stored on-chain, to allow anybody to access the information ensuring that the information is tamper proof, while the actual certificate files are stored off-chain. Blockchain enables SSI to achieve the highest degree of security and scalability required. On top of that, smart contracts can be used to manage the access policies and privacy settings. The first step to having an SSI solution is to provide a trustworthy signing issuer that issues VCs, a digital file that contains one or more credentials about a person from another source, authenticated by the verifier. In addition, in the case of SSI, a digital wallet enables private repositories of users to secure information such as keys, identities, and credentials and ensuring that only authorized individuals have access to the wallet [61].

Regarding access control, currently in the literature many commonly used access control models have been defined such as the Mandatory Access Control (MAC), the Discretionary Access Control (DAC), the RBAC and the ABAC). Static access control models such as MAC, DAC and RBAC usually provide a list of permissions that each subject has on certain objects. ABAC is by nature dynamic and there are not static lists of permissions that associate subjects with objects, but instead there are “snapshots” of such associations that can be generated and dynamically change based on the current context. On top of that, ABAC can work in tandem with the eIDAS infrastructure for the attribute/credential issuing and verification.

5.2.2 Distributed Ledger Technologies

The overall goal of PISTIS is the provision of a secure and auditable data exchange (e.g., trading actions) and transfer environment based on the use of Distributed Ledger Technologies (DLT) and smart contracts to capture data sharing, while complying with the prevailing GDPR legislation. Blockchain, as the most widely adopted DLT mechanism, will be used as the underlying medium for supporting data trading actions while achieving the necessary security and privacy requirements. Blockchain technology is supported and secured by cryptographic primitives and protocols that guarantee integrity, authenticity, and non-repudiated stored data [62]. Three main blockchain types exist: (i) permissioned (e.g., private or consortium), (ii) permissionless (e.g., public) and (iii) hybrid. A permissioned network is a private blockchain network with only authorized members (one member or group of members), whereas a permissionless network is a public blockchain network setup intended to allow public participation. In a private blockchain network, accessibility is controlled by the owner or any organization who can limit the data access to their users by utilising an access control mechanism (e.g., ABAC or RBAC). In general, the permissioned blockchain can offer high efficiency and fast transaction speeds, indicating that it can manage enough network operations in a limited amount of time, while the public blockchain can offer enhanced security due to its decentralized nature and the active participation of users. Today, a plethora of blockchain platforms is available, each designed for a different real-world use-case. In PISTIS, we will focus on open-source platforms that support the deployment and execution of smart contracts. Below we provide a list with some known blockchain platforms:

- Quorum[3] is an open source blockchain, created by JP Morgan and is built on top of Ethereum[4] but offers more efficiency for permissioned approaches, which is a handy solution for managing private transactions. Due to its use of multiple algorithms and a vote-based system, Quorum can process hundreds of transactions per second. Also, Quorum offers private and public transactions to maintain the confidentiality of data.

- MultiChain[5] is an open source blockchain, created by Coin Sciences and is seen as a fork of Bitcoin. MultiChain uses its special round robin validation as a consensus algorithm, where the blockchain nodes are pseudo-randomly selected to create blocks, but a node must wait several block-creation cycles to add another new block.

- Corda[6] is an open source blockchain, created by R3 and designed to operate better privacy level on transaction. Corda enables different business entities to transact without incurring expensive transactional costs. It is designed for managing private transactions by creating business network membership allowed to access their data.

- HyperLedger Fabric[7] is an open source blockchain, created by IBM and further developed by the Hyperledger Foundation. Fabric supports pluggable consensus algorithms (e.g., PBFT, Raft, and Kafka), allowing one to choose the consensus protocol that best suits their needs. Additionally, it offers privacy for communication via the design of channels and a membership mechanism to restrict channel access.

[3] https://consensys.net/quorum/
[4] https://ethereum.org/
[5] https://www.multichain.com/
[6] https://www.r3.com/corda-platform/
[7] https://www.hyperledger.org/use/fabric

- HyperLedger Besu[8] is an advanced Ethereum client that facilitates confining a permissioned and privacy enabled Ethereum network using the privacy group feature. It is more suitable to develop applications that require security, high-performance in private transaction processing since it is scalable, reliable, and offers secured off-chain privacy [63]. It supports both Proof of Work (PoW) and Proof of Authority (PoA) consensus protocols.

The previously introduced Blockchain platforms have respective pros and cons. Quorum and MultiChain use private transactions to hide the data so that a limited number of users can access them. In Corda transactions can be seen by a limited set of users rather than the entire network, while the credentials are kept secret to hide transaction identities. However, this may bring some risks for the transactions since they cannot be verified by network nodes. Fabric enables one to configure privacy via channels and private data access. Finally, Besu utilises the privacy group feature and offers off-chain privacy. Hyperledger Besu is one of the best option due to the user-friendly configuration, the privacy group feature, and the precise documentation to set up the server environment. On top of that, it is compatible with both public and private chains leading to broader and diverse set of applications and use cases using also the Ethereum MainNet [63]. Considering these, we believe that Hyperledger Besu as the best option to be adopted in PISTIS project. However, the current versions of Hyperledger Besu fail to provide necessary features for key management privacy [63]. Thus, a secure and scalable way to support key management for data sharing is necessary. To overcome this issue the prominent solution of OpenDSU and its inherent way of managing keys and SSIs, overcomes the underline Blockchain's weaknesses. The next subsection highlights and provides more details on this noteworthy solution.

5.2.3 Data Sharing Units

OpenDSU provides a data-centric approach to decentralization that can be integrated with different DTLs. In the specifications of OpenDSU[9], data (shared among blockchains) is viewed as individual Data Sharing Units. OpenDSU acts as a layer on top of the underlying Blockchain, providing additional features and functionalities to help developers build dApps. More specifically, OpenDSU provides a set of tools and services that can be used to create, store, and manage data on top of the blockchain, including a decentralized data storage system, access control mechanisms, and secure communication channels.

Anchoring in OpenDSU refers to the process of storing a hash of a piece of data on a blockchain. In essence, it provides a way to ensure the integrity and authenticity of the stored data. It must be noted here, that with anchoring only the metadata (e.g., hashes) are stored on the blockchain, keeping sensitive or privacy related information off the chain, in compliance with GDPR. In parallel, the purpose of the KeySSI concept is to provide blockchain-anchored identities. Both DIDs and KeySSIs are two different methods of creating SSIs. However, the concept of KeySSI is more general than the concept proposed by the W3C DID web standard. On top of that, an important innovation introduced by KeySSI is that they are used as symmetric encryption/decryption keys for DSUs (or for parts of the DSUs). KeySSI is also a self-certifying SSI, thus when a DSU is obtained from the resolver, it is possible to verify

[8] https://www.hyperledger.org/use/besu
[9] https://opendsu.com/rfc001

the authenticity of the DSU, considering it is digitally signed by the owner. Another important features of KeySSI are that it supports derivation of other lower-level KeySSIs as a form of access delegation and is compatible with the DIDs, since it is inspired from it.

OpenDSU empowers the user to have the control of his/her data in alignment with the GDPR. In the context of PISTIS, we consider OpenDSU as a core technology to be adopted for managing SSI digital identities to be stored on the PISTIS buyer’s wallet, keys and data stored on- and off-chain in seamless and agnostic to the underline blockchain technology way.

5.3 Data Security & Privacy

Organizations typically follow several processes to handle security and privacy issues related to data management and sharing. While specific approaches may vary depending on the industry, regulatory requirements, and organizational structure, the most common processes are listed and briefly described in this section.

Risk Assessment: Organizations conduct risk assessments to identify potential security and privacy risks associated with data management and sharing. This process involves evaluating the types of data collected, stored, and shared, as well as the potential threats and vulnerabilities that could compromise data security or privacy [64]. In more details, risk assessment regard several steps such as:

- Identify Assets and Data: The first step in risk assessment is to identify the assets and data that are relevant to the organization. This includes identifying the types of data collected, stored, and shared, as well as the systems, applications, and infrastructure that support data management and sharing processes.
- Identify Threats and Vulnerabilities: Once the assets and data are identified, the next step is to identify potential threats and vulnerabilities. Threats can include malicious activities such as hacking, data breaches, insider threats, or natural disasters that could compromise data security or privacy.
- Assess Impact: The impact assessment involves evaluating the potential consequences of a security or privacy incident. This includes assessing the financial, legal, reputational, operational, and regulatory impact that could arise from the compromise of data or the failure to protect it adequately.
- Assess Likelihood: The likelihood assessment involves evaluating the probability or likelihood of a threat exploiting a vulnerability and causing harm.
- Risk Analysis: In this step, the identified risks are analyzed by considering both their impact and likelihood. The goal is to prioritize risks based on their potential impact and the likelihood of occurrence.
- Risk Evaluation: The risks identified in the previous step are evaluated to determine the organization's tolerance level for each risk.
- Risk Treatment: Risk treatment involves selecting and implementing appropriate risk mitigation measures to reduce the identified risks to an acceptable level.
- Monitor and Review: Risk assessment is an ongoing process, and it is essential to continuously monitor and review the effectiveness of the implemented risk mitigation measures.

Data Classification: Organizations classify their data based on its sensitivity, value, and regulatory requirements. This classification helps determine the appropriate security controls and privacy measures that need to be applied to each category of data [65]. In more details, the data classification processes regard:

- Identify Data Categories: The first step in data classification is to identify and define the different categories or types of data within the organization.
- Define Classification Criteria: Organizations need to establish criteria for classifying data based on its sensitivity and importance.
- Assign Classification Labels: Once the classification criteria are established, data is labeled or tagged according to its assigned category or level of sensitivity.
- Determine Handling Controls: Each classification label should have associated handling controls and protection measures.
- Implement Technical and Organizational Measures: To enforce the handling controls, organizations need to implement appropriate technical and organizational measures. This can include access control mechanisms, encryption, data loss prevention (DLP) solutions, network segmentation, secure storage solutions, employee training programs, and regular audits to ensure compliance with the defined handling controls.
- Document Data Classification Policies: Organizations document their data classification policies and guidelines in a formal document or policy.
- Regular Review and Updates: Data classification is not a one-time activity but an ongoing process. It is essential to periodically review and update the classification of data as the organization evolves, new types of data are introduced, or regulatory requirements change.

By classifying data, organizations can better understand the sensitivity and value of their data assets, prioritize protection efforts, and apply appropriate security controls and privacy measures. Data classification helps organizations streamline their data management processes, enhance data governance, and ensure compliance with applicable laws and regulations. Organizations may adopt established frameworks such as ISO/IEC 27001, NIST SP 800-60, or industry-specific guidelines to develop their data classification strategies and processes.

Data Security Policies: Organizations establish data security policies that outline guidelines and best practices for protecting data. These policies address areas such as access controls, encryption, authentication mechanisms, data backup, and incident response procedures. Policies should align with relevant legal and regulatory frameworks, industry standards, and organizational requirements [66].

Privacy Policies: Organizations develop privacy policies that detail how personal and sensitive information is collected, used, stored, and shared. These policies typically include information on consent, data retention, data subject rights, and mechanisms for addressing privacy concerns or complaints [67].

Access Control: Access control mechanisms are implemented to ensure that only authorized individuals can access and manipulate data. This involves using techniques such as user authentication, role-based access control (RBAC), and implementing least privilege principles to restrict access to data on a need-to-know basis [68].

Data Encryption: Organizations employ encryption techniques to protect sensitive data both at rest and in transit. Encryption helps safeguard data from unauthorized access even if it is intercepted or compromised [69]. The policy may identify encryption standards, key management practices, and encryption requirements for portable devices, remote access, and communication channels.

Data Sharing Agreements: When sharing data with external entities, organizations establish data sharing agreements or contracts. These agreements define the terms and conditions for data sharing, including the purpose, scope, security requirements, and responsibilities of each party involved [70].

Employee Training and Awareness: Organizations conduct regular training programs and awareness campaigns to educate employees about data security and privacy best practices. This helps ensure that employees understand their roles and responsibilities, are aware of potential risks, and know how to handle data appropriately [71].

Incident Response and Data Breach Management: Organizations establish incident response plans to address data breaches, security incidents, or privacy breaches effectively. These plans outline the steps to be taken in case of a breach, including containment, investigation, notification, and remediation procedures [72].

Compliance Monitoring and Auditing: Organizations implement monitoring and auditing processes to ensure ongoing compliance with security and privacy requirements. This involves regularly assessing the effectiveness of security controls, conducting internal and external audits, and addressing any identified gaps or non-compliance issues [73].

It's important to note that the above processes serve as general guidelines, and organizations tailor their approach to data security and privacy based on their specific needs, applicable laws and regulations.

5.4 DATA TRADING AND VALUE EXCHANGE

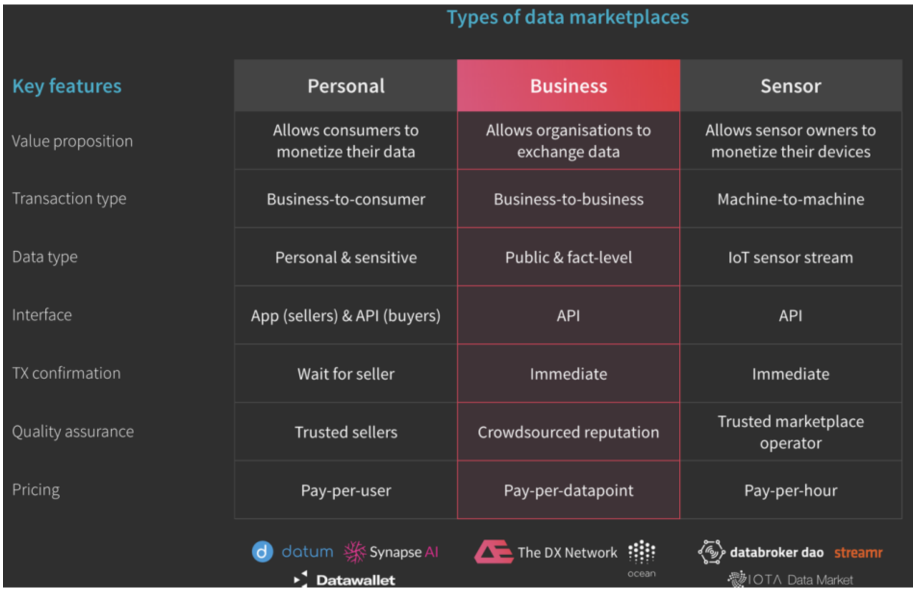
5.4.1 Data markets

Data markets are platforms or marketplaces that facilitate the exchange of data by providing opportunities to individuals, organisations, and businesses for data monetization, data analysis, and research enabling them to search and acquire, buy, and sell various types of data. Currently available data markets focus on different areas depending on the types of data they offer and the specific audience they refer to; however, it shall be noted that the availability and focus of the data marketplaces may change over time as new platforms may emerge.

Data brokers are companies that specialise in gathering, aggregating, and selling user data to their clients or customers; there are more than 4,000 data brokering companies worldwide [74]. Major data brokers are commonly known as ‘information resellers. Data acquisition and monetisation are the two pillars of data brokers industry [75]. Acxiom LLC, which provides data to about 2.5 billion customers, Epsilon Data Management LLC, Oracle America Inc., Equifax Information Services LLC and Experian LLC are among the most profitable companies worldwide [76].

They acquire information from a variety of both online and offline sources. The information obtained by brokers can be related to demographics (age, gender, place of birth and residence etc.), or more sensitive data including financial status and income, criminal record, health or behavioural issues, religious beliefs or even sexuality ([75], [77]). They come in different types and, thus, obtain and utilise data for different purposes; financial information, personal health, marketing and advertising, and people search sites (websites that collect personal information of a user and share it to other users online), are the most prominent types , [77].

Due to the wide range of data stakeholders and the countless potential uses for data, a convenient approach to classify data marketplaces is based on the type of data that participants can exchange. These categories include personal data, business data, and sensor data, as each marketplace caters to specific data types.



Types of data marketplaces

Key features	Personal	Business	Sensor
Value proposition	Allows consumers to monetize their data	Allows organisations to exchange data	Allows sensor owners to monetize their devices
Transaction type	Business-to-consumer	Business-to-business	Machine-to-machine
Data type	Personal & sensitive	Public & fact-level	IoT sensor stream
Interface	App (sellers) & API (buyers)	API	API
TX confirmation	Wait for seller	Immediate	Immediate
Quality assurance	Trusted sellers	Crowdsourced reputation	Trusted marketplace operator
Pricing	Pay-per-user	Pay-per-datapoint	Pay-per-hour
	datum Synapse AI Datawallet	The DX Network ocean	databroker dao streamr IOTA Data Market

Figure 7: Some of the key properties of blockchain-powered data marketplaces categorised by data type[10]

- **Personal Data Marketplaces** are in essence platforms where individuals can profit from their own data by selling it. This data can encompass various aspects, such as location, food preferences, preferred website designs, etc. Individuals can either set a price for their data and await potential buyers or accept incentives like sign-up bonuses or gift cards provided by these marketplaces. These personal data marketplaces fully comply with GDPR regulations, as individuals willingly share their data. Examples include Datum[11], DataWallet[12] , etc.

B2B Data Marketplaces., serve as centralised platforms that gather and store data from various providers, offering it to other organizations. These data marketplaces allow data consumers (i.e., other businesses) to access curated information from multiple sources, which

[10] https://dx.network/
[11] https://datum.org
[12] https://www.datawallet.com

can be utilised for business intelligence purposes. In contrast to the personal data marketplaces, within B2B marketplaces larger datasets are typically shared. Examples include Datarade [13], Ocean Protocol [14], SAP Data Marketplace [15], Bloomberg Enterprise Access Point[16], etc.

- **Sensor/IoT Data Marketplace:** One effective solution for monetising IoT data deriving from various distributed devices, is to sell it to third parties. This is typically accomplished through sensor/IoT data marketplaces which enable organisations to buy or sell real-time data collected from IoT devices. Indeed, data acquired from sensors can assist organisations in understanding consumer behaviour, enhancing sales, and formulating improved marketing strategies. Examples include Dawex[17], Streamr[18], Zenodys[19]

In a more fine-grained approach, data markets can be classified based on the business context of the data they provide. In that sense, such data market classes regard financial data, health data, geospatial data social data and sensor data.

Financial Data Markets specialise in delivering financial and economic data. Assets, liabilities, income, and expenditure, as well as cash flow, are some examples of ‘traditional’ financial data; in addition to that, useful information can be gathered by analysing alternative data such as web data, social media data, credit card transactions etc. [78]. Some indicative notable economic data providers include Bloomberg, Thomson Reuters, Veraset, SafeGraph, Quandl, and stock exchanges like NASDAQ. These systems provide financial institutions, traders, and researchers with historical and real-time market data, corporate financials, economic indicators, and other pertinent information.

Health Data Markets collect and distribute health-related data. Electronic health records (EHRs), medical claims data, health surveys, and data extracted from clinical trials comprise some of the datasets that are frequently used in these industries (Health Sciences Library - the University of Washington, n.d.). Companies such as IQVIA and Practise Fusion have been active in this market.

Geospatial Data Marketplaces specialise in providing geographic and location-based datasets. These sites provide satellite imagery, aerial pictures, digital maps, and other geospatial data. Examples of companies handling such data include ESRI and Mapbox. Geospatial data providers gather, and share information related to demographics, properties and real estate data, speed limits, streets names, ZIP codes, and even more fundamentally, coordinates [79].

IoT Data Marketplaces have formed to ease the exchange of IoT data. These markets enable different businesses to share, monetise and trade data created by connected devices, sensors, and other Internet of Thing’s endpoints [80]. Platforms for organising and exchanging IoT data

[13] https://datarade.ai/

[14] https://oceanprotocol.com/

[15] https://www.sap.com/

[16] https://www.bloomberg.com/professional/product/enterprise-access-point/

[17] https://www.dawex.com

[18] https://streamr.network

[19] https://www.zenodys.com

are available from major technology companies such as Amazon (via Amazon Web Services Marketplace), Microsoft (via Azure), and Bosch (via Bosch IoT Suite).

Social Media Data Markets specialise in collecting and analysing social media data, which are commonly found in the form of posts, likes and reactions, comments, shares, hashtags, followers, shares and reposts, or reviews [81]. Companies such as Cision Inc, Clarabridge Inc, NETBase and Hootsuite Inc, are among the pioneers in this sub-market [82]. Usually, a platform provides access to social media APIs for collecting enormous amounts of social media data. This information can be used for customer sentiment analysis and procced to personalised advertising, while it can also prove helpful in predicting resource planning and possible budget alterations [83].

Open Data Portals: Open data portals focus on making publicly available government-generated datasets which do not bear privacy and confidentiality restrictions, and everyone can (re-)use and distribute [84]. These platforms give academics, developers, and citizens access to a wide range of data from multiple government organisations, allowing them to use the data for analysis, study, and application creation [85].

To provide more content to the reader Table 26 presents a non-exhausting list of popular data markets operating both in the EU and worldwide, along with their description and their area of focus.

Table 26 Data markets

Data Marketplace	Description	Area of Focus
European Data Portal (EDP)[20]	The European Data Portal serves as a central hub for discovering and accessing open datasets from EU member states, promoting data-driven innovation across different sectors.	Open data from various European countries to facilitate data sharing and innovation within the EU.
Advaneo[21]	Advaneo Data Marketplace offers a wide range of open as well as commercial data from all industries, with the aim to create a sustainable portal as a place for innovation. This data marketplace serves as a data catalogue for the metadata provided by the data providers, while it functions as a secure trading portal and offers functions for data processing and administration.	Advaneo Data Marketplace sets particular emphasis on data security, governance and data sovereignty.
European DataWarehouse[22]	The European DataWarehouse collects, validates, and disseminates data related to asset-backed securities and structured finance products, offering transparency and reliable data for the securitization market in the EU.	Focuses on the securitization market and structured finance data, providing transparency and standardized reporting for investors, regulators, and market participants in the EU.

[20] https://data.europa.eu/en

[21] https://www.advaneo.de/

[22] https://eurodw.eu/

Amazon Web Services (AWS) Data Exchange[23]	AWS Data Exchange is a data marketplace provided by Amazon Web Services. It allows data providers to publish and sell their datasets, while data consumers can browse and purchase these datasets.	The marketplace focuses on a wide range of industries, including finance, healthcare, weather, and more.
Microsoft Azure Data Share[24]	Microsoft Azure Data Share enables businesses and organizations to securely share data with other businesses, customers, and partners in just a few clicks.	Within Azure Data Share, data owners can control who has access to their data, duration of access, and data terms of use; focusing on full visibility into data sharing relationships with a user-friendly interface.
Dataverse[25]	Dataverse is an open-source web application that enables researchers to publish, share, and find datasets.	It focuses on promoting data sharing in the academic and research community.
OpenDataSoft[26]	OpenDataSoft enables organisations to publish, share, and analyse their data, fostering the exchange of data and promoting transparency and collaboration in the EU.	Data marketplace platform supporting European countries, empowering governments, businesses, and non-profit organizations to create data portals and open data initiatives.
Oracle Data Marketplace[27]	It's a platform that provides access to quality data to target audiences at any stage of the purchase funnel.	It focuses on setting the standard for open and transparent audience data trading, with over 30,000 data attributes to power branding or direct marketing initiatives.
Snowflake Data Marketplace[28]	It is a platform in the Cloud Data Warehouse area connecting data-driven business leaders to over 360 providers, offering more than 1,700 live, ready-to-query data sets, data services, and applications.	This platform enables users explore, evaluate and purchase the data, data services, and applications needed to innovate their business, while eliminating the costs and delays associated with traditional ETL processes and integration.

5.4.2 Data valuation methods & processes

In today's data-driven world, organisations are increasingly recognising the value of their data as a strategic asset that dan provide additional value on top of their business operations. Data valuation and data monetisation have emerged as key practices to unlock the untapped potential of data assets. In this direction, the following section aims to describe and provide insights into the various processes/ methodologies (such as data inventory, analytics, market assessment, and risk evaluation) followed by organisations willing to perform data valuation and data monetization towards determining an appropriate target value for their data.

[23] https://aws.amazon.com/data-exchange/
[24] https://azure.microsoft.com/
[25] https://dataverse.org/
[26] https://www.opendatasoft.com/en/
[27] https://www.oracle.com/marketingcloud/products/data-management-platform/
[28] https://www.snowflake.com/en/data-cloud/marketplace/

Overall, data valuation refers to the process of assigning a financial worth to an organisation's data assets (being datasets, models, algorithms, etc.). On the other hand, data monetisation, involves leveraging the value of data to generate economic returns. Data monetization can be generally seen as "a way of companies increasing their revenue by selling their intangible data assets" [86]. Also, data monetization has been defined as "the act of exchanging information-based products and services for legal tender or something of perceived equivalent value" [87].Both these practices are important for organisations which are seeking to harness their data's potential towards gaining competitive advantage, informed decision-making, and generate new revenue streams [88].

The first step in data valuation is to conduct a comprehensive data inventory, which involves identifying and categorising different types of data assets available within the organization, such as operational data, transactional data, customer data, etc. This data inventory should capture essential details such as data volume, quality, uniqueness, and relevance to the organisation's business objectives [89]. All this information will ultimately serve as the foundation for subsequent valuation and monetization efforts.

Once the data inventory is established, organisations can leverage the potentials of data analytics, to extract valuable insights from their data; by uncovering patterns, trends, and correlations that can enhance decision-making, improve operational efficiency, and drive innovation [90]. Common data analytics techniques that can be used to generate actionable insights include, analysis of historical data, predictive modelling, and use of machine learning (ML), Deep Learning (DL) algorithms [91].

Conducting a market assessment is crucial to determine the value of data in the context of the broader marketplace; this involves understanding the demand for specific data types, identifying potential buyers or partners, and assessing market dynamics, including pricing models, and emerging trends [92]. Organisations may also explore partnerships, data exchanges, or participate/engage into data marketplaces to facilitate data monetisation and access new revenue streams [93].

The process of data valuation also entails evaluating the risks associated with data assets, by assessing legal and regulatory compliance, privacy and security risks, and reputational risks. It is imperative for organisations to implement robust data governance frameworks, data protection measures, and ensure compliance with applicable laws and regulations, such as GDPR. Evaluating risks helps mitigate potential liabilities and increases the overall value of the data [94].

Based on the data inventory, analytics insights, market assessment, and risk evaluation processes, organisations can finally articulate a target value for their data. This value represents the estimated monetary worth or potential return on investment that the data can generate. There are different valuation methods available that may be employed, such as market-based approaches (comparable sales or licensing transactions), cost-based approaches (replacement or reproduction costs), or income-based approaches (discounted cash flows or revenue multiples) [95]. To determine an appropriate target value for their data, organisations should consider a combination of quantitative and qualitative factors, such as the data's uniqueness, relevance to the market, the competitive advantage it offers, potential revenue streams, and associated risks. It is important to regularly reassess and update the valuation as market conditions, data quality, and regulatory landscapes continuously evolve.

According to [96], while there is still no standard way to assign value to data, nonetheless they identify three main categories for classifying data valuation models: *market-based models*, which see data value as cost and revenue; *economic models*, which target financial and public benefit; *dimensional models*, where data value is based on different dimensions.

Considering *information capacity* as one of those dimensions, it can be defined as “the current stock of understandings informed by a given installed base” [97] representing “the potential of a digital information asset that can be defined and evaluated independently from the usage.” The information capacity of an organization, on the one hand, determines “the economic utility of a digital information asset”; on the other hand, it enables capabilities that are” the possibility and/or right of the user or a user community to perform a set of actions on a computational object or process “.

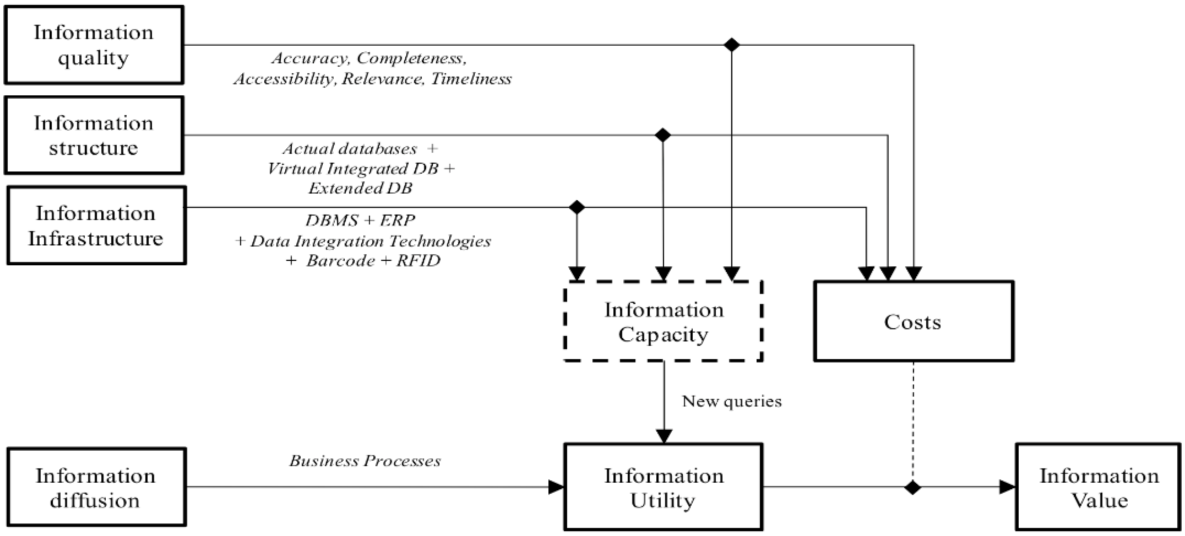

Information quality
Accuracy, Completeness, Accessibility, Relevance, Timeliness
Information structure
Actual databases + Virtual Integrated DB + Extended DB
Information Infrastructure
DBMS + ERP + Data Integration Technologies + Barcode + RFID
Information Capacity
Costs
New queries
Information diffusion
Business Processes
Information Utility
Information Value

Figure 8: A model for the assessment of information value, adapted from Batini et al. (2018)

As shown in the model of Figure 8 proposed in [98], the information capacity of an organization is made up of different dimensions that point out to crucial areas like *data quality*, the degree of *integration of data*, the available *infrastructure* (Data Base Management Systems, Enterprise Systems, Data Integration technologies, Information and Communication technologies like Barcode, RFID, etc.). Those dimensions also represent the *costs* that should be considered for the data valuation together with the *information utility* coming from the diffusion on the organization process of the information associated with the data. As to the data, they must have relevant attributes such as *timeliness*, *contents*, and *format* [99].

A way to understand the information utility for organizations' processes is to identify the core activities they are part of. To this end, the state-of-the-art literature has modeled them in terms of the *data value chain* [100]. Besides that model authors in [100] have also identified in that literature a *linked data value chain* (LDVC) and a *data value chain network* model, then proposing a *big data value chain* (BDVC) on which business models for data monetization can be developed around four main axes: “Data extracted from customers’ activities which could be in its raw format"; "Data providers that collect and sale primary and secondary data"; "Data aggregators that provide customers with aggregated services"; "Technical platforms, based on infrastructure, analysis, computing and cloud capabilities that enable to process, consume and share data" [100]. Further considering the state-of-the-art literature on data-value chains, Lim et al. (2018) propose a data–value chain as a nine-factor framework for data-based value creation in information-intensive services, whose spectrum goes from the *data source* through *data collection* and *data analysis* to the *use* made by the final user. Once organizations have assessed the value of their data according to the models mentioned above and dimensions, a data monetization strategy can benefit from access to *data markets* because they provide access to a network of prospective buyer [101].

As to the issue of creating and capturing value from data assets, particularly when they are big data, [102] used the resource-based view (RBV) theory [103] to frame their impact on firms' performance when they can't be "easily imitated or substituted by competitors" and to claim that "value creation and capture arising from big data depends on the benefits being greater than the costs of collecting, storing and using this resource" [104]. To this end, they considered in an empirical study of mobile device applications, the "amount of data collected" or the *volume* dimension of big data, measured by the "number of mobile device applications downloaded", together with the "assortment of data collected per observation", or the *variety* dimension of big data, measured by "number of types of data collected per application", and the "reliability and insightfulness of data" or the *veracity* dimension of big data, measured by "percentage of employees devoted to big data analysis" [104]. The findings of their study show that value creation is enabled by the coupling of big data volume and variety, which may "produce benefits that outweigh the costs, and positively affect firm performance". At the same time, veracity allows "firms to capture value by developing insights, from the volume and variety of data" [104].

Table 27 Data monetization options, adapted and elaborated from Parvinen et al. (2020).

Data monetization options	Target use	Source
Selling data	External	(Najjar and Kettinger, 2013; Wixom and Ross, 2018; Baecker *et al.*, 2020)
Charging (*"Data Packaging"*)	External	(Najjar and Kettinger, 2013; Baecker *et al.*, 2020)
Data-as-a-service	External	(Baecker *et al.*, 2020)
Creating new offerings through products and service innovation ("*Wrapping Information Around Products*")	External	(Wixom, 2014; Wixom and Ross, 2018; Baecker *et al.*, 2020)

Data Privacy and Control Guarantee	External	(Baecker *et al.*, 2020)
Improving existing products, and services (*"Bartering"*)	Internal/External	(Wixom, 2014; Wixom and Ross, 2018; Baecker *et al.*, 2020)
Optimization of operations and business process improvement (Costs reduction)	Internal	(Najjar and Kettinger, 2013; Baecker *et al.*, 2020)

Finally, authors in [105], through an analysis of the state literature on data monetization, has identified a set of options that can be classified based on their having a target use of data that is internal or external to the focal organization (see an adapted summary in Table 27. Furthermore, through another thematic analysis of the literature, [106]have proposed a data monetization configurational model whose monetization layer comprises three main components: *goods*, *trading model*, *end consumer*.

As seen from above, data valuation and data monetisation are critical practices for organisations, which are seeking to unlock the value of their data assets, and by adopting a holistic approach, considering factors such as data inventory, analytics insights, market assessment, and risk evaluation, organisations can systematically assess and articulate a target value for their data that aligns with their strategic objectives and maximises the potential benefits of their data assets.

6 TECHNOLOGY RADAR

This section contains a comprehensive list of external initiatives, projects, standards and frameworks that are relevant to the core functionalities of the PISTIS project and will serve, together with the data management and trading processes overview of section 5, as an initial knowledge base that can be used for the detailed state of the art analysis that will conducted in the first deliverables of WP2 and WP3 as well as the definition of the core functionalities of the PISTIS platform.

6.1 DATA MANAGEMENT

6.1.1 Initiatives related to ETL

This section is focused on initiatives related to data management, considering the entire pipeline of data operations, from ingestion to transformation, using specialized repositories for storing data models, metadata, data transformations and so on. The section is also including all the initiatives related to data quality management and data lineage tracking, which allows to track how data is used throughout the organization, providing valuable insights into data flow and usage patterns. Any kind of initiative related to analytics/insight/AI-based engine used for automating the data management processes should also be included in this list.

Some of the relevant concepts include Data Check-in, Data Enrichment, Data Transformation, Data Pipeline Storages, CIM Repository, Data Model Repository, Metadata Repository, Data Schemas, Metadata Repository, Data Quality, Data Lineage, Analytics/Insights, among others.

Very briefly we are referring to the following concepts:

Data Check-in: Data check-in refers to the process of validating and verifying data as it enters a system or database. This can include formal processes including:

- Format Check: Verifying that the data is in the expected format, such as CSV, JSON, or XML, and conforms to a predefined structure.
- Integrity Check: Ensuring that the data meets integrity constraints, such as primary key uniqueness, foreign key references, or data type consistency.
- Completeness Check: Verifying that all the required fields or attributes have been provided and are not missing.
- Quality Check: Assessing the quality of data based on predefined criteria, such as accuracy, consistency, validity, and timeliness.
- Duplication Check: Identifying and flagging any duplicate or redundant data entries to maintain data uniqueness and avoid inconsistencies.
- Validation Check: Applying business rules or validation criteria to ensure that the data adheres to specific standards or regulations.

Data Enrichment: Data enrichment involves enhancing or augmenting existing data with additional information to improve its quality, completeness, or usefulness. This process typically involves adding relevant data to the original data including:

- Geolocation Data: Adding geographic coordinates (latitude and longitude) to data records based on addresses or location information. This can enable spatial analysis and mapping.
- Demographic Data: Appending demographic attributes such as age, gender, income, education level, or household size to existing data. This information can provide valuable insights for segmentation and targeting.
- Social Media Data: Integrating social media data, such as user profiles, followers, or engagement metrics, into existing customer or user data. This can help in understanding social media behaviour and preferences.
- Behavioural Data: Incorporating behavioural data, such as browsing history, purchase patterns, or interaction logs, into customer profiles. This can enable personalized marketing or recommendation systems.
- External Data Sources: Integrating data from external sources, such as third-party APIs, public datasets, or industry databases, to supplement existing data. This can provide additional context and enrich the understanding of the data.
- Sentiment Analysis: Analysing textual data, such as customer reviews or social media posts, to determine sentiment scores or emotional tones. This can provide insights into customer satisfaction or brand perception.
- Data Standardization: Standardizing and normalizing data formats, units, or naming conventions to ensure consistency and compatibility across different data sources or systems.
- Data Aggregation: Combining and aggregating data from multiple sources or at different granularities to create consolidated views or summary statistics. This can facilitate reporting and analysis.
- Data Transformation: Data transformation refers to the process of converting data from one format, structure, or representation to another. Data transformation is a critical step in data management and is often performed to prepare data for analysis, reporting, or storage in a particular database or system. It ensures that data is in the appropriate format and structure required for the intended purpose It involves:
- Data Cleaning: Removing or correcting errors, inconsistencies, or inaccuracies in the data. This can involve tasks such as removing duplicates, handling missing values, or standardizing data formats.
- Data Filtering: Selecting and extracting a subset of data based on specific criteria or conditions. This can involve filtering out irrelevant or unwanted data to focus on the relevant subset.
- Data Aggregation: Combining multiple data records or observations into a summary representation. This can involve grouping data based on certain attributes and performing calculations such as sum, average, or count.
- Data Integration: Combining data from different sources or databases into a unified format. This can involve resolving schema or format differences, mapping attributes, and consolidating the data into a single, coherent dataset.

- Data Normalization: Restructuring and organizing data to eliminate redundancy and ensure data consistency. This can involve decomposing data into separate tables, establishing relationships, and applying normalization techniques.
- Data Formatting: Converting data from one format to another, such as converting dates from one date format to another or converting numerical values to a standardized format.
- Data Encoding/Decoding: Converting data between different character encodings or binary formats. This is particularly relevant when dealing with internationalization or compatibility across different systems.
- Data Calculations and Derivations: Performing mathematical, statistical, or logical operations on data to derive new variables or calculate aggregated metrics.
- Data Schema Transformation: Modifying the structure or schema of a database or dataset, such as adding or removing columns, changing data types, or altering relationships between tables.
- Data Serialization: Converting data structures or objects into a serialized format for storage or transmission, such as converting data to JSON, XML, or binary formats.

Data Pipeline Storages: Data pipeline storages refer to the repositories or storage systems used to store data during different stages of a data pipeline. A data pipeline is a sequence of processes that extracts, transforms, and loads data from various sources to a target destination, such as a data warehouse or analytics platform. The storage systems can include databases, data lakes, or cloud storage solutions.

CIM Repository: CIM stands for Common Information Model. A CIM repository is a centralized repository that stores and manages the information models based on the CIM standard. CIM is an industry standard for modelling and managing data in utility networks, such as electric power or telecommunications. The CIM repository facilitates interoperability and standardization of data models within the utility industry.

Data Model Repository: A data model repository is a centralized location where data models are stored and managed. Data models define the structure, relationships, and constraints of data elements in a database or information system. The repository helps in version control, collaboration, and reuse of data models across different projects or applications.

Metadata Repository: A metadata repository is a centralized storage system that contains metadata, which is data about data. It stores information about the structure, attributes, relationships, and other characteristics of data elements. The metadata repository enables data cataloguing, data discovery, and provides context for understanding and managing data assets.

Data Schemas: A data schema is a logical structure or blueprint that defines how data is organized, stored, and accessed in a database or data system. It specifies the tables, fields, data types, relationships, and constraints that govern the data. Data schemas provide a framework for data integrity and consistency within an organization.

Data Quality: Data quality refers to the reliability, accuracy, completeness, consistency, and timeliness of data. It measures the fitness of data for its intended use and ensures that data meets the required standards and expectations. Data quality management involves processes, tools, and techniques to assess, improve, and maintain data quality throughout its lifecycle.

Data Lineage: Data lineage is the record of the origin, movement, and transformations that data undergoes from its source to its destination. It traces the data's journey across systems, processes, and transformations, providing visibility into how data is derived, aggregated, and used. Data lineage helps in understanding data dependencies, impact analysis, compliance, and troubleshooting.

Analytics/Insights: Analytics or insights refer to the process of extracting meaningful information, patterns, and trends from data. It involves applying statistical, mathematical, or machine learning techniques to uncover insights, make predictions, or support decision-making. Analytics can range from simple descriptive analysis to advanced predictive or prescriptive analysis, depending on the goals and complexity of the data.

Table 28 Initiatives related to ETL

#	INITIATIVE/ PROJECT NAME	SHORT DESCRIPTION AND RELEVANCE FOR PISTIS
1	Apache NiFi [29]	Apache NiFi is an open-source project of Apache Foundation developing the same named software to support powerful and scalable directed graphs of data routing, transformation, and system mediation logic. It is a mature project and software.
2	ASSURED [30]	ASSURED designed and implemented a novel policy-driven, formally verified, runtime assurance framework in the complex domain of Cyber-Physical System (CPS). Attribute-based Encryption and Attribute-based Access Control mechanism were adopted that can also be utilised in PISTIS.
3	DataVaults [31]	DataVaults aims to create a framework and architecture for managing personal data coming from different data sources. This framework also provides the control to share that data, with the possibility of monetizing when sharing with third parties. Additionally, third-party organizations are given the mechanisms to request and access the data shared with them, generating a collaborative data-sharing relationship with individuals. All the data sharing is performed in an ecosystem managed by smart contracts to guarantee protection, ownership, and privacy of the data exchange. This project is highly related with Pistis considering the data Lineage tracking, where a big effort is being put. Additionally, the approach followed in this project regarding data analytics, providing an analytics playground where different processing tasks can be carried out over the owned data in a secured locally isolated solution, can be an example to be extended in PISTIS.

[29] https://nifi.apache.org

[30] https://www.project-assured.eu/

[31] https://www.datavaults.eu/

4	Glass [32]	The GLASS project introduces a citizen-centric e-governance model that enables beneficiaries to participate in a network for big data exchange and service delivery, which is by design digital, efficient, cost-effective, interoperable, cross-border, secure and promotes the once-only priority. OpenDSU were adopted to support a wallet implementation, which is like the PISTIS case.
5	I3-Market [33]	The i3-MARKET project aims to meet the demand for a unified European Data Market Economy by facilitating secure and privacy-preserving data sharing across various data spaces and marketplaces. Its primary goal is to provide technologies that enable trustworthy collaboration and federation of existing and future marketplace platforms, with a particular focus on industrial data. i3-MARKET seeks to provide a solution in form of federating data marketplace by developing technologies and solutions for a trusted, interoperable, and decentralized infrastructure. To that end, a backplane has been developed as the foundation for federation, enabling interoperability among different data spaces and marketplaces. It utilizes trusted, federated, and decentralized software components to integrate various marketplaces and facilitates secure and privacy-preserving data sharing. The strategy for data transformation and enrichment described in i3Market may be applicable to the Pistis project. In particular, the definition, creation, and collection of semantic data models that allow for the sharing of a common description of data assets, as well as the tools that aided in the development and implementation of (meta)data management systems and registries catalogues for managing information and meta data descriptions. On the other side, the i3Market common semantic data model repository may be of interest to PISTIS to facilitate interoperability across application domains, data platforms, and stakeholders.
6	MARVEL [34]	The MARVEL project focuses on the generation and processing of large-scale and diverse datasets in the context of smart city environments. It aims to leverage technologies such as AI, analytics, multimodal perception, software engineering, and high-performance computing to create an Edge-Fog-Cloud Computing Continuum paradigm. This paradigm goes beyond traditional Big Data approaches by integrating distributed resources, heterogeneous data sources, and privacy preservation techniques. The project's goal is to support real-time data-driven applications and decision-making in cities, addressing various societal challenges such as public safety, traffic analysis, and behaviour. MARVEL aligns with the EU Data Economy vision by addressing challenges in the Big Data Value chain, promoting open science and data sharing, investing in research and innovation, and involving citizens to drive breakthrough innovation.

[32] https://www.glass-h2020.eu/

[33] https://www.i3-market.eu/

[34] https://www.marvel-project.eu/

		This project is developing a framework for the data processing based on the use of Apache NiFi that could be of great interest at the design phase for the data workflow to be developed in PISTIS
7	PharmaLedger [35]	The PharmaLedger project delivered a blockchain-based platform for the healthcare sector, using the supply chain, clinical trials, and health data as case studies. Its architecture is based on a minimal use of smart contracts for anchoring off-chain data and code, with emphasis placed on encapsulation of self-sovereign code and data in Data Sharing Units (DSUs). The released PharmaLedger libraries, namely OpenDSU, can be utilised for the implementation of PISTIS wallet and the key management.
8	Piveau Consus [36]	Piveau Consus is a subproject of the Fraunhofer FOKUS open-source project piveau. Piveau Consus provides you a high performant and high scalable solution based on microservices and container technology to fetch data or metadata from a source. It is mature tool used in several open data projects, for example data.europa.eu and in some research projects.
9	PolicyCLoud [37]	The PolicyCloud project aims to enhance policy management by means of using the capabilities of big data, and cloud technologies. To do that, a framework that supports the whole workflow of the data processing from its ingestion to its consumption by the policy makers, by means of several ways to visualize the resultant data has been provided by the project. In relation to Pistis, the data processing carried out in the PolicyCloud project is of interest, as well as the different analytical components developed and integrated in the whole data flow implemented in the different use cases presented in PolicyCloud.
10	TANGO [38]	The TANGO project aims to create a framework for cross-sector data sharing in a citizen-centric, secure, and trustworthy manner. The project will develop a platform that promotes user-friendly, secure, compliant, and environmentally sustainable data management. It will leverage emerging digital technologies to strengthen privacy, reduce costs, and improve productivity. The project's objectives include designing a holistic framework for responsible and green data management, ensuring security, privacy, and data ownership. It will also develop trust management mechanisms, prioritize green data operations, and provide tools to ease the creations of machine learning models to generate an added value to the data managed in the project. This project will develop solutions regarding privacy assessment that would be of interest related to the data lineage tracking area to be addressed in the Pistis project. Regarding data analysis, this project expects to contribute to several areas, including Machine Learning

[35] https://pharmaledger.eu/

[36] https://doc.piveau.de/consus/

[37] https://policycloud.eu/

[38] https://tango-project.eu/

		operations, providing tracking for the predictive model generation process in form of experiments as well as the model trustworthiness that could be of interest in the project.
11	W3C [39]	The World Wide Web Consortium (W3C) is the committee with the agenda to standardize web technologies. Among many web related endeavours, they also dedicated a part of their work to create an ontology for storing provenance information. The PROV-O framework can be utilized to efficiently track data lineages in RDF-format.
12	XMANAI [40]	XMANAI is an EU funded project, which aims at providing explainable AI solutions for businesses in manufacturing. As the involved data is sensitive and subject to frequent manipulation, keeping track of its lineage is a substantial feature of the platform. In this context, a Provenance Engine, which documents who performed which type of operation when and on which dataset in a detailed and efficient manner was developed.
13	OpenLineage [41]	OpenLineage is an open-source lineage tracking project, providing an open standard for meta data- and lineage collection. It is integrable flexibly while offering ist services via various interfaces, depending on the needs of the user.

6.1.2 Initiatives related to Privacy, Security and Trust

This section is focused on data encryption initiatives that are relevant to the data privacy and security of access to data factories and trading platforms, especially with reference to projects in the distributed encryption space, without compromising the privacy of the data and making it easy for users to access encrypted data while maintaining a high level of security. Besides encryption/decryption and public key management, other projects relevant under this section are the ones related to identity and access management, cyberthreat detection and all those initiatives related to user authentication. Additional initiatives related to automation/assessment of data privacy are included within this section (those projects related to GDPR assessment tools, anonymization, and so on).

Table 29 Initiatives related to privacy, security and trust

#	INITIATIVE/ PROJECT NAME	SHORT DESCRIPTION AND RELEVANCE FOR PISTIS
1	DataVaults [42]	DataVaults provides a platform to support Secure, Privacy Preserving Data Sharing. In this context with respect to both Security and Privacy DataVaults provides the following support: Cryptography, data anonymisation, remote attestation and trusted data exchange with TPM technologies between the Personal DataVaults and the DataVaults cloud-based engine. Privacy risk assessment for individuals revealing the true risk exposure factor of

[39] https://www.w3.org/TR/prov-o/

[40] https://ai4manufacturing.eu/project/

[41] https://github.com/OpenLineage

[42] https://www.datavaults.eu/

		individuals based on the shared data. With respect to, PISTIS, the data anonymization technologies from DataVaults are readily transferable and can be adapted based on the specific requirements of the data and use cases. The anonymization techniques available include K-Anonymity, Differential Privacy, The Generation of Synthetic Data with the same statistical properties as the original data (specifically potentially applicable for location data) and the replacement of actual PII with randomized aliases as a mechanism for obfuscating Personally Identifiable Information (PII).
2	DOOR eSSIF-Lab EU Project [43] [44]	DOOR extends the eSSIF-Framework by building a new component on the Holder side that enables the use of hardware-based keys and offers the possibility to bind Verifiable Credentials (VCs) to the wallet of the holder. In this way, it transfers the root of trust of the SSI ecosystem purely on the digital wallet by considering an underlying Trusted Component as part of the wallet, without making any assumptions on the trustworthiness of the other layers. These SSI enhancements can support the eIDAS framework and be adopted in PISTIS.
3	EUDI [45]	The European Digital Identity (EUDI) Wallet Architecture and Reference Framework provides a set of the specifications needed to develop an interoperable European Digital Identity Wallet Solution based on common standards and practices. All the proposed guidelines and practices will be adopted in PISTIS.
4	HELayers [46]	The Open-source IBM Fully Homomorphic Encryption Toolkit. Provides a software development kit (SDK) for the practical and efficient execution of encrypted workloads using fully homomorphic encrypted data. HElayers enables application developers and data scientists to seamlessly use advanced privacy preserving techniques. It can be used, adapted and integrated within PISTIS to support some of the searchable encryption capability requirements.
	EU Project/ PRISMACLOUD[47]	Privacy and Security Maintaining Services in the Cloud: PRISMACLOUD focuses on developing privacy-enhancing technologies for cloud-based services. It includes research on secure identity management, access control mechanisms, and cryptographic solutions. Technologies and approaches followed and generated by the project can be considered for the PISTIS identity and access control management.

[43] https://essif-lab.eu/

[44] https://www.ngi.eu/funded_solution/essi_ioc_41/

[45] https://digital-strategy.ec.europa.eu/en/library/european-digital-identity-wallet-architecture-and-reference-framework

[46] https://github.com/IBM/fhe-toolkit-linux

[47] https://prismacloud.eu/

	EU Project/ LIGHTest [48]	A Federated Testbed for Testing and Validating Long-Term Security: LIGHTest is a project that aims to develop a framework for testing and validating electronic identification and trust services. It focuses on interoperability and trustworthiness of digital identities across different systems. Technologies and approaches followed and generated by the project can be considered for the PISTIS secure data sharing across the various data providers.
	EU Project/ ARIES [49]	Adaptive and Resilient European Identity and Access Management Systems: ARIES aims to enhance identity and access management systems to be adaptive, resilient, and privacy friendly. It focuses on improving the usability and security of identity solutions. Technologies and approaches followed and generated by the project can be considered for the PISTIS identity and access control management.
	EU Project/CREDENTIAL[50]	Secure Cloud Identity Wallet: CREDENTIAL focuses on developing a user-centric identity wallet that enables secure and privacy-enhanced authentication and access control. It aims to provide users with control over their personal data while ensuring strong security. Technologies and approaches followed and generated by the project can be considered for the PISTIS identity and access control management along with data ownership issues.
	Initiative / OpenDP [51]	OpenDP is a community effort to build trustworthy, open-source software tools for statistical analysis of sensitive private data. These tools, which we call OpenDP, will offer the rigorous protections of differential privacy for the individuals who may be represented in confidential data and statistically valid methods of analysis for researchers who study the data. Technologies and approaches followed and generated by the project can be considered for the PISTIS platform with respect to data ownership and processing.
	Open Source Project / ABY Framework [52]	ABY framework combines secure computation schemes based on Arithmetic sharing, Boolean sharing, and Yao's garbled circuits and makes available best-practice solutions in secure two-party computation. It allows to pre-compute almost all cryptographic operations and provides efficient conversions between secure computation schemes based on pre-computed oblivious transfer extensions. Technologies and approaches followed and generated by the project can be considered for the PISTIS platform with respect to data ownership and processing.
	Initiative / Confidential Computing Initiative [53]	The Confidential Computing Consortium is an industry collaboration that aims to promote the adoption of confidential computing technologies, including secure multi-party computation. The

[48] https://www.lightest.eu/

[49] https://aries-project.eu/

[50] https://credential.eu/

[51] https://opendp.org/

[52] https://github.com/encryptogroup/ABY

[53] https://confidentialcomputing.io/

		consortium brings together technology companies, researchers, and developers to develop open standards and tools for secure computation. Technologies and approaches followed and generated by the project can be considered for the PISTIS platform with respect to data ownership and processing.
	Initiative / ZKProof Standards [54]	Several initiatives and organizations, such as the Zero-Knowledge Proof Standardization Workshop, have emerged to foster collaboration and standardization in the field of zero-knowledge proofs. These efforts aim to promote interoperability and security among different zero-knowledge proof systems. Technologies and approaches followed and generated by the project can be considered for the PISTIS platform with respect to trustfull data sharing and exhange interoperabiliy
	Authority /ENISA [55]	ENISA (European Union Agency for Cybersecurity): ENISA is an EU agency that provides guidance and recommendations on various aspects of cybersecurity. It includes guidelines on data anonymization and pseudonymization techniques to protect personal data and ensure secure data sharing. PISTIS can utilize guidelines that are published by ENISA with respect to data privacy and security.
	EU Project / pdp4e [56]	PDP4E focus on spreading the creation of products, systems and services that better protect the privacy and personal data of EU citizens. To achieve this, PDP4E focues on delivering methods and software tools on data protection principles applications. A thorough review of the project could support PISTIS on the final tools and technologies to be adopted regarding secure and trustful data sharing.

6.2 Data Trading

6.2.1 Initiatives related to Distributed Ledgers

This section is focused on initiatives and projects for smart contract generation, transaction execution and data exchange on distributed ledgers, managing a process that transforms legal requirements for a data transaction into a smart contract. This usually means managing smart contract templates, collecting description of data assets, generating NFTs and designing a specific transaction (a single purchase, a subscription, a data investment plan, etc.).

Table 30 Initiatives related to distributed ledgers

#	INITIATIVE/ PROJECT NAME	SHORT DESCRIPTION AND RELEVANCE FOR PISTIS
1	Hyperledger ARIES [57]	Hyperledger Aries provides a shared, reusable, interoperable tool kit designed for initiatives and solutions focused on creating, transmitting, and storing verifiable digital credentials. It is infrastructure for

[54] https://zkproof.org/
[55] https://www.enisa.europa.eu/
[56] https://www.pdp4e-project.eu/
[57] https://www.hyperledger.org/use/aries

		blockchain-rooted, peer-to-peer interactions. This project consumes the cryptographic support provided by Hyperledger Ursa, to provide secure secret management and decentralized key management functionality. These concepts for storing and transmitting verifiable credentials as well as the key management can be used as a basis for PISTIS.
2	Hyperleger INDY [58]	Hyperledger Indy provides tools, libraries, and reusable components for providing digital identities rooted on blockchains or other distributed ledgers so that they are interoperable across administrative domains, applications, and any other silo. Indy is interoperable with other blockchains or can be used standalone powering the decentralization of identity. These decentralization concepts can be used as a basis in PISTIS.

6.2.2 Initiatives related to Data Monetization

This section is focused on the projects and initiatives related to data monetization and data pricing assessment. These projects are focused on the development of technologies enabling insights, data valuation services, data usage and intentions analytics offering to all the stakeholders a transparent and unbiased valuation of data, identifies potential revenue streams, and analyses how data is being used to create value.

Table 31 Initiatives related to data monetization

#	INITIATIVE/ PROJECT NAME	SHORT DESCRIPTION AND RELEVANCE FOR PISTIS
1	5GMETA [59]	The EU-funded 5GMETA project (H2020 ID: 957360) is developing an open platform to leverage car-captured data to stimulate and facilitate innovative products and services. It will empower the automotive ecosystem, from industry players to new entrants such as small and medium-sized enterprises and high-tech start-ups. Granting access to data from relevant geographical regions, the project will create new opportunities and business models from valuable services where data liability and billing will rely on an accountability dashboard of data-flow subscription and volume consumption. PISTIS can exploit the car data monetization services of 5GMETA for the design and development of its Data Trading and Value Exchange Platform.
2	Ataccama [60]	Ataccama ONE unifies Data Governance, Data Quality, and Master Data Management into a single, AI-powered fabric across hybrid and cloud environments. The dimensions it claims to be analysing are some of those relevant for quantifying the value of data.

[58] https://www.hyperledger.org/use/hyperledger-indy

[59] https://5gmeta-project.eu/

[60] https://www.ataccama.com/platform/master-data-management

3	CitizenMe [61] [62]	Marketplace for “ethically sourced” personal data, claiming to put the control over data in the hands of “digital citizens” who generate it. It then sells these data to businesses which are interested in using it.
4	Datacoup [63] [64]	Blockchain-enabled apps and tools for monetization of personal data. Not clear if the company still exists. Their webpage asks us to stay tuned for new features, but at the same time it states is now part of another company, called ODE.
5	Datum [65] [66] [67]	“Datum is a decentralized and distributed high performance NoSQL database backed by a blockchain ledger. This technology allows anyone to securely and anonymously backup structured data from social networks, wearables, smart homes, and other IoT devices. Datum provides a marketplace where users can share or sell data on their own terms.” Relevant to PISTIS because it developed a data market (personal data only!) that functions over the blockchain.
6	DATAMITE [68]	The EU-funded DATAMITE project (HORIZON EUROPE ID: 101092989) will develop a modular, open-source and multi-domain framework to improve technologies and solutions for data monetisation, interoperability, trading, and exchange in the form of software modules, training and business materials. The framework will provide users with the tools they need to improve the quality management of their data, adhering to FAIR principles and allowing upskilling on technical and business aspects that make data more reliable. At the external level, it will provide new sources of revenue and opportunities for interaction with other stakeholders. PISTIS may create a synergy with DATAMITE towards exchanging ideas in the data trading and monetization services.
7	Eurecat Data Valuation Platform [69]	A process and a platform for establishing the value of a structured data set. The process relies on i) the context in which the data set will be used (user provided, covering the areas of System & Economics, Legal & Obligations, Data Science, Data Properties, Business uses), ii) a data quality assessment, iii) a data utility assessment (including model performance and chance estimators) and iv) an assessment of the risk of deanonymisation (if applicable). The

[61] https://www.citizenme.com/

[62] https://www.citizenme.com/case-studies/

[63] https://www.odeinfinity.com/

[64] https://datacoup.com/

[65] https://datum.org/

[66] https://datum.org/assets/Datum-WhitePaper.pdf

[67] https://www.youtube.com/watch?v=NOTlH2sS1C4

[68] https://datamite-horizon.eu/

[69] https://safe-deed.eu/wp-content/uploads/2022/01/Safe-DEED_D4_5.pdf

		reporting of the data value is done in a top-down fashion and uses a visual scoring metaphor like that of energy labels (A-E).
8	FAME [70]	The EU-funded FAME project (HORIZON EUROPE ID: 101092639) will develop a secure federated data marketplace for embedded finance (EmFi). The aim is to demonstrate the full potential of the data economy. FAME will develop a federated cloud environment with multiple providers of EmFi data assets, including datasets, AI/ML models, and more. By connecting with more than a dozen data marketplaces, the project will roll out seven pilots. PISTIS may create a synergy with FAME towards exchanging ideas in the data trading and monetization services.
9	i3Market Data Pricing Recommendation Tool [71]	H2020 i3-Market (funded under EU H2020 Grant no. 871754) is a data pricing tool aimed at SMEs that are looking to participate in a data market either as buyers or sellers of data assets. The authors outline 10 data value properties, combining elements of collection cost, market insights (e.g., data scarcity, trust), data quality, data utility (in the form of analytics, estimated value to the customer). They use experts to establish the weights of each of these properties and plug them into a set of formulae for calculating the price of a data asset. While the method is simple and traceable, it is not completely clear how did the authors derive the formulae, nor does their validation seem to be very robust.
10	Informatica [72]	Their main product is an Intelligent Data Management Cloud, which provides cloud-native, AI-based data management capabilities including the following, relevant to the PISTIS solutions: data catalog, data integration, data quality, master data management, data marketplace. They claim to be able to manage data of any type, pattern, or complexity.
11	KNOW Centre Data Value Check [73]	This tool applies a cost-benefit analysis to assess the risks of implementation of data-driven use cases. The tool is not fully digitalized (it consists of a set of spreadsheets) and the methodology is heavily reliant on the interaction between assessors (consultants who know the tool) and the beneficiary (e.g., a company that needs its data to

[70] https://cordis.europa.eu/project/id/101092639

[71] https://www.i3-market.eu/data-pricing-recommendation-tool/

[72] https://www.informatica.com/platform.html

[73] https://www.know-center.at/en/business-success-story/reval-identifying-and-prioritizing-ai-use-cases/

		be assessed). The tool also lacks almost any kind of documentation or online presence and seems to be an in-house project.
12	Oracle Enterprise Data Quality [74]	"Oracle Enterprise Data Quality provides a comprehensive data quality management environment, used to understand, improve, protect, and govern data quality. The software facilitates best practice Master Data Management, Data Governance, Data Integration, Business Intelligence, and data migration initiatives, as well as providing integrated data quality in CRM and other applications and cloud services."
13	Permission.io [75]	Permission.io is a blockchain startup that pays you in crypto for watching advertisements, shopping online, and using data. The company has its own cryptocurrency, Permission Coin (ASK), that it offers as payment to users who engage with different brands and advertisers online.
14	Precisely [76]	This company has a suite of products providing some of the capabilities that the PISTIS platform will develop data enrichment, data integrity, data governance, data quality, master data management. These are elements that can contribute to data valuation. However, precisely offers them individually and its objective doesn't seem to be that of quantifying data value.
15	SAP Master Data Quality Management [77]	The Master Data Quality Management suite is part of the SAP Master Data Governance on SAP S/4HANA. It allows users to automatically extract data quality rules, manually define their own, run data quality assessment and data remediation. These are all desirable features for a data quality assessment system, such as that we wish to include in the PISTIS Platform.
16	Syniti [78]	They develop a suite of products that address some of the desired capabilities of the PISTIS Platform, which can impact data valuation: data quality assessment, data matching, master data management. However, these are not sufficient for full data valuation, and this doesn't seem to be part of their goals.
17	Snowflake Marketplace [79]	A solution for discovery, evaluation and purchase of data, data services and applications. Users can sample a dataset and test it against preset or custom business cases (in forms of data queries) before they decide to purchase it. The platform offers some form of data quality assessment under-the-hood, but it doesn't compute any scores. An embedded demo video nicely explains the capabilities of the platform.

[74] https://www.oracle.com/middleware/technologies/enterprise-data-quality.html

[75] https://permission.io/

[76] https://www.precisely.com/products

[77] https://assets.cdn.sap.com/sapcom/docs/2020/01/3cc3266c-7c7d-0010-87a3-c30de2ffd8ff.pdf

[78] https://www.syniti.com/solutions/

[79] https://www.snowflake.com/en/data-cloud/marketplace/

18	Talend [80] [81]	Talend offers solutions for Data Integration (transformation, mapping, enrichment with external sources), Data Quality (ML-based deduplication, validation, standardization), Data Integrity and Governance (quality checks, data catalogue, data lineage). They develop their own Talend Trust Score, to quantify and trace the reliability of a dataset. The company seems to be interested in the topic of data monetization and their tools appear to be valuable components in such a process, although not clear if sufficient.
19	Unidata [82]	Unidata is a multifunctional platform for building corporate data management systems, providing centralized data collection (inventory and accounting resources), standardization of information (normalization and enrichment), accounting current and historical information (control of record versions, periods data relevance), data quality and statistics.
20	UPCAST [83]	The EU-funded UPCAST (HORIZON EUROPE ID: 101093216) will provide a set of universal, trustworthy, transparent, and user-friendly data market plugins. These will enable actors in the European data spaces to design and deploy data exchange and trading operations. Four real-world pilots across Europe will operationalise a set of working platform plugins for data sharing, monetisation, and trading. PISTIS may create a synergy with UPCAST towards exchanging ideas in the data trading and monetization services.

6.2.3 Data Sources and Services

This section is focused on projects and initiatives related to data transfer using API interfaces, dataspace connectors and others emerging standards for interfacing with data ecosystem. This section include any projects for data discovery and for distributed querying, cataloguing and matchmaking. Other sets of initiatives to be investigated are those focused on data service configuration, providing a wide range support to the lifecycle of data models, including their creation, versioning, and deployment, and enabling users to discover and reuse existing data models as well as the propagation of semantics/ ontologies within data spaces to promote data reuse and consistency.

Table 32 Initiatives related to data sources and services

#	INITIATIVE/ PROJECT NAME	SHORT DESCRIPTION AND RELEVANCE FOR PISTIS
1	BDVA[84]	The Big Data Value Association (BDVA) is an industry-driven organisation with a mission to develop an innovation ecosystem

[80] https://www.talend.com/products/data-fabric/

[81] https://www.talend.com/resources/data-monetization/

[82] https://unidata-platform.com/products/

[83] https://www.upcast-project.eu/

[84] https://www.bdva.eu/

		that enables the data-driven digital transformation of the economy and society in Europe. PISTIS may benefit from BDVA for the design and development of its Data Management Platform services.
2	Catena-X / Tractus-X [85] [86]	Catena-X is an open and collaborative data ecosystem and exchange platform for the automotive industry. It is based on IDSA and Gaia-X specifications. All participants have the same rights and sovereignty. The main objective of the initiative is to create resilient and flexible supply chains in the automotive industry. The technical solution published as Open Source under the name Eclipse Tractus-X. It is well documented and offers a variety of services for reuse. Catena-X and Tractus-X are one of the most mature data space projects and implementations. It overlaps with many aspects and concepts of PISTIS and should be considered as a blueprint regarding implementation, governance, and documentation.
3	data.europa.eu - The official portal for European data [87]	The official portal for European data contains open data from the European public sector. PISTIS may benefit from such data for the training of its analytic services.
4	DSBA [88]	BDVA, FIWARE, Gaia-X and IDSA launched the Data Spaces Business Alliance (DSBA) with a common objective to accelerate business transformation in the data economy. One of the joint working areas of the DSBA is supporting the existing organisations and data spaces by pooling their tools, resources, and expertise in a focused way. PISTIS may benefit from DSBA for the design and development of its Data Management Platform services.
5	DATAMITE [89]	DATA Monetization, Interoperability, Trading & Exchange (DATAMITE) is a sister Horizon 2023 project started in January 2023, with similar topics and technical architecture as in PISTIS. DATAMITE delivers a modular, open-source and multi-domain Framework to improve data monetising, interoperability, trading, and exchange in the form of software modules, training and business materials for European companies, empowering them to become new relevant players in the data economy. The project will provide contributions relevant for PISTIS: • Technological stack to exchange quality data in data spaces • Integration of several IDSA (or alternative) tools, including connectors in the Data Sharing Module

[85] https://catena-x.net/en/

[86] https://eclipse-tractusx.github.io

[87] https://data.europa.eu/en

[88] https://data-spaces-business-alliance.eu/

[89] https://datamite-horizon.eu/

		• Easy-to-use interoperability and sovereignty tools for data exchange DATAMITE outcomes are relevant for many PISTIS tasks.
6	DataBri-X [90]	Data Process & Technological Bricks for expanding digital value creation in European Data Spaces (DataBri-X) is a Horizon 2023 project started in October 2022 with a total duration of 36 months. Its ambition is to realise a cross-border, cross-sectoral sharing data space and enable platforms to process proprietary, personal, and open public data requires overcoming technical, legal and business challenges along the data life cycle. The project will deliver a holistic, energy-efficient, and user-friendly toolbox of practical, robust, and scalable bricks/Bri-X to improve data and metadata interoperability, usability, discoverability, quality, and integrity and expand digital value creation in the context of European data spaces. The toolbox will align with accountability, fairness, privacy and confidentiality regulations and FAIR principles. The project will provide tools to support a holistic approach of the data lifecycle in compliance with FAIR principles. These tools may be of high relevance for PISTIS. DataBri-X will build on results of relevant past and current initiatives, data management tools, systems and processes that enable automated creation and maintenance of common ontologies, vocabularies, and data models, as well as automated authoring, co-creation, curation, annotation and labelling of data. The project will address issues, such as data ownership, data provenance veracity and verification, decentralised data sharing, confidentiality, and digital rights management. All these topics are highly relevant for PISTIS. On the DataBri-X use cases is an Energy Data space: enabler of a large-scale energy community data market. It is recommended that PISTIS takes this work into consideration regarding identity related tasks within the projects.
7	DataHub by LinkedIn [91]	DataHub is a metadata search and discovery tool developed by LinkedIn and introduced as "an open-source metadata platform for the modern data stack". As the data grows, it becomes more challenging for LinkedIn employees to discover, understand and take advantage of the available data. To overcome these challenges and enable them to discover data that matters to them, they re-architected their data catalogue system resulting in what we know now as DataHub, which consist of two distinct stacks: a modular UI frontend and a generalized metadata architecture backend. Datahub has been running in LinkedIn production since 2019 and many companies have adopted it to manage their data.

[90] https://databri-x.eu

[91] https://datahubproject.io/

		DataHub current architecture is designed to handle a huge amount of data in what people consider now as the golden age of data. Apart from its potential to be a data catalogue in PISTIS, its metadata management architecture and use cases are worth to be explored to see how PISTIS can benefit from this tool.
8	Dat Ecosystem [92]	Dat Ecosystem is a global community of many projects, which most of them are self-funded. These projects are built on top of Hypercore protocol, a secure transport protocol that makes it easy to build fast and scalable peer-to-peer applications. They work side by side on open and secure protocols for the web of commons. Some of the projects contribute maintenance and development to core pieces of the Dat Ecosystem while others create high level applications based on the peer-to-peer protocols. This community is relevant to PISTIS since they promote secure peer-to-peer applications. It might be interesting to see their use cases to get more insight and learn from them.
9	DIF [93]	DIF is a alliance of organizations, who build standards and technologies for decentralized identities. The objective is to achieve interoperability in the scope of digital identities through harmonized protocols, data formats and reference implementations. DIF is organized in several working groups, such as DID communication, wallet security, claims and credentials. It is recommended that PISTIS takes this work into consideration regarding identity related tasks within the projects.
10	DWeb [94]	DWeb is an umbrella organization that connects people, projects, and organizations in the domain of the decentralized Web. They organize regular events and meetups to exchange ideas and latest trends. For PISTIS the initiative is relevant as a channel for dissemination and feedback.
11	EDC [95]	The Eclipse Dataspace Components (EDC) will implement the International Data Spaces (IDS) standard as well as relevant protocols and requirements associated with Gaia-X, and thereby provide implementation and feedback to these initiatives. However, it will be extensible in a way that it may support alternative protocols. Similar with the works by IDSA, EDC is also highly relevant for PISTIS, since it provides the technical components to develop PISTIS data ecosystem.

[92] https://dat-ecosystem.org/

[93] https://identity.foundation/

[94] https://getdweb.net/

[95] https://github.com/eclipse-edc/

12	EBSI [96]	European Blockchain Service Infrastructure (EBSI) is a partnership of all EU Member States, Norway, Liechtenstein, and the European Commission. It constitutes the first pan-European Blockchain initiative with active nodes running distributed in many countries in Europe. The objective is to provide a public Blockchain infrastructure for cross-border data sharing and application development. The focus is the domain of public services. EBSI is based on Hyperledger Fabric, Hyperledger Indy and the W3C standard for Verifiable Credentials. Currently many pilot use cases are being developed. An application for an early adopters' program is possible. EBSI is a relevant blueprint, how to govern and deploy a cross-border Blockchain network. PISTIS can adopt relevant processes and artefacts.
13	FAME [97]	Federated decentralized trusted data Marketplace for Embedded finance (FAME) is a Horizon 2023 sister project running parallel to PISTIS. It will provide the means for integrating, pricing, and trading data assets from interconnected, federated data management infrastructures, including heterogeneous data spaces and data marketplaces. The project will provide the means for aggregating and integrating data assets from different providers in a federated catalogue. FAME outcomes are relevant for many PISTIS tasks.
14	FIWARE [98]	FIWARE Foundation drives the definition - and the open-source implementation - of key open standards that enable the development of portable and interoperable smart solutions in a faster, easier, and affordable way, avoiding vendor lock-in scenarios, whilst also nurturing FIWARE as a sustainable and innovation-driven business ecosystem. The open APIs offered by FIWARE may assist PISTIS in its data management operations.
15	Gaia-X [99]	Gaia-X is a European data infrastructure initiative to ensure European digital sovereignty. Its objective is not to be a cloud service provider, but to establish a federated, secure, and trustworthy ecosystem that links many cloud service providers and users together to share their data. Within this ecosystem, it will be possible to provide, share, and use data within a trustworthy environment. Thus, spurring innovation and creating added value for the data economy to all who share data. Gaia-X is highly relevant for PISTIS, since its very existence is to enable secure data sharing among trusted participants in Europe, which is one of the core values of PISTIS. PISTIS may

[96] https://ec.europa.eu/digital-building-blocks/wikis/display/EBSI/Home

[97] https://www.fame-horizon.eu/

[98] https://www.fiware.org/

[99] https://gaia-x.eu/

		benefit from Gaia-X for the design and development of its Data Management Platform services.
16	GXFS [100]	The Gaia-X Federation Services (GXFS) can be seen as a toolbox to achieve the objective of Gaia-X. It provides a set of open-source software components to set up a Gaia-X compliant federated and secure ecosystem. The Federation is an ecosystem where individual participants work together to develop and offer services to be used within the Federation. Since it is a toolbox for Gaia-X, GXFS is also highly relevant for PISTIS to implement the data ecosystem.
17	Hyperledger [101]	Hyperledger is an initiative and collection of open-source solutions to build and maintain blockchain networks and related technologies. Hyperledger is part of the Linux Foundation. It currently consists of six mature projects, covering a variety of use cases and scenarios: Aries provides tools for managing Verifiable Credentials, Besu is an Ethereum client, Fabric is a generic framework to build permissioned distributed ledger networks, Indy focuses on the decentralized management of identities, Iroha emphasizes on IoT domain, and Sawtooth on a modular approach for building enterprise distributed ledgers. The Hyperledger project is highly relevant for PISTIS, since it can be a criticial building block for implementing the decentralized features of the solution.
18	IDSA	Data spaces enable organizations to securely share data with others. The International Data Spaces Association (IDSA) is a non-profit coalition of companies, scientist, lawmakers and other relevant stakeholders of International Data Spaces (IDS), a secure, sovereign system of data sharing in which all participants can take full advantage of their data. IDS follows the same vision as Gaia-X, proliferating data sovereignty and create an ecosystem of trust for data sharing. The IDSA has established the important technical standard, such as the architecture, interfaces and sample code for an open and secure data ecosystem of trusted partners, in order to create the future of the global, secure and digital economy. The members of the International Data Spaces Association develop use cases and even entire IDS-based data spaces that can host a wide range of use cases. These front-running data spaces and use cases show how the IDS standard adopts in real-life challenges and becomes a widely agreed and applied de facto standard. The works that has been done by this association are highly relevant to PISTIS due to their vision. They have established the standard that could help PISTIS in term of technical

[100] https://www.gxfs.eu/

[101] https://www.hyperledger.org/

		development. PISTIS may benefit from IDSA for the design and development of its Data Management Platform services.
19	LinkedDataHub [102]	LinkedDataHub is an open-source platform for collaboratively working and managing linked data, which does not require the data providers to have an extensive knowledge of web application development and deployment to set up their own data management application. By default, the application includes a SPARQL endpoint and a Linked Data API which do not require yet another development efforts. It is claimed to be the WordPress for managing Knowledge Graphs. One can run his/her own instance of LinkedDataHub or simply use an existing online instance. This platfrom enables the data owners and/or providers to manage their linked data easily without being overwhelmed by the installation and technical process. Their approach to make linked data management as easy as possible is something that PISTIS can learn from. Since it is an open-source project, it has the potential to be a data catalogue in PISTIS and there might be rooms to further extend the project to meet PISTIS objectives.
20	Linking Data in Europe [103]	This project engages EuroSDR (European Spatial Data Research) members and their network in the linkage of datasets across Europe. The publication of metadata as linked data and the creation of links between metadata records are studied. The motivation is to set up a Metadata Knowledge Graph (MKG) ecosystem to support the joint usability of different digital assets that exist in Europe. PISTIS may benefit from this project for the design and development of its data matchmaking services.
21	Ocean Protocol [104]	The Ocean Protocol is a decentralized, Blockchain-based, Web3-inspired and open marketplace for data. The basic concept is the notion of data NFTs (non-fungible tokens) and smarts contracts, that manage the publication and consumption of data services. Currently, the Ocean Protocol runs on the public Ethereum blockchain. The protocol supports the management of private data, that never leaves the premises of the owner. The objective of the initiative is to monetize data trading, while maintaining sovereignty. It is possible to create custom marketplaces. The Ocean Protocol is highly relevant for PISTIS, since is employs a Blockchain as backend. Methodologies and standards can be transferred to PISTIS.

[102] https://atomgraph.com/products/linkeddatahub/
[103] https://www.eurosdr.net/research/project/linking-data-europe
[104] https://oceanprotocol.com/

22	piveau [105]	Piveau is an open-source metadata catalogue solution. It is highly scalable and covers the essential life cycle of your metadata: harvesting, storage and quality assurance. Piveau was designed and developed around Semantic Web technologies, the W3C standard DCAT and the European standard for Open Data DCAT-AP. It closes the gap between formal metadata specifications and their application in production. Piveau is a relevant foundation for building metadata and schema repositories for PISTIS. It can be extended to fit specific needs and the Linked Data support allows to integrate specifications from the broader Semantic Web ecosystem.
23	Solid [106]	Solid is a protocol for managing personal data in a decentralized manner. It is built upon existing W3C specifications. The basic concept is the storage of structured and unstructured data in Personal Online Data Stores – so called Pods. These Pods are accessible through the Web. It is possible to develop apps on top of the Pods. Users have fine-grained control over their data via Access Control Lists (ACL). Solid does not rely on centralized services for identity or data descriptions. A reference implementation is available as open-source software. Solid is relevant for PISTIS, since the standard combines decentralization and concepts of the Semantic Web to securely exchange data. It can be adopted beyond the scope of personal data.
24	TANGO [107]	Digital Technology for Secure and Trustworthy Data Flows (TANGO) is a Horizon 2023 sister project started in November 2022. The project aims to provide a trustworthy data management and sharing solution to ensure data sovereignty, governance and provenance to citizens, businesses, and public administration around Europe. The project will develop technologies for industrial data sharing: • ensuring privacy, security & trustworthiness • Data & AI toolkit • Legal and ethical compliance. TANGO outcomes are relevant for many PISTIS tasks.

[105] https://doc.piveau.io

[106] https://solidproject.org/

[107] https://tango-project.eu

7 CONCLUSION

D1.1 encloses multifaceted information that will be used throughout the execution of the project as a reference. The related content span from the end – to – end scenarios and data life cycle of the use cases to rules and regulations that follow the data management and trading processes nowadays in Europe. Moreover, a thorough study on the use case scenarios and a respective business analysis have generated a list of business requirements that will be used to guide the generation of the technical requirements (functional and non – functional). The outcome of the process will be the final design of the architecture and the implementation of the platform.

While D1.1 is initiated by the internal processes of the PISTIS project, as a public deliverable is expected to have significant contribution to the literature of the domain since it reports popular approaches and technologies that are followed nowadays on data management and trading processes by various organizations in multiple application domains.

8 REFERENCES

[1] T. Madiega, *Digital sovereignty for Europe,* European Parliamentary Research Service, July 2020.

[2] E. Commission, 2030 Digital Compass: the European Way for the Digital Decade, 9 March 2021.

[3] E. P. a. E. Council, *European Declaration on Digital Rights and Principles for the Digital Decade,* Decision on the Digital Decade Policy Programme 2030, 23 January 2023.

[4] E. Commission, *Digital Decade 2030 Objectives.*

[5] E. Commission, *A European Data Strategy,* 19 February 2020.

[6] *The EU's Cybersecurity Strategy for the Digital Decade,* European Commission and the High Representative of the Union for Foreign Affairs and Security Policy, 16 December 2020.

[7] *Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union..*

[8] *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)..*

[9] *COM(2020) 66 final: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European Strategy for Data.*

[10] *European Commission (2022). Data Governance Act explained. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained.*

[11] *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.*

[12] *Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use.*

[13] *European Commission (2022). Data Act: Commission Proposes measures for fair and innovative data economy. Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113..*

[14] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal*

data and on the free movement of such data, and repealing Directive 95/46/EC (General Da.

[15] *European Data Protection Supervisor, Opinion 3/2020 on the European Strategy for Data, 16 June 2020..*

[16] *European Data Protection Board, Guidelines 07/2020 on the concepts of controller and processor in the GDPR, Version 2.1, 7 July 2021.*

[17] *European Data Protection Board, Guidelines 07/2020 on the concepts of controller and processor in the GDPR, Version 2.1, 7 July 2021.*

[18] *40 General Data Protection Regulation, Recital 90.*

[19] *General Data Protection Regulation, Recital 91.*

[20] *Article 29 Data Protection Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, As last Revised and Adopted on 4 October 2017.*

[21] *General Data Protection Regulation, Article 35.*

[22] *General Data Protection Regulation, Article 35.*

[23] *Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).*

[24] *Digital Market Act, Chapter IV.*

[25] *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act.*

[26] *Digital Services Act, Article 1.*

[27] *Digital Services Act, Article 52.*

[28] *Digital Services Act, Article 18.*

[29] *Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market.*

[30] *COM(2020) 593 final: Proposal for a Regulation of the European Parliament and of the Council on Markets in Crypto-assets, and amending Directive (EU) 2019/1937.*

[31] *European Parliament legislative resolution of 20 April 2023 on the proposal for a regulation of the European Parliament and of the Council on Markets in Crypto-assets and amending Directive (EU) 2019/1937, Article 1(1).*

[32] *European Parliament legislative resolution of 20 April 2023 on the proposal for a regulation of the European Parliament and of the Council on Markets in Crypto-assets and amending Directive (EU) 2019/1937, Article 93.*

[33] *Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (E.*

[34] *NIS2 Directive, Article 3 (1) (a) and (c).*

[35] *COM(2022) 454 final: Proposal for a Regulation of the European Parliament and of the Council on horizontal cybersecurity requirements for products with digital elements and amending Regulation (EU) 2019/1020.*

[36] *Proposal for the Cyber Resilience Act, Article 5.*

[37] *COM(2021) 206 final: Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.*

[38] *Proposal for AI Act, Article 1.*

[39] *EU Parliament Briefing on AI Act, First edition, January 2022.*

[40] *Access to the European Blockchain Regulatory Sandbox, https://ec.europa.eu/digital-building-blocks/wikis/display/EBSI/Sandbox+Project.*

[41] *Thibault Schrepe, Smart Contracts and the Digital Single Market Through the Lens of a "Law + Technology" Approach, European Commission, Directorate-General for Communications Networks, 2021..*

[42] *Adam J Kolber, Not-So-Smart Blockchain Contracts and Artificial Responsibility, STAN. TECH. L. REV, 2018..*

[43] *Proposal on Data Act, Recital 80 and Article 2.*

[44] *Proposal on Data Act, Article 30.*

[45] *Olli Pitkanen and Juhani Luoma-Kyyny, Rulebook for a Fair Data Economy, The Finnish Innovation Fund Sitra, 31 August 2022..*

[46] *European Commission, Guidance on Ethics By Design and Ethics of Use Approaches for Artificial Intelligence, 25 November 2021..*

[47] *European Commission, Guidance on Sharing Private Sector Data in the European Data Economy, Commission Staff Working Document, 25 April 2018..*

[48] *Birch, K., Cochrane, D., & Ward, C. (2021). Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. Big Data & Society, 8(1). https://doi.org/10.1177/20539517211017308.*

[49] *Global Partnership for Sustainable Development Data, What Do We Know About the Value of Data?, Jenna Slotin, 2018..*

[50] *Short, J.E., and Todd, S. What's Your Data Worth? MIT Sloan Management Review. 2017.*

[51] *European Data Protection Board, Guidelines 2/2019 on the processing of personal data in the context of online services to data subjects, 16 October 2019 parag. 54.*

[52] *European Data Protection Board, Statement 05/2021 on the Data Governance Act in light of the legislative developments, 19 May 2021.*

[53] *European Data Protection Board and European Data Protection Supervisory, Joint Opinion 2/2022 on the Proposal of the European Parliament and of the Council on harmonised rules on fair access to and use of data, 4 May 2022.*

[54] *G. Action, GODAN Action Online Course on Open Data Management in Agriculture and Nutrition (Version v1.0), 2019..*

[55] *"Data on the Web Best Practices," W3C Recommendation, [Online]. Available: https://www.w3.org/TR/dwbp/. [Accessed 9 June 2023]..*

[56] *D. S. K. P. Z. Y. a. K. V. Colin Puri, "Implementing a Data Lineage Tracker," Berlin, Heidelberg, 2012..*

[57] *F. A. Hofmann, "Tracer: A Machine Learning Approach to Data," Massachusetts , 2020..*

[58] *Y. Jolly, "www2.deloitte.com," Deloitte AG, [Online]. Available: www2.deloitte.com. [Accessed 06 08 2023]..*

[59] *Praitheeshan, Purathani et al. "Private and Trustworthy Distributed Lending Model Using Hyperledger Besu." SN Computer Science 2 (2021): 1-19.*

[60] *Lesavre, L. et.al.: A Taxonomic Approach Understanding Emerging Blockchain Identity Management Systems: National Institute of Standards and Technology, 2020..*

[61] *Shuaib M, Hassan NH, Usman S, Alam S, Bhatia S, Agarwal P, Idrees SM. Land Registry Framework Based on Self-Sovereign Identity (SSI) for Environmental Sustainability. Sustainability. 2022; 14(9):5400..*

[62] *Guo H, Yu X (2022) A survey on blockchain technology and its security. Blockchain Res Appl 3(2):1–15. https://doi.org/10.1016/j.bcra.2022.100067.*

[63] *Praitheeshan, Purathani et al. "Private and Trustworthy Distributed Lending Model Using Hyperledger Besu." SN Computer Science 2 (2021): 1-19..*

[64] *NIST SP 800-30 Rev. 1 - Guide for Conducting Risk Assessments.*

[65] *ISO/IEC 27001:2013 - Information technology - Security techniques - Information security management systems - Requirements.*

[66] *SANS Institute - Security Policy Project Templates.*

[67] *General Data Protection Regulation (GDPR) - Article 13 - Information to be provided where personal data are collected from the data subject.*

[68] *NIST SP 800-53 Rev. 5 - Security and Privacy Controls for Information Systems and Organizations.*

[69] *NIST SP 800-53 Rev. 5 - Security and Privacy Controls for Information Systems and Organizations.*

[70] *International Association of Privacy Professionals (IAPP) - Data Sharing Agreements.*

[71] *National Cyber Security Centre (NCSC) - Staff awareness: the importance of training.*

[72] *SANS Institute - Incident Response Policy.*

[73] *ISACA - COBIT 2019 Framework.*

[74] *Rouse, M. (2019, December 20). Data brokering. Techopedia. https://www.techopedia.com/definition/34111/data-brokering.*

[75] *Crain, M. (2018). The limits of transparency: Data brokers and commodification. new media & society, 20(1), 88-104..*

[76] *Ferraro, A. (2022, August 19). What are data brokers? everything you need to know. Privacy Blog. https://privacy.com/blog/what-are-data-brokers.*

[77] *Burdova, C. (2021, December 1). Data Brokers: Who They Are and How They Work. AVG Signal. https://www.avg.com/en/signal/data-brokers.*

[78] *Stsiopkina, M. (2021, August 18). What is financial data? Oxylabs. https://oxylabs.io/blog/what-is-financial-data.*

[79] *Guy, J.-S. (2021, April 19). The rise of geodata marketplaces. Korem. https://www.korem.com/the-rise-of-geodata-marketplaces/.*

[80] *Sober, M., Scaffino, G., Schulte, S., & Kanhere, S. S. (2022). A blockchain-based IoT data marketplace. Cluster Computing, 1-23..*

[81] *Barnhart, B. (2018, August 28). How to mine your social media data for a better ROI. Sprout Social. https://sproutsocial.com/insights/social-media-data/.*

[82] *10 best companies operating in the social media analytics industry. Fortune Business Insights. (2021, May 7). https://www.fortunebusinessinsights.com/blog/10-best-companies-in-the-social-media-analytics-industry-10557.*

[83] *Gaikwad, M. (2020, July 3). What is social media analytics? definition, best practices, and examples. Spiceworks. https://www.spiceworks.com/marketing/advertising/articles/what-is-social-media-analytics/.*

[84] *Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. Information systems management, 29(4), 258-268.*

[85] *Open Data Essentials. The World Bank. (n.d.). http://opendatatoolkit.worldbank.org/en/essentials.html.*

[86] *Kelly, L. (2023) 'Ultimate Guide to The Data Marketplace in 2023', Datarade. Available at: https://about.datarade.ai/data-marketplaces (Accessed: 16 June 2023)..*

[87] *Wixom, B.H. (2014) 'Cashing In on Your Data', CISR Research Briefing, XIV(8), pp. 1–4..*

[88] *LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). "Big data, analytics and the path from insights to value." MIT Sloan Management Review, 52(2), 21-32..*

[89] *Redman, T. C. (2008). "Data driven: Creating a data culture." IT Metrics Strategies, 2(6), 1-12..*

[90] *Marr, B. (2017). "Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results." John Wiley & Sons..*

[91] *Gandomi, A., & Haider, M. (2015). "Beyond the hype: Big data concepts, methods, and analytics." International Journal of Information Management, 35(2), 137-144..*

[92] *Ross, J. W., Beath, C. M., & Sebastian, I. M. (2017). "How to develop a great data strategy." MIT Sloan Management Review, 58(2), 1-9..*

[93] *Manyika, J., Chui, M., & Brown, B. (2011). "Big data: The next frontier for innovation, competition, and productivity." McKinsey Global Institute..*

[94] *Laney, D. (2001). "3D Data Management: Controlling Data Volume, Velocity, and Variety." META Group Research Note, 6(70), 1-9..*

[95] *Marr, B. (2017). "Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results." John Wiley & Sons..*

[96] *Fleckenstein, M., Obaidi, A. and Tryfona, N. (2023) 'A Review of Data Valuation Approaches and Building and Scoring a Data Valuation Model', Harvard Data Science Review, 5(1). Available at: https://doi.org/10.1162/99608f92.c18db966..*

[97] *Viscusi, G. and Batini, C. (2014) 'Digital Information Asset Evaluation: Characteristics and Dimensions', in L. Caporarello, B. Di Martino, and M. Martinez (eds) Smart Organizations and Smart Artifacts..*

[98] *Batini, C. et al. (2018) 'Digital information asset evaluation: A case study in manufacturing', ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 49(3), pp. 19–33..*

[99] *Ahituv, N. (1980) 'A Systematic Approach Towards Assessing the Value of an Information System', MIS Quarterly, 4(4), pp. 61–75.*

[100] *Faroukhi, A.Z. et al. (2020) 'Big data monetization throughout Big Data Value Chain: a comprehensive review', Journal of Big Data, 7(1), p. 3. Available at: https://doi.org/10.1186/s40537-019-0281-5..*

[101] *Kelly, L. (2023) 'Ultimate Guide to The Data Marketplace in 2023', Datarade. Available at: https://about.datarade.ai/data-marketplaces (Accessed: 16 June 2023)..*

[102] *Cappa, F. et al. (2021) 'Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance*', Journal of Product Innovation Management, 38(1), pp. 49–67. Available at: http.*

[103] *Barney, J.B., Wright, M. and Ketchen, D.J. (2001) 'The resource-based view of the firm: Ten years after 1991', Journal of Management, p. 17..*

[104] *Cappa, F. et al. (2021) 'Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance*', Journal of Product Innovation Management, 38(1), pp. 49–67. Available at: http.*

[105] *Parvinen, P. et al. (2020) 'Advancing Data Monetization and the Creation of Data-based Business Models', Communications of the Association for Information Systems, 47(1), pp. 25–49. Available at: https://doi.org/10.17705/1CAIS.04702..*

[106] *Hanafizadeh, P. and Harati Nik, M.R. (2020) 'Configuration of Data Monetization: A Review of Literature with Thematic Analysis', Global Journal of Flexible Systems Management, 21(1), pp. 17–34. Available at: https://doi.org/10.1007/s40171-019-00228-3..*

[107] J. Smith, "How to create references using the bibliography tool in ms word," *A nice journal,* pp. 12-14, 1990.

[108] *Free Flow of Non-Personal Data Regulation, Article 2.*

Appendix A: Workshops' Miro Boards

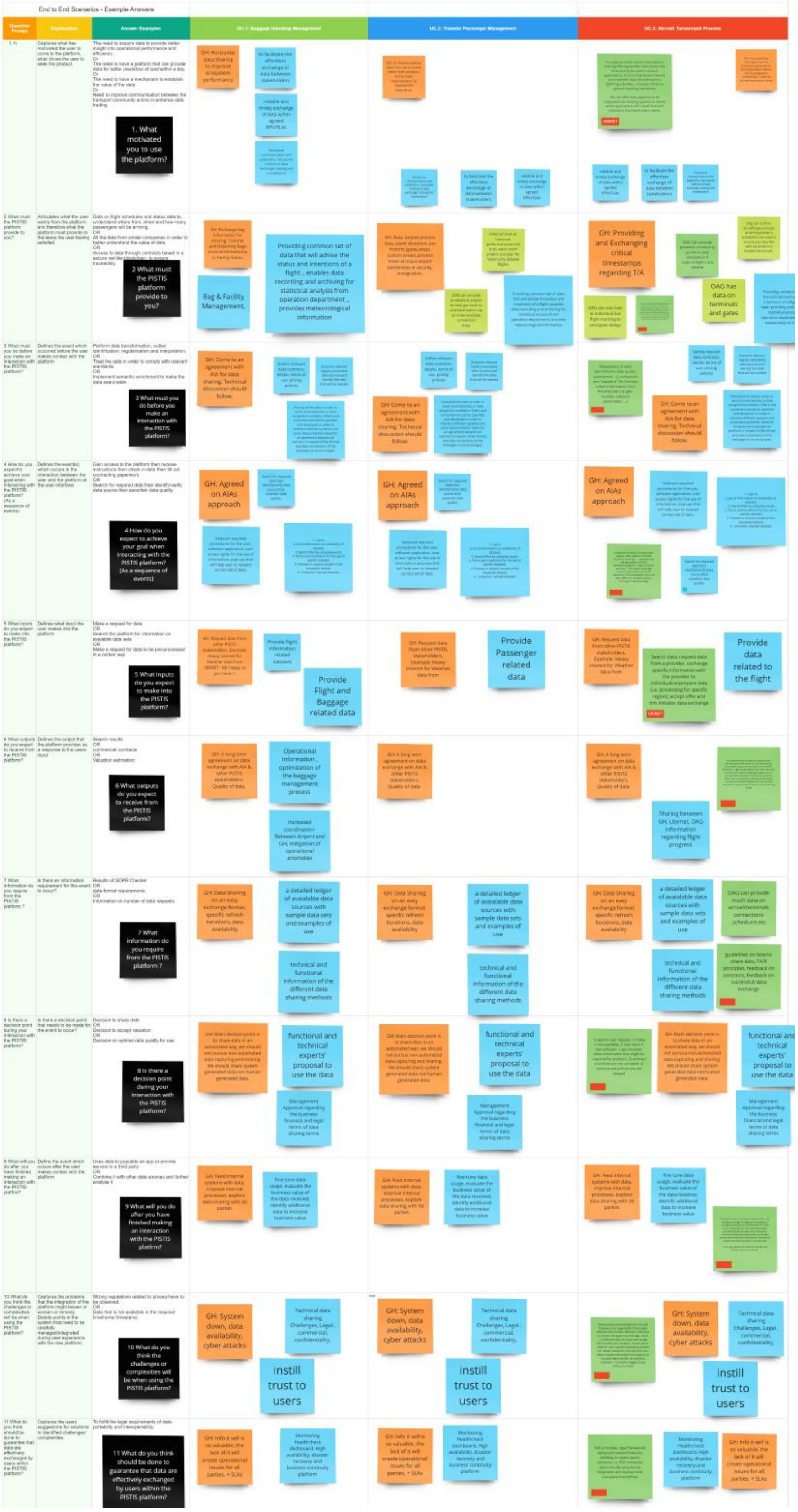
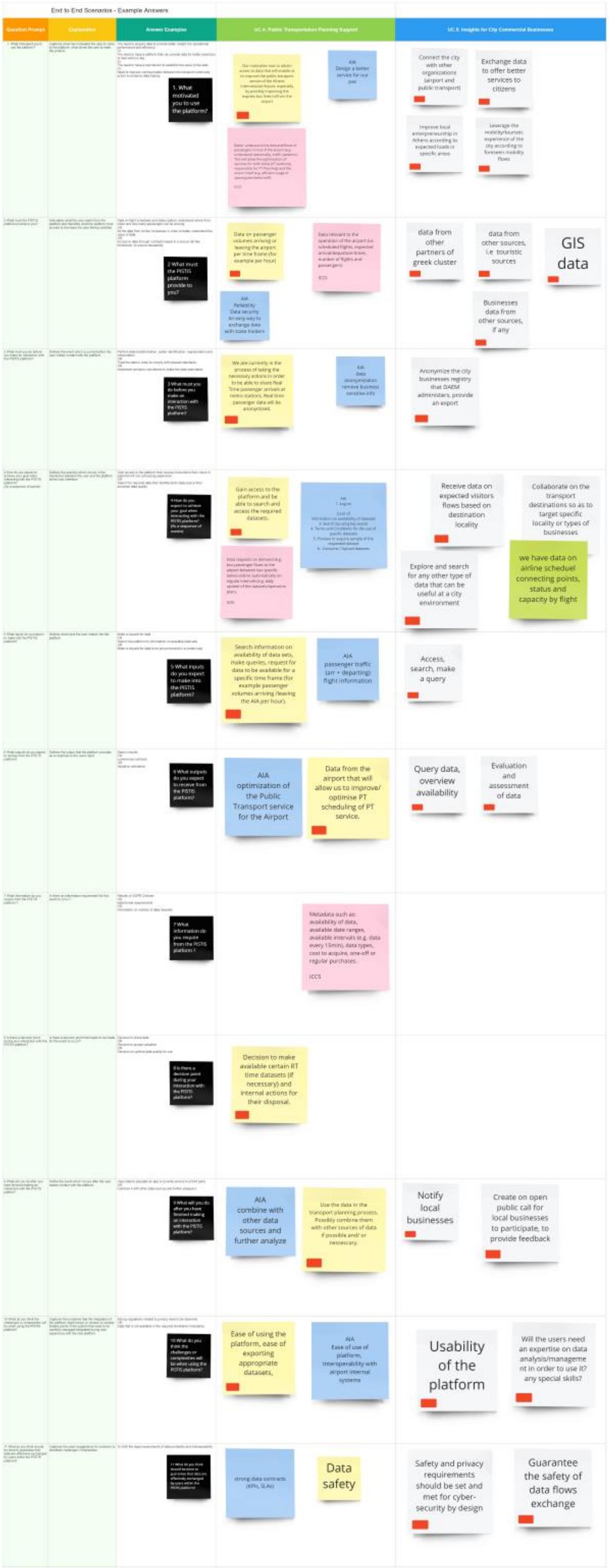

End to End Scenarios - Example Answers
Question Prompt
Explanation
Answer Examples
UC 1: Baggage Handling Management
UC 2: Transfer Passenger Management
UC 3: Aircraft Turnaround Process
1. What motivated you to use the platform?
GH: Horizontal Data Sharing to improve ecosystem performance
to facilitate the effortless exchange of data between stakeholders
reliable and timely exchange of data within agreed KPI/SLAs
2 What must the PISTIS platform provide to you?
Bag & Facility Management.
Providing common set of data that will advise the status and intentions of a flight.. enables data recording and archiving for statistical analysis from operation department , provides meteorological information
GH: Providing and Exchanging critical timestamps regarding T/A
OAG has data on terminals and gates
3 What must you do before you make an interaction with the PISTIS platform?
GH: Come to an agreement with AIA for data sharing. Technical discussion should follow.
4 How do you expect to achieve your goal when interacting with the PISTIS platform? (As a sequence of events)
GH: Agreed on AIAs approach
5 What inputs do you expect to make into the PISTIS platform?
Provide Flight and Baggage related data
Provide Passenger related data
Provide data related to the flight
6 What outputs do you expect to receive from the PISTIS platform?
GH: A long term agreement on data exchange with AIA & other PISTIS stakeholders. Quality of data.
Operational information , optimization of the baggage management process
Increased coordination Between Airport and GH, mitigation of operational anomalies
Sharing between GH, Ubimet, OAG information regarding flight progress
7 What information do you require from the PISTIS platform ?
GH: Data Sharing on an easy exchange format, specific refresh iterations, data availability
a detailed ledger of avalaiable data sources with sample data sets and examples of use
technical and functional information of the different data sharing methods
OAG can provide much data on arrival/terminals connections /scheduels etc
guidelines on how to share data, FAIR principles, feedback on contracts, feedback on successful data exchange
8 Is there a decision point during your interaction with the PISTIS platform?
functional and technical experts' proposal to use the data
Management Approval regarding the business financial and legal terms of data sharing terms
9 What will you do after you have finished making an interaction with the PISTIS platfrm?
GH: Feed internal systems with data, improve internal processes, explore data sharing with 3d parties
fine tune data usage, evaluate the business value of the data received, identify additional data to increase business value
10 What do you think the challenges or complexities will be when using the PISTIS platform?
GH: System down, data availability, cyber attacks
Technical data sharing Challenges, Legal , commercial, confidentiality,
instill trust to users
11 What do you think should be done to guarantee that data are effectively exchanged by users within the PISTIS platform?
GH: Info it self is so valuable, the lack of it will create operational issues for all parties. + SLAs
Monitoring Healthcheck dashboard, High availability, disaster recovery and business continuity platform

Figure 9: Miro board from mobility hub workshop for UC 1-2-3



End to End Scenarios - Example Answers

Question Prompt	Explanation	Answer Examples	UC 4: Public Transportation Planning Support	UC 5: Insights for City Commercial Businesses
[illegible]	[illegible]	[illegible] 1. What motivated you to use the platform?	Our motivation was to obtain access to data that will enable us to improve the public transport service of the Athens International Airport, especially by possibly improving the express bus lines to/from the airport AIA Design a better service for our pax Better understand the demand/flows of passengers in/out of the airport (e.g. understand seasonality, traffic patterns). This will allow the optimisation of services for both OASA (PT authority responsible for PT Planning) and the airport itself (e.g. efficient usage of [illegible]). ICCS	Connect the city with other organizations (airport and public transport) Exchange data to offer better services to citizens Improve local enterpreneurship in Athens according to expected loads in specific areas Leverage the mobility/touristic experience of the city according to foreseen mobility flows
[illegible]	[illegible]	[illegible] 2 What must the PISTIS platform provide to you?	Data on passenger volumes arriving or leaving the airport per time frame (for example per hour) Data relevant to the operation of the airport (i.e. scheduled flights, expected arrival/departure times, number of flights and passengers) ICCS AIA Reliability Data security An easy way to exchange data with stake holders	data from other partners of greek cluster data from other sources, i.e touristic sources GIS data Businesses data from other sources, if any
[illegible]	[illegible]	[illegible] 3 What must you do before you make an interaction with the PISTIS platform?	We are currently in the process of taking the necessary actions in order to be able to share Real Time passenger arrivals at metro stations. Real time passenger data will be anonymized. AIA data anonymization remove business sensitive info	Anonymize the city businesses registry that DAEM administers, provide an export
[illegible]	[illegible]	[illegible] 4 How do you expect to achieve your goal when interacting with the PISTIS platform? (As a sequence of events)	Gain access to the platform and be able to search and access the required datasets. AIA 1. Log in 2. List of information on availability of datasets 3. Search (using key words) 4. Terms and Conditions for the use of specific datasets 5. Preview or acquire sample of the requested dataset 6. Consume / Upload datasets Data requests on demand e.g. bus passenger flows at the airport between two specific dates; link to automatically on regular intervals e.g. daily update of the datasets/operation plans. ICCS	Receive data on expected visitors flows based on destination locality Collaborate on the transport destinations so as to target specific locality or types of businesses Explore and search for any other type of data that can be useful at a city environment we have data on airline scheduel connecting points, status and capacity by flight
[illegible]	[illegible]	[illegible] 5 What inputs do you expect to make into the PISTIS platform?	Search information on availability of data sets, make queries, request for data to be available for a specific time frame (for example passenger volumes arriving /leaving the AIA per hour). AIA passenger traffic (arr + departing) flight information	Access, search, make a query
[illegible]	[illegible]	[illegible] 6 What outputs do you expect to receive from the PISTIS platform?	AIA optimization of the Public Transport service for the Airport Data from the airport that will allow us to improve/ optimise PT scheduling of PT service.	Query data, overview availability Evaluation and assessment of data
[illegible]	[illegible]	[illegible] 7 What information do you require from the PISTIS platform ?	Metadata such as: availability of data, available date ranges, available intervals (e.g. data every 15min), data types, cost to acquire, one-off or regular purchases. ICCS	
[illegible]	[illegible]	[illegible] 8 Is there a decision point during your interaction with the PISTIS platform?	Decision to make available certain RT time datasets (if necessary) and internal actions for their disposal.	
[illegible]	[illegible]	[illegible] 9 What will you do after you have finished making an interaction with the PISTIS platform?	AIA combine with other data sources and further analyze Use the data in the transport planning process. Possibly combine them with other sources of data if possible and/ or neseccary.	Notify local businesses Create on open public call for local businesses to participate, to provide feedback
[illegible]	[illegible]	[illegible] 10 What do you think the challenges or complexities will be when using the PISTIS platform?	Ease of using the platform, ease of exporting appropriate datasets, AIA Ease of use of platform, Interoperability with airport internal systems	Usability of the platform Will the users need an expertise on data analysis/manageme nt in order to use it? any special skills?
[illegible]	[illegible]	[illegible] 11 What do you think should be done to guarantee that data are effectively exchanged by users within the PISTIS platform?	strong data contracts (KPIs, SLAs) Data safety	Safety and privacy requirements should be set and met for cyber-security by design Guarantee the safety of data flows exchange

Figure 10: Miro board from mobility hub workshop for UC 4-5

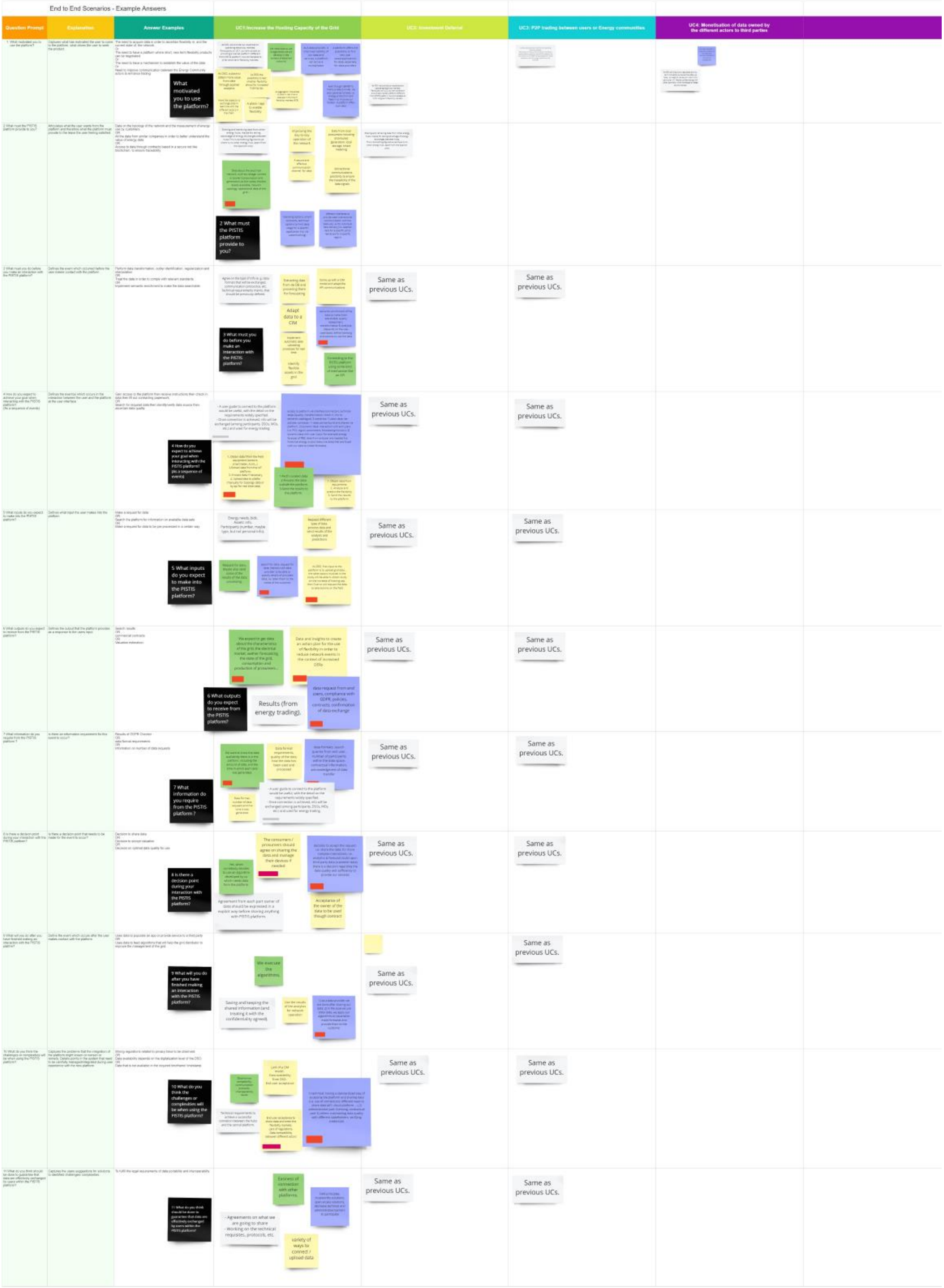

End to End Scenarios - Example Answers
Question Prompt
Explanation
Answer Examples
UC1:Increase the Hosting Capacity of the Grid
UC2: Investment Deferral
UC3: P2P trading between users or Energy communities
UC4: Monetisation of data owned by the different actors to third parties
What motivated you to use the platform?
2 What must the PISTIS platform provide to you?
3 What must you do before you make an interaction with the PISTIS platform?
4 How do you expect to achieve your goal when interacting with the PISTIS platform? (As a sequence of events)
5 What inputs do you expect to make into the PISTIS platform?
6 What outputs do you expect to receive from the PISTIS platform?
Results (from energy trading).
7 What information do you require from the PISTIS platform?
8 Is there a decision point during your interaction with the PISTIS platform?
9 What will you do after you have finished making an interaction with the PISTIS platform?
10 What do you think the challenges or complexities will be when using the PISTIS platform?
11 What do you think should be done to guarantee that data are effectively exchanged by users within the PISTIS platform?
Same as previous UCs.
Adapt data to a CIM

Figure 11: Miro board from energy hub workshop

Figure 12: Miro board from automotive hub workshop

APPENDIX B

Table 33 EDPB's nine criteria for high-risk data processing operations

No	Criteria	Description	Example
1	Use of Innovative Technologies	This refers to processing involving both the use of new technologies, and the novel application of existing technologies such as DL/ML models of AI technology. This is because the personal and social consequences of the use of such technology may be unknown and could result high risk for data subjects.	• Smart Mobility Systems. • Public Infrastructure Management based on ML/DL models.
2	Automated-Decisions with Legal or Similar Significant Effect	Fully automated decision-making procedure that has an impact on legal rights, legal status, social status, or the life standards of individuals.	• Entitlement to or denial of a particular social benefit or public services granted by law.
3	Systemic Monitoring	This type of processing covers organised or methodical tracking of individual's behaviour or geolocation in publicly accessible space including the online realm. It is often conducted according to a pre-arranged organised system and as part of a general plan for data collection.	• The use of a camera system to monitor driving behaviour on public roads. • A company systematically monitoring its employees' workstation.
4	Processing of Data with Highly Personal Nature	The concept of "data with highly personal nature" includes the special categories of personal data defined under Article 9(1) of the GDPR, the data relating to criminal convictions or offences and other categories of sensitive data which can be considered as increasing the possible threats to the rights and freedoms of individuals.	• A hospital processing its patients' genetic and health data. • The use of facial recognition systems.
5	Matching Datasets	This means combining, comparing, or matching personal data obtained from multiple sources or originating from two or more data processing operations performed for different purposes.	• Monitoring personal use/uptake of statutory services or benefits. • Automated fraud prevention techniques. • Federated identity assurance services.
6	Processing on a Large-Scale	When assessing whether the processing is performed on a large scale, the EDPB recommends data controllers to consider i) the number of data subjects concerned, either as a specific number or as a proportion of the relevant population; ii) the volume of data and/or the range of different data items being processed; iii) the duration, or permanence, of the data processing activity; iv) the geographical extent of the processing activity.	• Processing of travel data of individuals using a public transport system such as tracking via public transportation cards.
7	Evaluation or scoring	This type of processing covers profiling and the use of these profiles to make predictions about the future of individuals such as their performance at work, economic situation, health, personal	• Credit checks for mortgage applications.

		preferences or interests, reliability or behaviour, location, or movements.	• Risk assessment for insurance premiums. • Identification of high net-worth individuals for the purposes of direct marketing.
8	Data Concerning Vulnerable Individuals	Vulnerable individuals refer not only to a certain segment of the population that are traditionally deemed vulnerable such as children, mentally ill persons, asylum seekers, or the elders but also to those that are in disadvantageous position in their relationship with the data controller such as employees. Due to the increased imbalance between data controller and subject in the data processing within such contexts, it is presumed that data subject is often unable to knowingly and carefully oppose or consent to the processing or exercise their rights. Therefore, processing of the personal data of vulnerable individuals is deemed to be high-risk processing.	• Data collection through connected toys • Monitoring of the employees' online activities at work.
9	Denial of Service or Right	This includes processing activities including but not limited to the automated decision making that aims at allowing, modifying, or refusing data subjects' access to a product, service, opportunity, or entry into a contract.	• A credit institution screens its customers against a credit reference database to decide whether to offer them a loan