Ref. Ares(2025)9082 - 02/01/2025

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

# D1.3
# PISTIS Technical Requirements and MVP v2

| | |
|---|---|
| **Editor(s)** | Kyriakos Stefanidis, Christos Panagiotou |
| **Lead Beneficiary** | ATHENA |
| **Status** | Final |
| **Version** | 1.0 |
| **Due Date** | 30/11/2024 |
| **Delivery Date** | 02/01/2025 |
| **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 fr them.

| | |
|---|---|
| **Project** | PISTIS – 101093016 |
| **Work Package** | WP1 - PISTIS Trusted and Interoperable Data Trading and Management Framework |
| **Deliverable** | D1.3 - PISTIS Technical Requirements and MVP v2 |
| **Contributor(s)** | FHG, ATOS IT, ATOSES, IDC, IMPERIAL, ATHENA, ASSENT, SUITE5, UNIMALTA, EURECAT, DBL, UBITECH, ALEGAL, SPH, POLIMI, ICCS, UBIMET, OAG, AIA, GOLDAIR, DAEM, OASA, CUERVA, BAMBOO, OMIE, CARTIF, VIF, CARUSO, TRAF |
| **Reviewer(s)** | Anastasis Tzoumpas (UBITECH)<br><br>Francesco Alessandro Cuzzola (POLIMI) |
| **Abstract** | This deliverable identifies the minimum set of technical requirements of the PISTIS platform that realise the business needs and expectations of the end-users, through a user stories-oriented methodology for defining PISTIS MVP. |

# Executive Summary

Deliverable D1.3 serves as the updated version of the PISTIS Technical Requirements and MVP, building upon the initial results of D1.2 and reflecting progress made in the T1.5 task, "PISTIS User Stories, Technical Requirements, and MVP Design," within WP1, "PISTIS Trusted and Interoperable Data Trading and Management Framework." This deliverable refines and extends the technical foundations of the PISTIS platform to better address business needs, user expectations, and evolving technical challenges, while incorporating feedback from early implementation and pilot activities.

The scope of D1.3 is to enhance the previously defined Minimum Viable Product (MVP) design, which is the minimum set of requirements that the final version of the platform must follow, by addressing gaps identified during the initial development phases and by aligning with updated user requirements and technical constraints. The iterative process leveraged a user stories-driven methodology, incorporating continuous collaboration between end-users and developers to ensure practical applicability and alignment.

The methodology remains grounded in agile principles, expanding upon the four steps established in D1.2. These include refining stakeholder roles and personas, updating user stories with insights from the alpha version, reprioritizing technical requirements through MoSCoW analysis, and refining the MVP design to include advanced functionalities. This approach ensures that the deliverable reflects both stakeholder feedback and technological feasibility.

D1.3 presents refined user stories with fine-tuned functionalities and address scenarios observed during the alpha version. These refinements build upon the original 76 requirements from D1.2 and address critical feedback from stakeholders to optimize performance, security, and regulatory compliance. The existing 76 technical requirements have been expanded to 88 (with some new added some deleted as not applicable and other spitted or merged) for PISTIS platform. These technical requirements are specific for expressing the added value and benefits provided by PISTIS and doesn’t include classical information systems technical requirements, such as accessibility and standard security measures.

D1.3 now integrates insights from the Innovation Radar, which highlights emerging technologies and methodologies critical to enhancing the PISTIS platform's competitiveness. These include advancements in data monetization, secure data sharing, and advanced analytics that align with evolving trends in federated data ecosystems. These innovations ensure the platform remains adaptive to market demands and technological evolutions.

Additionally, the Generic Terms of PISTIS Data Sharing Contracts section introduces standardized contractual frameworks. These frameworks ensure that all data exchanges or co-creation activities within the PISTIS ecosystem are conducted transparently, securely, and in compliance with legal and regulatory requirements. They are designed to provide clarity and consistency for stakeholders, fostering trust and collaboration in data transactions.

# Table of Contents

1 Introduction ..... 9
1.1 Scope of The Deliverable ..... 9
1.2 Structure ..... 9
2 Methodology ..... 10
2.1 Actors Definition ..... 12
3 User Stories ..... 13
3.1 User Stories Identification ..... 13
3.2 User Stories Definition Template ..... 15
3.3 Operational User Stories ..... 17
3.3.1 Data check-in ..... 18
3.3.2 User registration and profile management ..... 21
3.3.3 Data transformation and analysis ..... 26
3.3.4 Data quality check ..... 28
3.3.5 Definition of access policies ..... 31
3.3.6 Monetization of the data ..... 33
3.3.7 Data Valuation ..... 35
3.3.8 Data usage and market analytics ..... 37
3.3.9 Data Query ..... 39
3.3.10 Data Transaction ..... 41
3.3.11 Data exchange monitor/audit ..... 43
3.4 Support Operations User Stories ..... 46
3.4.1 PISTIS platform monitor/audit ..... 46
3.4.2 PISTIS platform configuration ..... 48
3.4.3 Data Factory Configuration/Setup ..... 50
3.4.4 Data Pipeline Definition, Execution and Fetching ..... 52
4 Technical Requirements ..... 55
5 MVP Design ..... 64
6 Generic Terms of PISTIS Data Sharing Contracts ..... 71
7 Technology Radar ..... 74
7.1 Data Management ..... 74
7.1.1 Initiatives related to ETL ..... 74
7.1.2 Initiatives related to Privacy, Security and Trust ..... 77
7.2 Data Trading ..... 80
7.2.1 Initiatives related to Distributed Ledgers ..... 80
7.2.2 Initiatives related to Data Monetization ..... 82

7.2.3 Data Sources and Services ..... 85
7.3 Technology Radar Summary ..... 94
8 Conclusion ..... 95
Annex 1 – PISTIS Generic Terms of Data Sharing ..... 96
Annex 2 – PISTIS Generic Terms of Data Sharing and Joint Dataset Creation Agreement ..... 98

# List of Figures

Figure 1: PISTIS Technical Requirements and MVP Definition Methodology ..... 10
Figure 2: Data lifecycle from the involved actors' perspective ..... 14
Figure 3: Actors and PISTIS high-level interactions ..... 14
Figure 4: Data check-in process ..... 19
Figure 5: User account creation process ..... 22
Figure 6: User profile self-management ..... 23
Figure 7: Process of PISTIS Organization Administrator manages a user's profile. ..... 24
Figure 8: Data transformation and analysis ..... 26
Figure 9: Data Transformation Execution Sub-process ..... 27
Figure 10: Data quality check process ..... 29
Figure 11: Definition of access policies ..... 32
Figure 12: Monetization of the data ..... 34
Figure 13: Data Valuation ..... 36
Figure 14: Data usage and market analytics ..... 38
Figure 15: Data Query ..... 40
Figure 16: Data Transaction ..... 42
Figure 17: Monitor Specific Dataset Exchange process. ..... 44
Figure 18: Auditing History process. ..... 45
Figure 19: PISTIS platform monitor/audit process ..... 47
Figure 20: PISTIS platform configuration ..... 49
Figure 21: Data Factory Configuration/Setup ..... 51
Figure 22. PISTIS Data pipeline definition and execution ..... 53
Figure 23: Positioning the Smart Contract Checker within PISTIS architecture ..... 71

# List of Tables

Table 1: List of PISTIS User Stories ..... 15
Table 2: User stories template ..... 16
Table 3: Data check-in User Story ..... 18
Table 4: User registration and profile management User Story ..... 21
Table 5: Data transformation and analysis User Story ..... 26
Table 6: Data quality check User Story ..... 28
Table 7: Definition of access policies User Story ..... 31
Table 8: Monetization of the data User Story ..... 33
Table 9: Data Valuation User Story ..... 35
Table 10: Data usage and market analytics User Story ..... 37
Table 11: Data Query User Story ..... 39
Table 12: Data Transaction User Story ..... 41
Table 13: Data exchange monitor/audit User Story ..... 43
Table 14: PISTIS platform monitor/audit User Story ..... 46
Table 15: PISTIS platform configuration User Story ..... 48
Table 16: Data Factory Configuration/Setup User Story ..... 50
Table 17: PISTIS Technical Requirements ..... 55
Table 18: Requirements Prioritising Vote Results ..... 65
Table 19: PISTIS MVP Requirements ..... 66
Table 20: Initiatives related to ETL ..... 74
Table 21: Initiatives related to privacy, security and trust ..... 77
Table 22: Initiatives related to distributed ledgers ..... 80
Table 23: Initiatives related to data monetization ..... 83
Table 24: Initiatives related to data sources and services ..... 86

# Terms and Abbreviations

| | |
|---|---|
| **AI** | Artificial Intelligence |
| **BPMN** | Business Process Model and Notation |
| **CA** | Certification Authority |
| **DLT** | Distributed Ledger Technology |
| **DSSE** | Dynamic Symmetric Searchable Encryption |
| **DQA** | Data Quality Assessment |
| **FTP** | File Transfer Protocol |
| **IAM** | Identity and Access Management |
| **JSON** | JavaScript Object Notation |
| **MoSCoW** | M - Must have, S - Should have, C - Could have, W - Won't have. |
| **MVP** | Minimum viable product |
| **NFT** | Non-Fungible Token |
| **RDF** | Resource Description Framework |
| **XML** | Extensible Markup Language |
| **eIDAS** | electronic IDentification And trust Services |

# 1 INTRODUCTION

## 1.1 SCOPE OF THE DELIVERABLE

This deliverable, D1.3 - PISTIS Technical Requirements and MVP v2, reflects the progression in the development of the PISTIS platform, building on the foundational work presented in D1.2. The progress of the project, marked by the release of the alpha version of the platform, has provided insights into the practical implementation and operational requirements of PISTIS. These advancements have informed the updates and refinements contained within this deliverable.

The PISTIS project is centered on fostering a federated and trusted ecosystem for secure, fair, and efficient data sharing, trading, and monetization. The release of the alpha version served as a technical milestone enabling the consortium to assess the initial realization of the platform's core functionalities, such as federated data management, secure data sharing, and monetization.

Deliverable D1.3 delves deeper into the technical and operational details of the PISTIS platform, presenting a more mature and comprehensive set of requirements. These updates are underpinned by a systematic methodology that incorporates lessons learned during the alpha phase. The enhanced approach places greater emphasis on scalability, interoperability, and user-centric design, ensuring the platform is both robust and adaptable to the dynamic needs of its ecosystem.

Central to this deliverable is the continued use of user stories as a tool for bridging the gap between conceptual design and practical implementation. This narrative-driven approach not only provides a shared understanding among stakeholders but also delivers a structured framework for prioritizing technical requirements.

The new "Technology Radar" section, provides a comprehensive overview of emerging technological trends relevant to the PISTIS platform ensuring that the platform remains at the forefront of innovation. Meanwhile, the Generic Terms of PISTIS Data Sharing Contracts section introduces standardized contractual frameworks, offering guidelines for ensuring fairness, legal compliance, and transparency in data transactions within the PISTIS ecosystem.

Furthermore, the Generic Terms of PISTIS Data Sharing Contracts section establishes the scope for standardized contractual frameworks that govern all data-sharing activities within the ecosystem. These frameworks outline clear principles and guidelines to ensure fairness, legal compliance, and operational transparency. By addressing both the technical and legal dimensions of data transactions, they create a structured and reliable foundation for managing data exchange and collaboration.

As the PISTIS progresses, this deliverable, as a revised version of D1.2, serves as a roadmap, weaving together the threads of user stories, technical requirements, and strategic prioritization to guide the PISTIS project implementation towards the successful realization of an impactful MVP. It should be noted that **MVP design** in the context of the work of WP1 and deliverables D1.2 and D1.3, refers to the **minimum set of requirements** for the envisioned PISTIS framework that is both desirable and operational for the end users and can be addressed during the project implementation **in the final delivered version** of the PISTIS platform.

## 1.2 STRUCTURE

The document is organized into several sections that collectively address the development, implementation, and refinement of the PISTIS platform. It begins with an introduction that establishes the scope, purpose, and objectives of the deliverable, highlighting updates and enhancements made

based on insights from the alpha version. This section also discusses the use of user stories to bridge the gap between user needs and technical requirements and introduces standardized contractual frameworks and a Technology Radar section to align the platform with emerging trends. The methodology section describes the approach used to define the PISTIS Minimum Viable Product (MVP). It includes stakeholder identification, enhancement of user stories, refinement of technical requirements, and prioritization of these requirements using the MoSCoW method to ensure alignment with end-user and technical needs. The user stories section elaborates on the narratives that capture user needs and expectations, providing a clear framework for implementing core functionalities. It includes operational user stories that detail key processes and support operation user stories that focus on administrative and maintenance tasks. The technical requirements section outlines the specific needs derived from the user stories, addressing both business and technical challenges. These requirements have been refined and expanded based on feedback from the alpha phase and are categorized for clarity. The MVP design section defines the core features and functionalities essential for the PISTIS platform. It details the prioritization process using the MoSCoW method and presents the final set of MVP requirements that guide the development process. The generic terms of PISTIS data sharing contracts section introduces contractual frameworks designed to ensure fairness, compliance, and transparency in data sharing and collaboration within the PISTIS ecosystem. These frameworks support secure and standardized transactions between stakeholders. The Technology Radar section analyzes external initiatives and emerging technologies that align with the goals of the PISTIS platform. This includes areas such as data management, privacy, security, trust, distributed ledgers, and data monetization, providing insights into how the platform can leverage these trends. The document concludes by summarizing its key outcomes, aligning the deliverable's contributions with the overarching objectives of the PISTIS project and providing a roadmap for future development. Annexes provide additional details, such as generic terms for data sharing and agreements for joint dataset creation. This comprehensive structure ensures the deliverable effectively addresses the platform's requirements and implementation strategy.

# 2 METHODOLOGY

The methodology adopted for defining the Minimum Viable Product (MVP) in D1.3 builds on the framework established in D1.2, incorporating enhancements informed by the progress of the project and feedback from stakeholders (Figure 1). As the PISTIS alpha version is now operational, the methodology integrates findings from its implementation to refine the technical requirements and ensure their alignment with the stakeholders' needs. The updated approach comprises the following key steps:

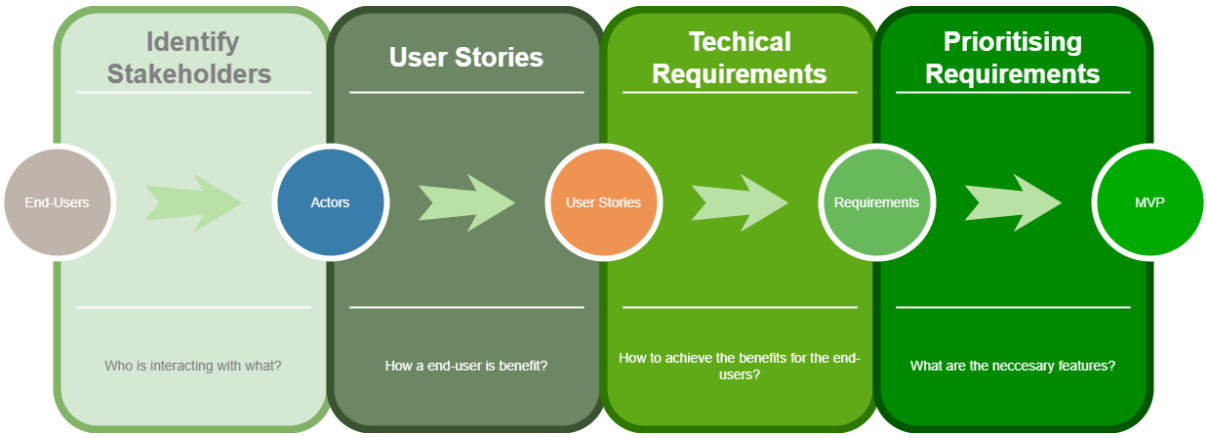



**Figure 1: PISTIS Technical Requirements and MVP Definition Methodology**

The technical requirements and MVP definition methodology comprised of the following steps:

1. **Identify Stakeholders and User Personas**: The identification of stakeholders and user personas remains foundational, with renewed efforts to capture insights from the alpha release. This iteration incorporates feedback from early adopters and pilot participants, ensuring that the evolving requirements reflect both technical feasibility and user expectations.
2. **User Stories enhancement**: User stories have been revisited and fine-tuned based on the lessons learned from the alpha version. Adjustments have been made to align more closely with operational challenges and expanded platform functionalities while preserving their original narrative structure of "As a [user persona], I want [an action] so that [benefit/value]." These updates maintain the narrative format of "As a [user persona], I want [an action] so that [benefit/value]," while introducing refinements to address identified gaps.
3. **Technical Requirements refinement**: The definition of the user stories, apart from the benefits for the users, also included the technical and procedural challenges that must be faced by the PISTIS. Thus, from each user story, a list of technical requirements was defined. The technical and procedural challenges identified during the alpha release have been systematically analyzed to refine the technical requirements.
4. **Prioritization of Requirements**: The final step was the prioritization of the technical requirements towards the identification of which of them had to be met in the PISTIS MVP. This led to the MoSCoW (Must-haves, Should-haves, Could-haves, Won't-haves)[1] after a weighted voting including participants from both groups of end-users and developers. Prioritisation was primarily based on business value, dependencies and customer needs, but also considered the technical solutions needed to deliver the solutions.

   Consider the technical requirement REQ_65: PISTIS supports secure data encryption on data transactions. This requirement ensures that all data exchanges within the PISTIS platform are protected using robust encryption protocols, safeguarding user data during transactions. During the prioritization process, stakeholders from two groups — end-users and technical experts — were asked to evaluate the requirement's importance. The group, consisting of partners operating the PISTIS platform emphasized the necessity of secure data encryption to meet regulatory compliance, enhance user trust, and mitigate risks associated with data breaches. They assigned high priority (4, "Must-Have") to this requirement, citing its critical role in ensuring the platform's credibility and operational success. The technical group, composed of developers and architects, assessed the feasibility of implementing this requirement. They acknowledged the importance of secure encryption but also considered the technical complexity and integration challenges. Nevertheless, they agreed that without robust encryption mechanisms, the platform's functionality and market adoption would be compromised. Consequently, they also assigned a high priority (4, "Must-Have") to this requirement. Using the weighted MoSCoW formula, where end-user votes were assigned a weight of 1.2 and technical group votes were weighted at 0.8, the prioritization score was calculated as follows:

$$\frac{(1.2 \times 4 \times N) + (0.8 \times 4 \times M)}{N + M}$$

   Here, $N$ and $M$ represent the number of participants in the end-user and technical groups, respectively. Assuming equal participation, the calculated score reaffirmed the requirement as a "Must-Have.". This example illustrates how collaborative input and structured

[1] R. Popli, N. Chauhan and H. Sharma, "Prioritising user stories in agile environment," 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 2014, pp. 515-519, doi: 10.1109/ICICICT.2014.6781336.

prioritization helped identify REQ_65 as essential for the PISTIS MVP, ensuring alignment between business needs and technical feasibility.

In conclusion, the refined definition of the PISTIS MVP in D1.3 updates the prioritization of the technical requirements derived from both the earlier deliverable and new insights gained during the alpha phase of development. This process has resulted in an updated categorization of the requirements presented in D1.2, enhanced with the newly identified in D1.3. These requirements are classified into must-haves and should-haves, serving as foundational inputs for the subsequent work packages WP2, WP3, and WP4.

## 2.1 ACTORS DEFINITION

To ensure clarity and alignment in the interactions between users and the PISTIS system, a precise definition of actors and their roles has already been established in D1.2. This section remains unchanged from D1.2 as the content accurately reflects the necessary definitions of actors and their roles within the PISTIS platform. Since no significant updates or modifications were required, the descriptions provided in D1.2 continue to serve their intended purpose effectively in D1.3. This consistency ensures alignment with previously defined interactions and roles, maintaining coherence throughout the deliverable.

The actor definition step is fundamental for articulating user stories that accurately capture the operational needs and expectations of the platform's users. Rather than focusing on the internal communication between system modules, the user stories emphasize the external interactions of users with the PISTIS system, providing actionable insights into user-centric functionalities.

The user stories are focused on determining the interaction of the users with the PISTIS system rather than analysing the internal communication and data exchange between PISTIS system's functional modules. Nevertheless, as the users are familiar with the PISTIS conceptual architecture and the main components, they can express in a high-level conceptualisation their opinion about how they can execute some of the functionalities to help the technology partners to understand better their needs. In the user stories **PISTIS Platform is the main actor** representing the whole PISTIS solution. It can be seen as the complex system including all the functional modules, regardless their topology and technology.

However, since end-users already have a familiarity with the conceptual architecture of PISTIS outlined in the project proposal, as well as the basic functionality of its main components, the decision was made to incorporate references to specific modules in the user stories. This inclusion of references to PISTIS modules serves to aid both end-users and developers in comprehending how a process will be executed with PISTIS, considering the distinct features and roles of certain functional components. The functional components that are referred to the user stories are:

- The **PISTIS Data Factory** module is an essential module of the PISTIS conceptual architecture. It is one of the two main modules of PISTIS. The Data Factory is an application which contains an organization's data, connects with the other Data Factories to execute data exchange transactions and it is fully managed by the owner organization.
- The **PISTIS Central Platform** is the second essential module of PISTIS. It is the centralized system that gathers the necessary information from the PISTIS data factories and builds an ecosystem for data publishing, monetization and financial transactions.
- The **PISTIS Identity and Access Manager (IAM)** is the module designed to ensure secure and organized access to PISTIS digital resources (data, users, etc.). It revolves around managing and controlling the identities of individuals or entities within a system, determining what actions

they are authorized to perform, and monitoring their activities. IAM systems play a crucial role in safeguarding sensitive information, applications, and systems by enforcing policies and permissions. This includes authentication processes to verify the identity of users, authorization mechanisms to grant appropriate access levels, and identity lifecycle management to handle changes in user status. The **PISTIS Identity and Access Manager** is an online platform that facilitates the buying and selling of data assets between multiple parties. As a data asset is defined a monetized dataset published for sale.

By analysing the business requirements as well as the scope of the PISTIS platform, different roles of users have been identified, based on their utilisation of the platform. Drawing on this, the PISTIS User entity used in the user stories can act in different roles according to the functional needs. A preliminary set of these roles is the following:

- The **Data Seller/Provider** is a part of a specific organisation and can manage the datasets (data and metadata) registered in the PISTIS platform, particular in the organisation's PISTIS Data Factory. Users acting with this role can also manage organisation assets in PISTIS marketplace including their monetisation and publication.
- The **Data Buyer/Consumer** is also a user belonging to an organisation acting as a potential buyer in the PISTIS ecosystem. Data Consumers can search for data assets to buy and complete financial deals with the Data Providers that will conclude to the acquisition of the data.
- The **Organisation Administrator** is managing the resources (users and datasets) of an organisation's PISTIS Data Factory.
- The **PISTIS Administrator** is responsible for managing the functionalities and resources of the PISTIS Central Platform, such as the management of the connected PISTIS Data Factories.
- The **Auditor** is an external or internal entity (maybe an administrator) who is responsible for auditing the behaviour of the systems or sub-systems of PISTIS framework.

Moreover, in the stories there an independent entity (system or service), the **Certificate Authority (CA)**, which is a trusted entity that issues digital certificates. The Certificate Authority plays a crucial role in the implementation of secure communication over the internet by verifying the authenticity of entities involved in a transaction, such as systems and users.

# 3 USER STORIES

To attain the ultimate objective of Task 1.5, which is to define the PISTIS MVP, the initial step involves bridging the comprehension gap between end-users and developers. To accomplish this, the User Stories tool has been chosen to articulate the functionalities of the PISTIS platform based on user needs and experiences. User stories are concise, easy-to-understand narratives that capture the "who," "what," and "why" of a feature or requirement. Furthermore, they provide a clear and simple way to communicate requirements by using natural language and avoiding technical jargon. This section presents the user stories for the PISTIS platform which will act as a bridge between the end-users and the developers, helping to align the PISTIS development process with user needs and expectations.

## 3.1 USER STORIES IDENTIFICATION

The identification of User Stories that will be analysed is based on the discussion made with the end-users for extracting their business needs. As stated in D1.1 PISTIS Operation Principles and Context Detailing, the end-users perceive the PISTIS solution as a platform that offers specific operations aligned with the data lifecycle illustrated in Figure 2.

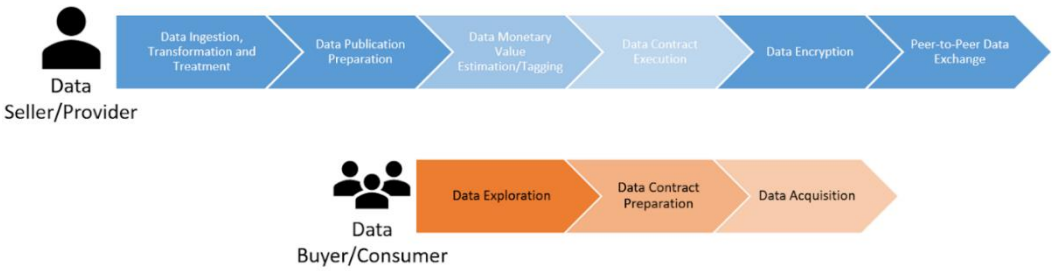



**Figure 2: Data lifecycle from the involved actors' perspective**[2]

The PISTIS Data lifecycle is defined for two major actors, the Data Seller/Provider and Data Buyer/Consumer. In the effort to determine the interactions of the users with PISTIS platform, a new group of users has been established, the administrators, which provide supportive operations to the systems of PISTIS framework. Figure 3 depicts the interactions in high level between actors and PISTIS.

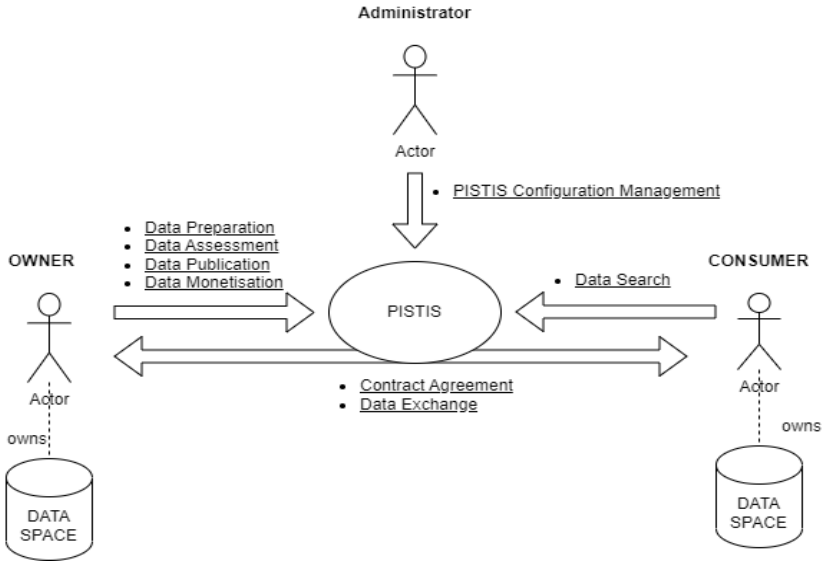



**Figure 3: Actors and PISTIS high-level interactions**

Derived from these interactions that articulate the core functionalities of PISTIS and taking into account general technological aspects of information systems, a list of potential user stories (Table 1) has been proposed. This includes the following:

**Operational User Stories** describing the basic functionalities provided to business end-users:

- User registration and profile management
- Data check-in
- Data transformation and analysis
- Data quality check
- Definition of access policies
- Monetization of the data
- Data Valuation
- Data usage and market analytics

[2] PISTIS Consortium, D1.1 PISTIS Operation Principles and Context Detailing, June 2023.

- Published Data management.
- Advanced Data Query
- Smart contract creation
- Data Exchange Initialisation
- Data exchange monitor/audit

**Supporting User Stories** describing mostly tasks of systems administrators:

- PISTIS platform monitor/audit
- PISTIS platform configuration
- Data Space Factory Configuration/Setup

The final step to the User Stories selection was a fine tuning of the suggested stories based on the following criteria:

1. They must meet all the business requirements as depicted in D1.1.
2. If they are including duplicate procedures, then they must be merged.
3. If they are very complex, they must be separated.
4. If a user story is very simple and part of another, they must be merged.

The final list of User Stories along with the business requirements they are dealing with, especially for the Operational User Stories is the following:

**Table 1: List of PISTIS User Stories**

| ID | User Story | Business Requirements (D1.1) |
|---|---|---|
| **OPERATIONAL USER STORIES** | | |
| **PISTIS.OUS.01** | Data check-in | BR02, BR09, BR11 |
| **PISTIS.OUS.02** | User registration and profile management | BR01, BR17 |
| **PISTIS.OUS.03** | Data transformation and analysis | BR05, BR11 |
| **PISTIS.OUS.04** | Data quality check | BR05, BR06, BR11 |
| **PISTIS.OUS.05** | Definition of access policies | BR15, BR16 |
| **PISTIS.OUS.06** | Monetization of the data | BR07, BR12, BR14, BR15 |
| **PISTIS.OUS.07** | Data Valuation | BR07 |
| **PISTIS.OUS.08** | Data usage and market analytics | BR07, BR12 |
| **PISTIS.OUS.09** | Data Query | BR03, BR04 |
| **PISTIS.OUS.10** | Data Transaction | BR07, BR08, BR09, BR10, BR14 |
| **PISTIS.OUS.11** | Data exchange monitor/audit | BR13, BR15 |
| **SUPPORT OPERATIONS USER STORIES** | | |
| **PISTIS.SOUS.01** | PISTIS platform monitor/audit | BR12, BR15, BR16, BR17 |
| **PISTIS.SOUS.02** | PISTIS platform configuration | BR01, BR15, BR16, BR17 |
| **PISTIS.SOUS.03** | Data Factory Configuration/Setup | BR02, BR11, BR16, BR17 |

## 3.2 User Stories Definition Template

To enhance the clarity and consensus in defining PISTIS functionalities from the user's perspective while ensuring understanding by technology partners, a blend of natural language description and formalism was chosen for PISTIS User Stories. These user stories adhere to the template outlined in Table 2. Each user story includes a precise definition of involved actors (human or system), a detailed

story description encompassing triggering events or conditions, and the benefits for users. The initial fields follow the widely used "As a [user persona], I want [an action] so that [benefit/value]" template.

Furthermore, each story is depicted as a process showing the interaction between the actors and the operations of the latter. The Business Process Model and Notation [3] (BPMN) was utilized as a standardized graphical notation that is used to create visual representations of business processes . BPMN provides a standardized and visually intuitive way to represent complex business processes. By using standard notation, BPMN ensures consistency in process documentation across end-users and developers. This consistency is crucial for clear communication and understanding. In summary, BPMN enhances procedure management by providing a standardized, visual, and communicative way to document, analyse, and improve business processes. Besides, BPMN contributes to organizational efficiency, transparency, and adaptability in an ever-changing business environment.

The BPMN-based definition of a user story allows the identification of technical challenges and issues that must be considered for the realization of the indicated process by PISTIS framework. These concepts would be used later in the definition of the technical requirements.

**Table 2: User stories template**

| **Story ID** | ***A unique identifier of the user story*** |
|---|---|
| **Story Title** | ***The title of the story*** |
| **Story Actors** | *PISTIS Administrator/PISTIS Platform/PISTIS User (Provider)/ PISTIS User (Consumer)/etc* |
| **Overview** | *A brief description of the story* |
| **Triggers** | *The motivation, action or event that triggers this story* |
| **Workflow** | *A graphical representation of the workflow (process) of the story. A BPMN (Business Process Model Notation) is preferable to be used here.* |
| **Story Sequence** | *A detailed depiction of the story in steps. Each step and sub-step are specifically numbered (1,2,3, 3.1, 3.2, ….)* |
| **User Benefits** | *Define, in bullets, the operations or functionalities that affect the user during this procedure.* |
| **Challenges** | *Define technical or business challenges that must be considered as constrains or requirements for the execution of the story* |
| **Security Issues** | *Enlist security related issues* |
| **Other issues** | *Any other issues or constrains that must be considered for the execution of the story.* |

[3]Object Management Group (OMG), Business Process Model And Notation, OMG specification, https://www.omg.org/spec/BPMN/2.0/

## 3.3 OPERATIONAL USER STORIES

In D1.3, the operational user stories have been significantly refined and expanded compared to D1.2, incorporating features and processes aimed at addressing practical challenges and enhancing platform capabilities based on feedback from the alpha phase.

For the Data Check-In story, new requirements were introduced to ensure robust data security and access management. Ratified datasets are now encrypted using advanced, hardware-rooted key management techniques provided by the Encryption/Decryption engine. This ensures confidential data is securely stored at rest. Additionally, the system now automatically generates default access policies during the check-in process, providing baseline protections for datasets. Users also have the option to override these default policies, allowing them to modify or exclude specific access criteria for greater flexibility.

In User Registration and Profile Management, integration with external identity providers such as eIDAS and other government-issued digital IDs has been added to streamline the user onboarding process. The workflow for updating user profiles has been streamlined to improve usability and reduce the complexity of managing account details.

The Data Transformation and Analysis story now supports schema evolution, allowing transformation workflows to adapt to changes in data structures dynamically. This ensures greater compatibility and flexibility when managing diverse datasets.

The Data Quality Check story has been enhanced to include lineage tracking, ensuring traceability throughout the quality check process. Quality-related metadata is now incorporated into the results, offering a richer understanding of dataset integrity. Reporting capabilities have been improved to provide more detailed insights into quality metrics and issues, while the workflow has been modified to dynamically handle incomplete metadata, ensuring the process remains adaptable and comprehensive.

For the Definition of Access Policies, attribute-level access controls have been introduced to enable fine-grained management of permissions. Dynamic policy enforcement has also been added, allowing access controls to adapt based on user behavior or specific contextual requirements. The workflows for defining and managing access policies have been simplified to reduce administrative overhead, and the policy conditions now include more granular options for managing dataset access.

The Monetization of Data story has seen the introduction of advanced pricing models, including subscription-based and pay-per-use options, enabling users to implement more flexible monetization strategies. Support for bundling datasets into customizable packages has been added to meet diverse market needs. Transaction templates have been modified to support a broader range of monetization strategies, making the process more adaptable to user requirements.

For Data Valuation, new guidelines have been introduced to define valuation criteria such as data quality, relevance, and utility. These updates ensure that valuation processes are clearer and more actionable for end-users.

In the Data Query story, a significant addition is the support for querying encrypted data using Dynamic Symmetric Searchable Encryption (DSSE). This functionality allows users to characterize encrypted data with metadata, enabling secure and efficient querying while maintaining data confidentiality.

The Data Transaction story has undergone multiple updates to enhance the exchange process. A Data Factory Connector has been introduced to facilitate seamless communication between the Data

Factories of two PISTIS users. Users' wallets now support hardware-based identity keys for enhanced security. A portion of the contract value is allocated as fees for the platform, providing a sustainable operational model. Additionally, the dataset's metadata is transferred from the market to the consumer's Data Factory during the transaction, ensuring that the buyer receives complete and actionable information.

These updates collectively enhance the operational user stories, ensuring that D1.3 reflects a more robust, scalable, and user-centered approach compared to D1.2. The changes address critical security, usability, and marketability concerns, aligning the platform with real-world use cases and stakeholder feedback.

The following sections depict the user stories related to the operations carried out to deliver the main functionalities of the PISTIS Platform.

### 3.3.1 Data check-in

**Table 3: Data check-in User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.01** |
| **Story Title** | **Data check-in** |
| **Story Actors** | PISTIS User (Provider), PISTIS Platform |
| **Overview** | The PISTIS User (Provider) wants to register a portion of data (dataset) to the PISTIS Platform. The user must point out the data to the PISTIS platform, which will select the appropriate way to access the data (uploaded file, FTP, API, etc.). The PISTIS Platform (Data Space Factory) will receive the data source access details and the data's metadata. The data source details will be used to access the data and inject them into the PISTIS platform. The data's metadata will be used to add the new dataset to the lineage tracker and data catalogue. Finally, the PISTIS platform will orchestrate the data check-in process and notify the user when it is finished. |
| **Triggers** | A PISTIS User (Provider) that has already installed PISTIS Data Space Factory in his/her owned data space wants to add a new portion of data to the PISTIS ecosystem. |

**Workflow**

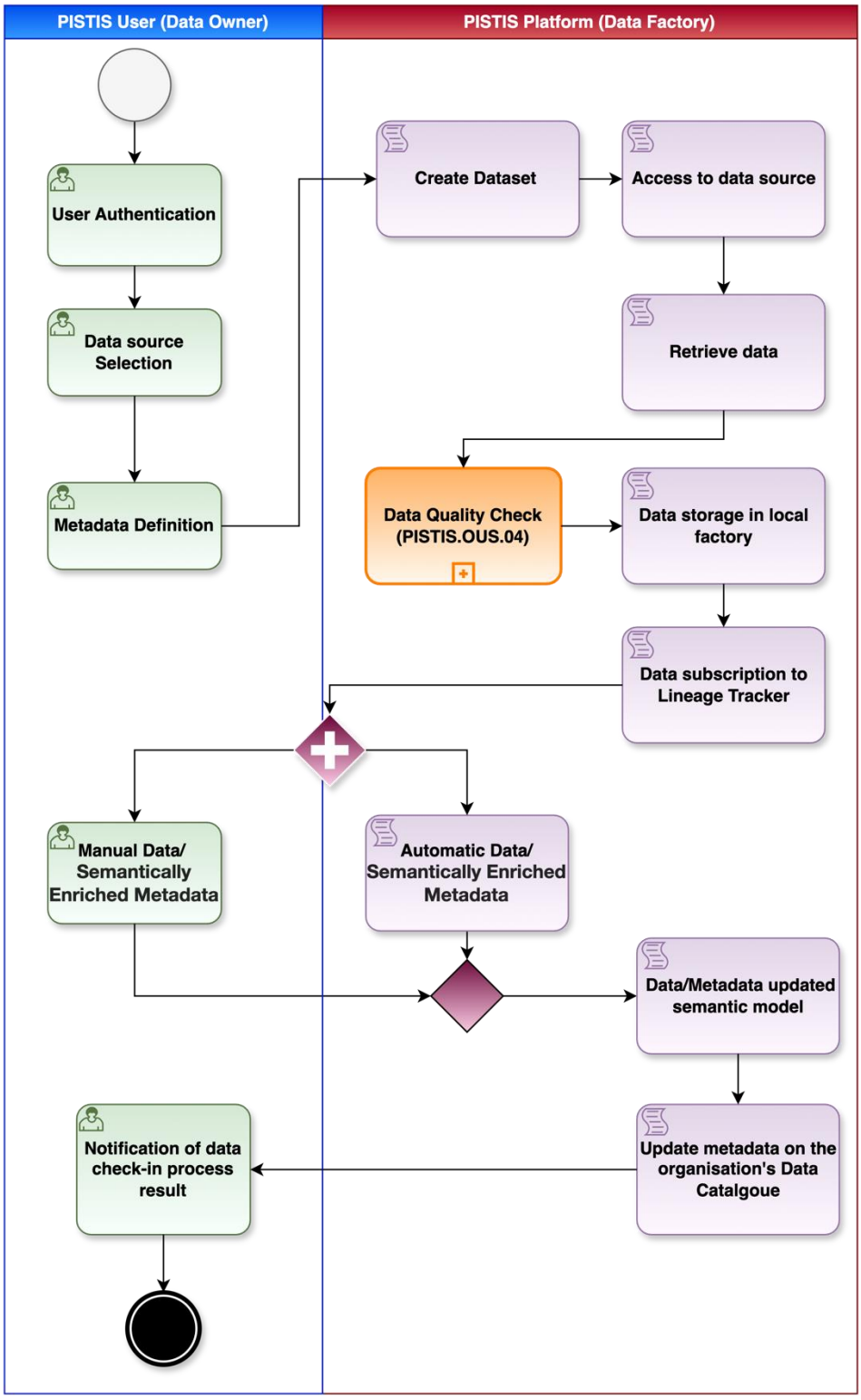



Figure 4: Data check-in process

**Story Sequence**

1. PISTIS User is authenticated by the PISTIS platform.
2. PISTIS User selects the data that wants to check-in. The selection includes the specification of data source (uploaded file, FTP, API, etc.).
3. PISTIS User defines data's metadata by completing a form.
4. PISTIS Platform creates a local dataset.

| | |
|---|---|
| | 5. PISTIS Platform accesses the data source according to the parameters defined in step 2.<br>6. PISTIS Platform retrieves data.<br>7. The data quality check sub-process is executed in the retrieved data (PISTIS.OUS.04).<br>8. Based on the output of the Data Quality check, ratified data is then encrypted for confidential data at rest management. Encryption is leveraging advanced (Hardware-Rooted) key management techniques offered by Encryption/Decryption engine.<br>9. PISTIS Platform registers the new dataset to lineage tracker.<br>10. PISTIS Platform semantically enriches data/metadata. It is done in two ways:<br>10.1. Manually by the user.<br>10.2. Automatically by analysing available metadata and especially the actual data.<br>11. PISTIS Platform the registers the new dataset to the organisation's local data catalogue.<br>12. PISTIS Platform creates default access policies (automatically generated) and overrides (user defined)<br>13. PISTIS Platform notifies PISTIS User that data check-in process is finished. |
| **User Benefits** | The user will be able to:<br>• Define the dataset that will be checked in.<br>• Define the metadata of the new dataset.<br>• Inject dataset into PISTIS Platform (internal storage).<br>• Be informed of possible errors during the procedure<br>• Redefine correct metadata in case of error.<br>• Override default Access Policies of the dataset (exclude criteria)<br>• Be informed of successful end of the procedure. |
| **Challenges** | • The data can be accessed/stored in various types of sources, i.e. HTTP APIs, FTP, storage volumes, cloud storages, etc.<br>• The data doesn't follow a specific format or structure.<br>• The metadata can be described in various description languages and formats, i.e. JSON, XML, TXT, etc., a specific metadata model is needed. |
| **Security Issues** | • Confidentiality and Authentication: Only the correct user, as the provider of the data, should be able to have access to perform this task.<br>• Accountability and non-repudiation of the user's actions regarding providing the necessary privileges to the PISTIS Platform to selected data sources. This will be achieved through the provision of appropriate signature mechanisms for verifying the correctness of user's actions. |
| **Other issues** | - |

### 3.3.2 User registration and profile management

**Table 4: User registration and profile management User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.02** |
| **Story Title** | **User registration and profile management** |
| **Story Actors** | PISTIS Platform, PISTIS User (Organization Administrator), PISTIS User, Trusted CA |
| **Overview** | A new PISTIS User wants to be part of PISTIS Platform. The administrator of the PISTIS Organization where User belongs to, manually registers the user either by creating an account within PISTIS Identity Access Manager or by using external Identity Providers (e.g., Google, eIDAS). PISTIS Organization Administrator is responsible for activating/deactivating User accounts. User is offered with optional preferences, i.e. two-factor authentication (2FA) for the account.<br><br>Additionally, a Digital DLT-FIAT Wallet is created for the newly designated user.<br><br>Each user who wants to use PISTIS Platform with an external Identity Provider will initially need to register with a Trusted CA (for example an eIDAS Trusted Provider) and obtain issued verifiable credentials. These verifiable credentials will then have to be linked to the HW-Based identity key of their identity wallet. This will also be propagated to their monetary wallet to extend the binding process to also associate a DLT Wallet address to the user's credentials.<br><br>At any time, the user will be able to manage their profile, within PISTIS Identity Access Manager, for the newly created account for the PISTIS platform, i.e., manage secondary details and 2FA preference.<br><br>PISTIS Organization Administrator will be privileged with the right to disable (or remove) the account, change its access right by modifying its set of roles. |
| **Triggers** | This story is divided in three sub-stories based on the functionality provided by the PISTIS Platform concerning the management of user's profiles. Each sub-story is initiated by different events.<br><br>1. User account creation: A PISTIS User account is created by a PISTIS Organization Administrator.<br>2. User profile self-management: An existing PISTIS User wants to manage his/her profile.<br>3. PISTIS Organization Administrator manages a user's profile: PISTIS Organization Administrator wants to alter an existing PISTIS User's profile. |
| **Workflow** | **User creation** |

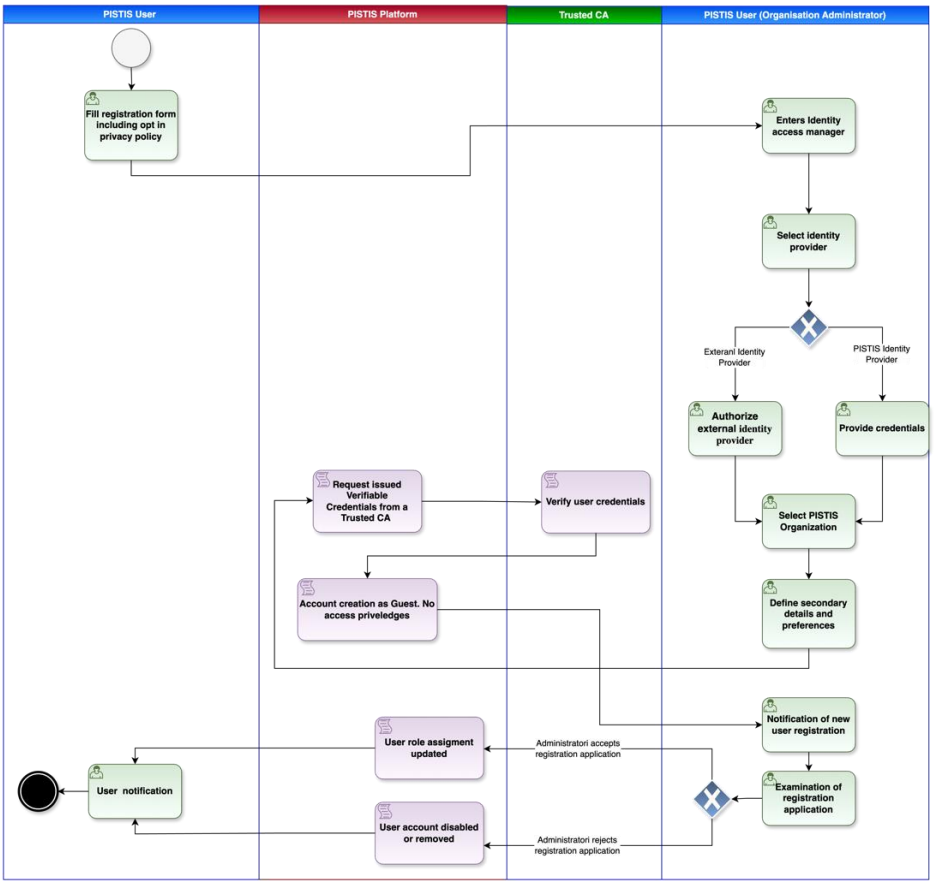



Figure 5: User account creation process

**User profile self-management**

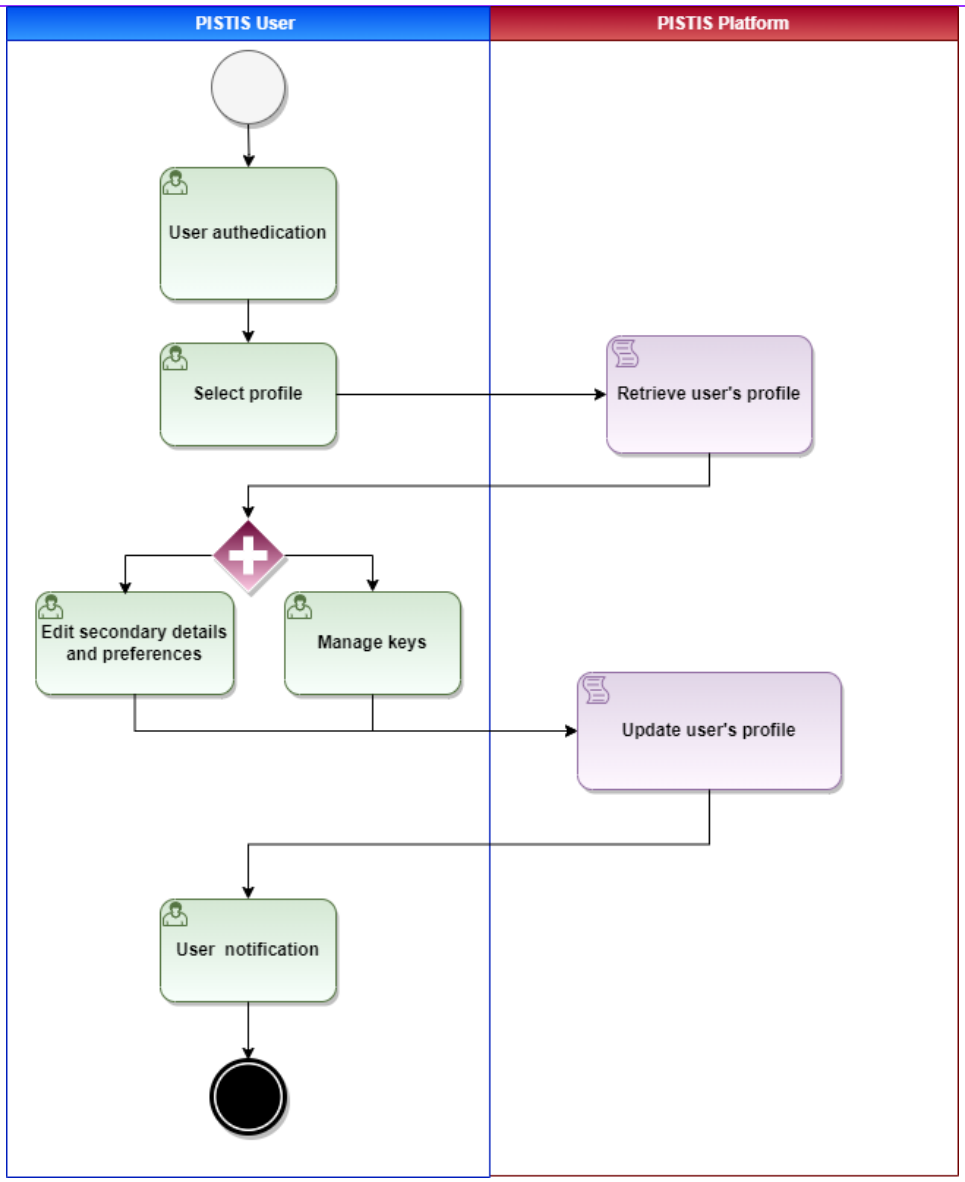



Figure 6: User profile self-management

**PISTIS Organization Administrator manages a user's profile.**

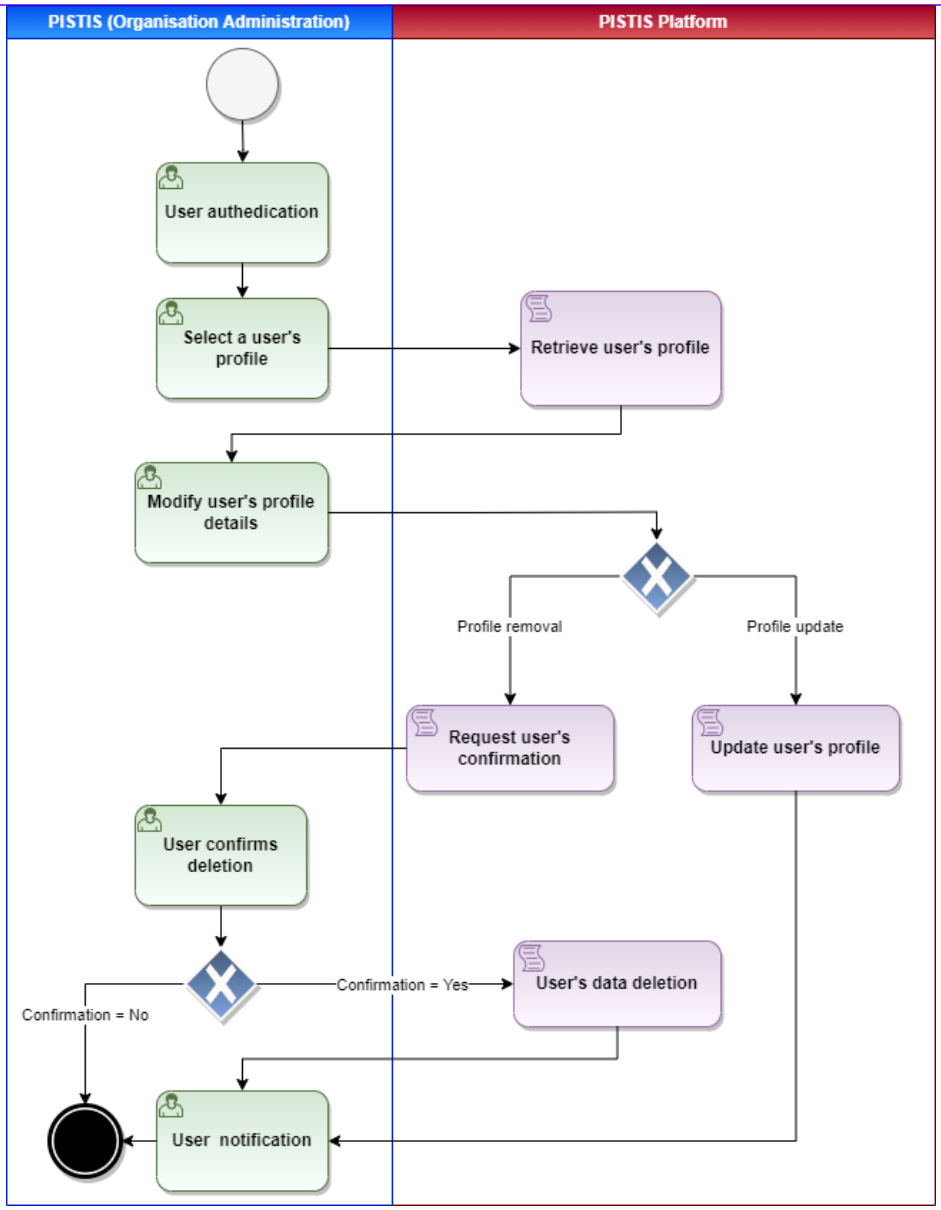



Figure 7: Process of PISTIS Organization Administrator manages a user's profile.

**Story Sequence**

**User account creation**

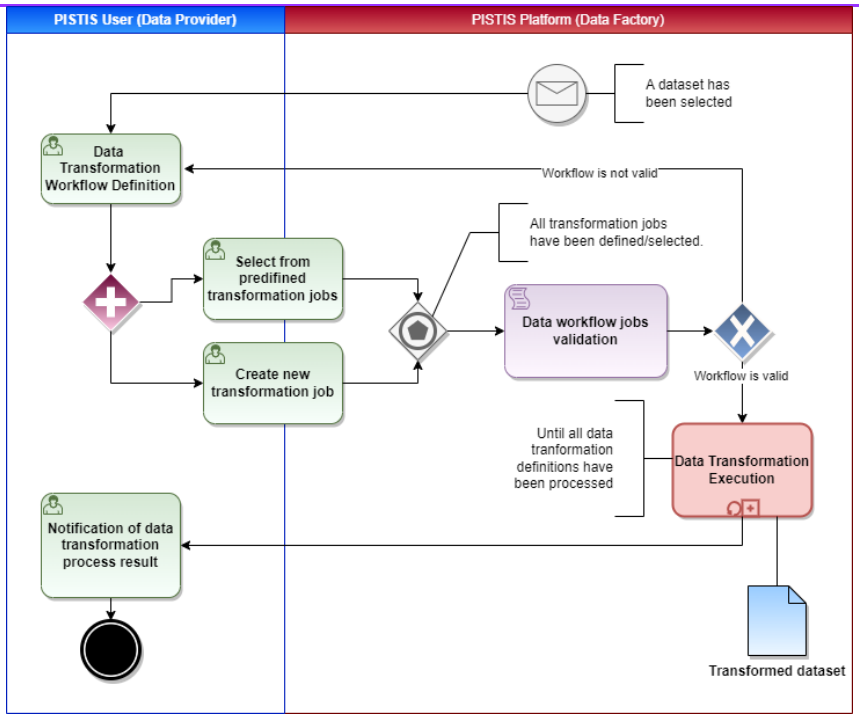
1. User accesses the registration form of PISTIS Platform, accepts PISTIS privacy policy and opts in for registering in the PISTIS platform.
2. A PISTIS Organization Administrator enters PISTIS Identity Access Manager and selects options for the new User's account, based on the following two options:
   - 2.1. Account will be created with username and password within PISTIS Identity Access Manager.
   - 2.2. Account will use User's existing identity in one of the available external identity providers. In this case it is required authorisation of the selected external identity provider that will be done by the User.
3. Account is connected to PISTIS Organization Administrator's Organization
4. Additional account details are specified, i.e. secondary profile details and other preferences (e.g., 2FA activation)
5. A Digital DLT-FIAT Wallet is created.

| | |
|---|---|
| | 6. PISTIS Platform requests and retrieves issued verifiable credentials from a Trusted CA for the specific user.<br>7. Bind a DLT Wallet Address with the user's verifiable credentials.<br>8. User account is created by default with a PISTIS_USER role<br>9. PISTIS Organization Administrator finally enables/disables new account<br><br>**User profile self-management**<br>10. PISTIS User logs in the PISTIS platform<br>11. PISTIS User accesses the profile management webpage.<br>11.1. PISTIS User edits secondary account details or preferences<br>11.2. PISTIS User manages their keys and/or certificates.<br>12. PISTIS Platform updates User profile.<br>13. PISTIS User is notified on the result.<br><br>**PISTIS Organization Administrator manages a user's profile**<br>14. PISTIS Organization Administrator logs in the PISTIS platform.<br>15. PISTIS Organization Administrator accesses the PISTIS Users management webpage.<br>16. PISTIS Organization Administrator selects an account to be edited:<br>16.1. PISTIS Organization Administrator disables/removes the account.<br>16.2. PISTIS Organization Administrator modifies the profile of the selected PISTIS User (assigned roles, related organisation, etc.).<br>17. PISTIS Platform updates User profile.<br>18. PISTIS Organization Administrator is notified on the result. |
| **User Benefits** | • A User will be able to obtain an account within PISTIS ecosystem and utilize, if desired, their existing identity in an external Identity Provider.<br>• A PISTIS User will be able to manage their account secondary details and preferences, as well as their keys/certificates.<br>• A PISTIS User will be able to bind their verifiable credentials with PISTIS Platform.<br>• PISTIS Organization Administrator will be able to manage a PISTIS User profile. |
| **Challenges** | • Ensure a user-friendly account creation process for PISTIS Organization Administrators.<br>• Integration with external Identity Providers.<br>• Integration with eIDAS Trusted CA<br>• Securely store and handle user credentials, personal details, and keys/certificates.<br>• 2FA implementation for enabling and disabling the option. |
| **Security Issues** | • Protect user credentials and sensitive data.<br>• Secure data exchange with the external Identity Providers.<br>• Ensure that User registration process complies with eIDAS requirements.<br>• Ensure 2FA implementation effectiveness.<br>• Identification and monitoring of unauthorized access and data breaches.<br>• Ensure immediate user role adjustment. |

| | |
|---|---|
| **Other issues** | - |

### 3.3.3 Data transformation and analysis

**Table 5: Data transformation and analysis User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.03** |
| **Story Title** | **Data transformation and analysis** |
| **Story Actors** | PISTIS Platform, PISTIS User (Provider) |
| **Overview** | The Pistis User (Provider) wants to perform a transformation process over a portion of data (dataset) in the PISTIS Platform (Data Factory). This process is defined as a data workflow of data transformation jobs. Each transformation job includes the triggering criteria for the transformation, the type of transformation and the transformation parameters. The transformation jobs defined that might apply (according to the triggering of each data transformation defined by the user) will be executed following the transformation criteria defined. The resultant dataset will be returned. |
| **Triggers** | The data transformation is a process that is executed on a selected dataset and can be part of other functional processes of the PISTIS Platform (i.e. Data check-in). |
| **Workflow** |  <br>**Figure 8: Data transformation and analysis**<br><br>**Data Transformation Execution Sub-process** |

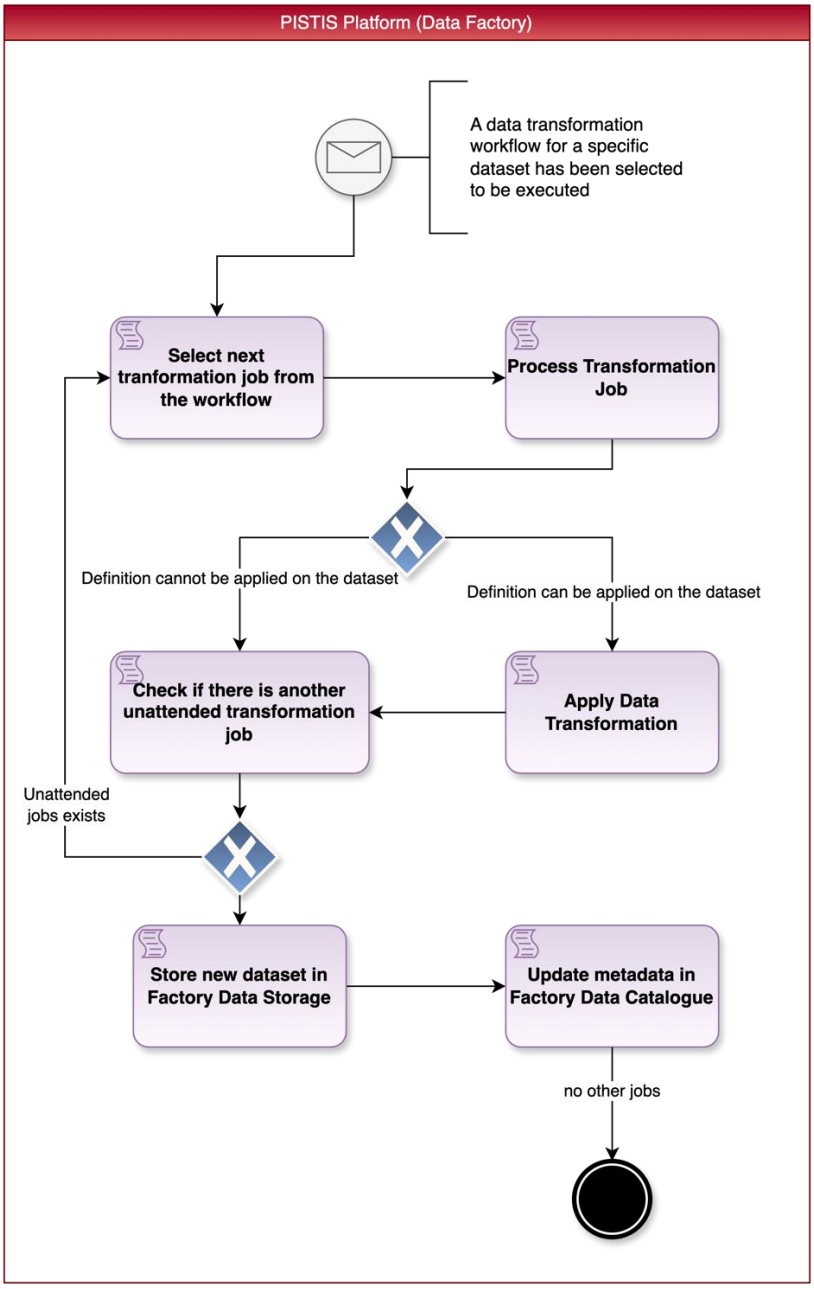



**Figure 9: Data Transformation Execution Sub-process**

**Story Sequence**

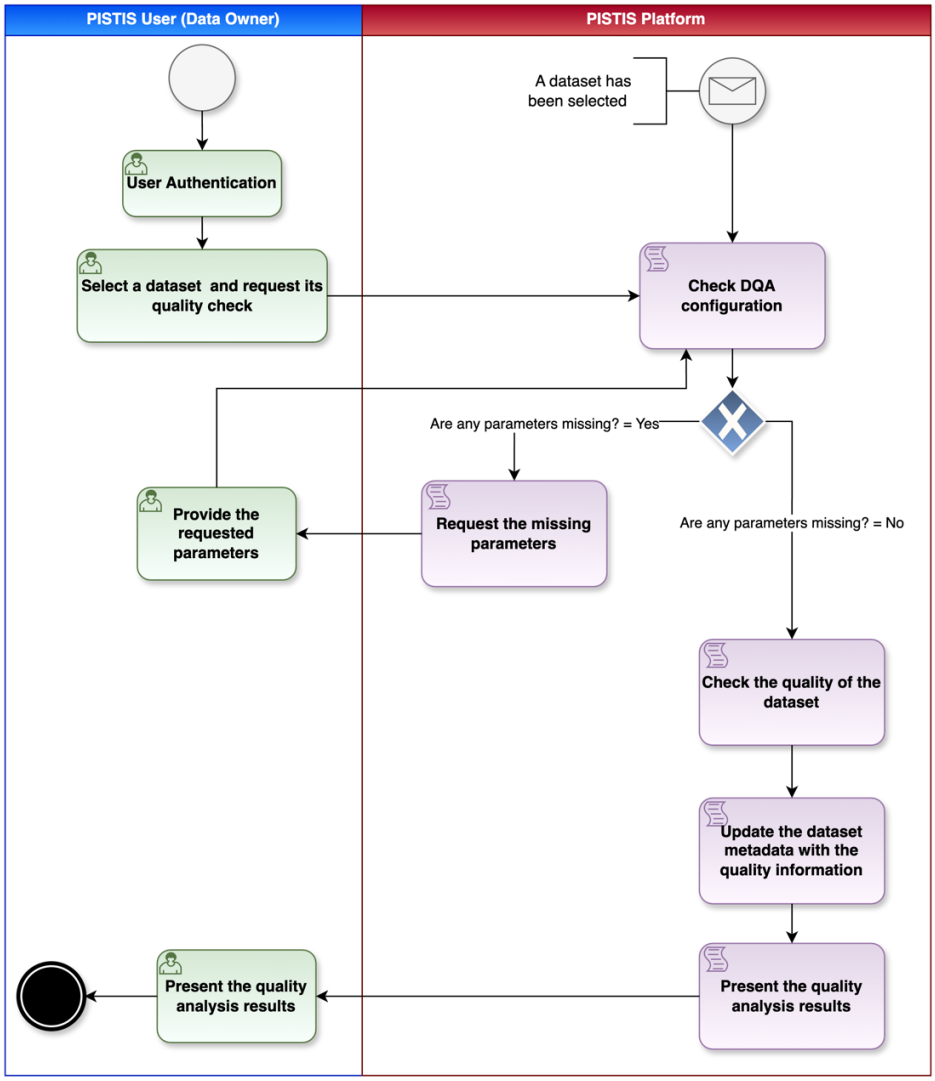
1. The process is started with the selection of a specific dataset. This selection is the result of another process, i.e. data check-in or PISTIS User's selection.
2. PISTIS User (Data Provider) defines a data workflow consisting of data transformation jobs. Each data transformation job includes a) a triggering criterion and b) a transformation definition including what needs to be transformed and how it needs to be transformed. Data transformation workflow is built by defining the transformation jobs following one the actions:
   2.1 Select from a set of predefined transformation jobs and provide the appropriate parameterisation.
   2.2 Create a new transformation job definition and provide the appropriate parameterisation.

| | |
|---|---|
| | 3. PISTIS Platform receives the transformation workflow and checks the validity of the workflow definition according to the definitions of the transformation jobs involved in the workflow.<br>4. PISTIS Platform executes the sub-process of data transformation execution:<br>4.1 *Data Transformation Execution Sub-process*:<br>4.1.1 PISTIS Platform selects the next unattended transformation job.<br>4.1.2 PISTIS Platform process the transformation job by checking if the triggering criteria are matched by the dataset:<br>4.1.2.1 Data transformation component performs the defined data transformation on the dataset in case the triggering criteria are matched by the dataset.<br>4.1.3 When the transformation is performed, or in case the dataset did not match the transformation triggering criteria, the platform checks if there are any other unattended transformation jobs.<br>4.1.3.1 If there are the process is re-executed by the step 4.1.1<br>4.1.4 New dataset obtained after applying transformations is stored in Factory Data Storage.<br>4.1.5 A new dataset distribution is created on the Factory Data Catalogue including metadata update.<br>4.1.6 Data transformation component ends processing all the data transformation definitions.<br>5. PISTIS User is informed that the dataset has been successfully transformed. |
| **User Benefits** | The user will be able to:<br>• Define or select the different transformations (e.g. regex replacement, row with missing values removal, normalization, etc.) to be carried out by the PISTIS Platform.<br>• Define the target dataset where the transformations will be applied.<br>• Define the triggering criteria to apply each one of transformations defined.<br>• Define the elements to be transformed in the target dataset.<br>• Define the transformation procedure |
| **Challenges** | • Create a definition of all the allowed transformations supported by the PISTIS Platform.<br>• Integrate a solution into the architecture to obtain (and provide) all the required resources for a proper functioning (i.e. transformation definitions, dataset repository for retrieval and storage, etc.). |
| **Security Issues** | - |
| **Other issues** | - |

### 3.3.4 Data quality check

**Table 6: Data quality check User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.04** |
| **Story Title** | **Data quality check** |
| **Story Actors** | PISTIS Platform (Data Factory), PISTIS User (Provider) |

| | |
|---|---|
| **Overview** | PISTIS User asks PISTIS Platform to check the quality of data and its metadata before the publication. PISTIS Platform checks the quality of (meta-)data and extends its metadata with the quality information. |
| **Triggers** | This process can be triggered by two events:<br>1. PISTIS User logs in and selects a dataset for quality checking.<br>2. A dataset has been already selected by the user in another process (i.e. Data check-in). |
| **Workflow** |  <br>**Figure 10: Data quality check process** |
| **Story Sequence** | 1. Selection of dataset for Quality check:<br>1.1. PISTIS User (Data Provider) selects a dataset stored in local Factory and requests the Platform to perform data quality assessment.<br>1.2. A dataset is already selected from another process that invokes quality check process.<br>2. The DQA component retrieves the Domain Data Model corresponding to the data set and automatically creates the first data quality requirements for format and domain validity of the fields in the dataset. |

| | |
|---|---|
| | 3. The user can confirm or correct the requirements extracted in Step 2, using the DQA User Interface.<br>4. The user selects and parametrises additional custom DQA metrics. Each metric requires different parameters (e.g., "Define the values that will be assimilated as missing values", "Define the values that are assimilated as defaults", "Define the degradation function and factor for timeliness", definition of complex dependencies between columns).<br>5. The PISTIS Platform checks that all information for the quality check is available.<br>5.1 Missing parameters in the DQA configuration: Return to step 4. PISTIS User is asked to provide the required information via the DQA-UI<br>6. The DQA Component maps the parametrised metrics to its internal Data Quality Language (currently Great Expectations lib) and instantiates a Data Quality process, consisting of tests related to each requirement.<br>7. PISTIS Platform checks the quality of the dataset by running the DQA process from Step 6.<br>8. The DQA process outputs a JSON file with a detailed report containing an itemised breakdown of the results for each data quality requirement. This includes: the name of the test run, the column(s) the test was run on, all test parameters, pass or failure of the test. Some tests provide further elucidation on failures: samples of failed values, percentage of failures etc.<br>9. The user receives a set of recommendations for improving the DQ of the dataset and decides whether to apply them or not.<br>9.1 If the user applies them, return to Step 6.<br>9.2 If the user doesn't apply them, continue to Step 10.<br>10. PISTIS Platform updates the metadata of the dataset with the quality information from Step 8.<br>11. PISTIS Platform parses the output JSON from Step 8 and presents the quality analysis results to the PISTIS User. |
| **User Benefits** | The user will get:<br>• The results of the dataset quality assessment according different KPIs<br>• The list of the identified issues<br>• The explanation of the dataset quality assessment methodology as reference information |
| **Challenges** | • The analysis of data may require a long time, so the workflow may become asynchronous with respect to the PISTIS user activity.<br>• Analyse the quality of dynamically changed data registered to PISTIS Platform cannot be achieved by continuous execution of the quality check especially if there is data that changes in a tight time frame. A solution including the storage of data snapshot and the execution of the data quality check on them is more feasible.<br>• Data quality requirements involve the definition of data quality rules, which involve a reasonable technical level on behalf of the user.<br>• Data quality assessment for data streams may require the application of sampling techniques or specific reporting (sample statistics, cumulative statistics etc.) |
| **Security Issues** | Secure data management and access need to be ensured. |

| Other issues | - |
|---|---|

### 3.3.5 Definition of access policies

**Table 7: Definition of access policies User Story**

| Story ID | **PISTIS.OUS.05** |
|---|---|
| **Story Title** | **Definition of access policies** |
| **Story Actors** | PISTIS User (Platform Administrator), PISTIS (Organisation Administrator), PISTIS Platform (IAM) |
| **Overview** | PISTIS modules at factory level are provided with corresponding access rights by PISTIS Administrator at the deployment time as described in PISTIS.OUS.15 Data Space Factory Configuration/Setup.<br><br>PISTIS Organisation Administrator wants to define a set of policies associated to a specific PISTIS role available for assignment to PISTIS Organization Users.<br><br>Each PISTIS access policy should describe: (a) who will be able to access a PISTIS Organisation resource (either a specific PISTIS Platform feature or a PISTIS Organisation's dataset), (b) the scope (rights) on the accessible PISTIS Organisation's resources and (c) refine accessibility on nested objects/attributes within a certain PISTIS Organisation's resource (e.g. allow Read of a data stream for all Organizations but exclude Trade scope for Organizations originating in a certain set of countries). |
| **Triggers** | PISTIS Organisation Administrator wants to refine certain access rights to a PISTIS Organization Role or create new policies to match PISTIS data exchange requirements. |

**Workflow**

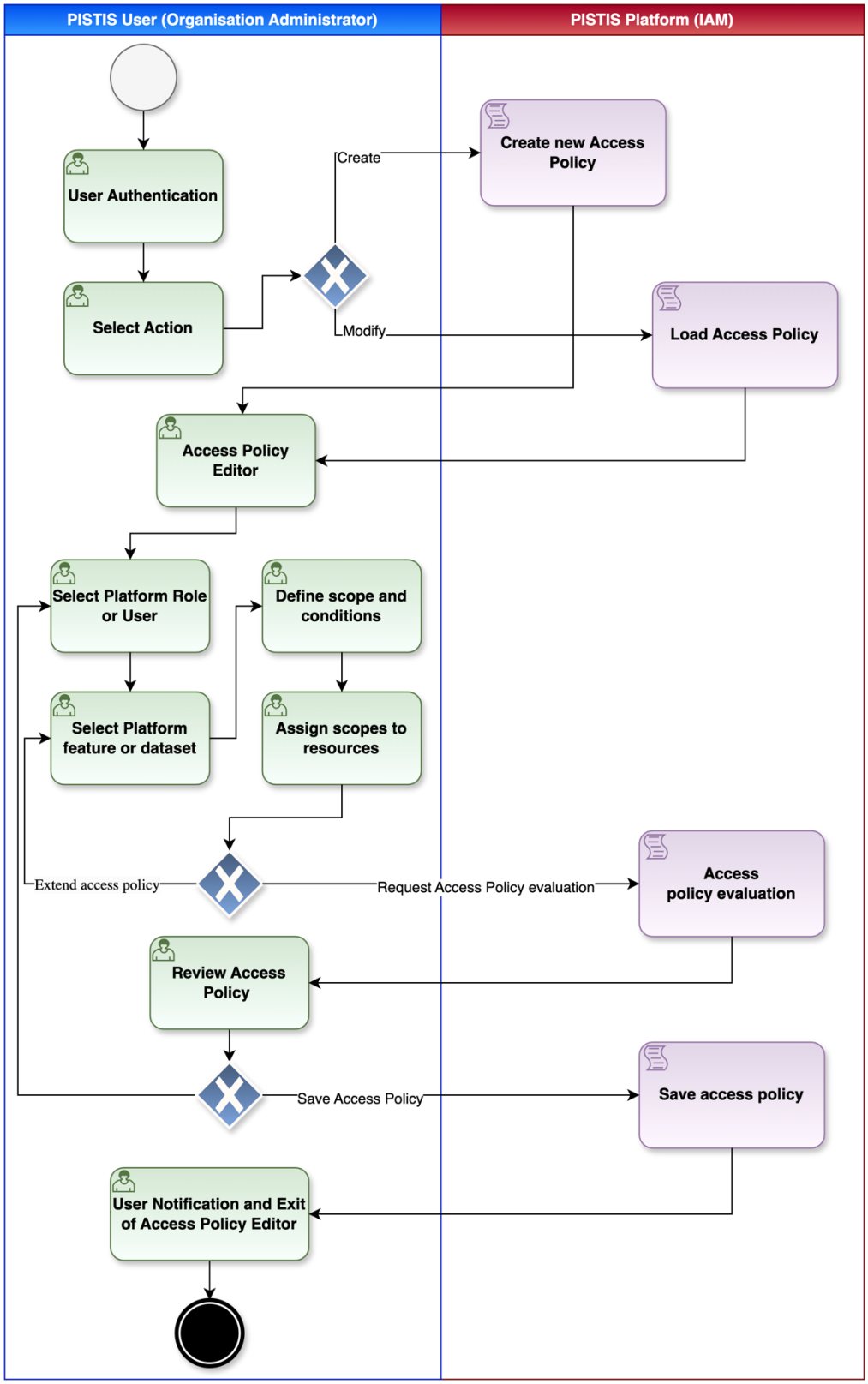



Figure 11: Definition of access policies

**Story Sequence**

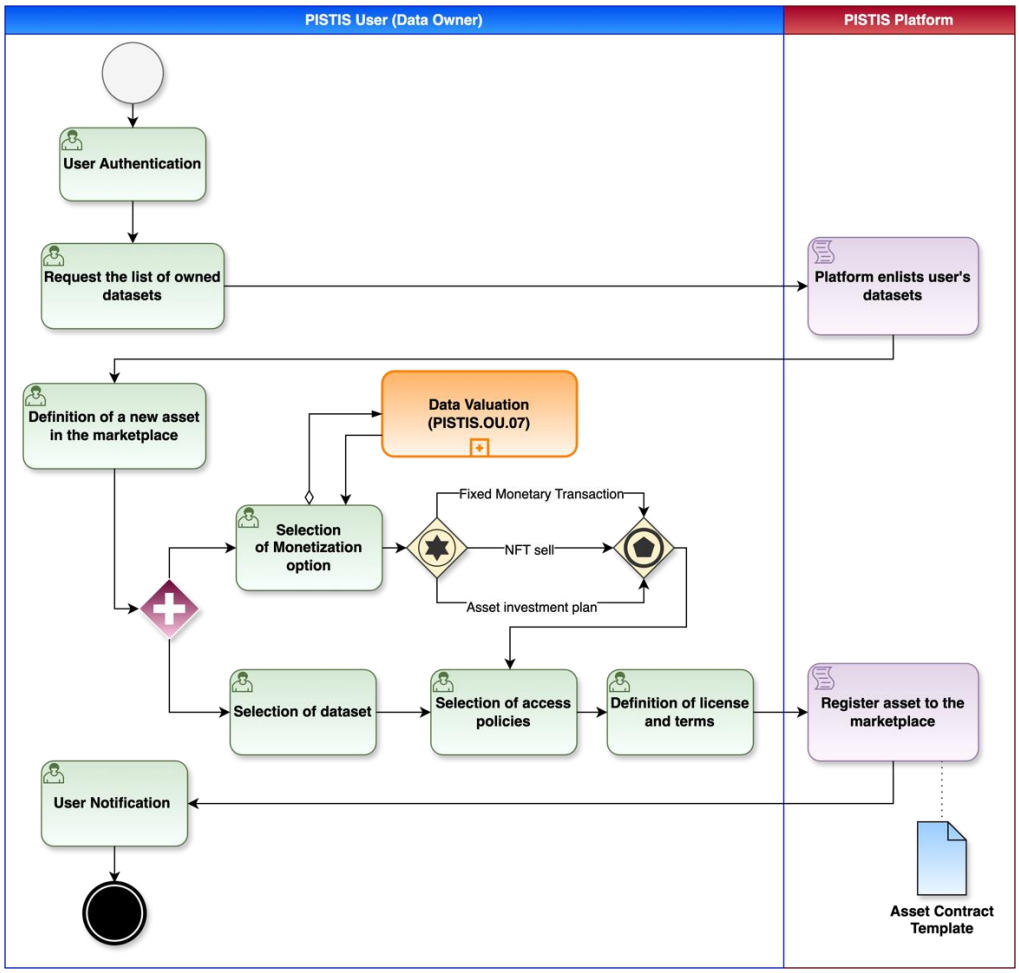
1. PISTIS User (Organisation Administrator) accesses a web interface that provides the necessary functionality for creating a new Access Policy for a PISTIS Organisation or to modify an existing one.

| | |
|---|---|
| | 2. PISTIS User adds/removes admission of access of a certain object (a PISTIS Organisation's data stream/dataset) to the selected PISTIS Organisation or to a matching set of criteria reflecting Organizations' attributes.<br>3. PISTIS User modifies the access policy to the selected resource in the following steps.<br>3.1. PISTIS User specifies the scope of access for each accessible Object.<br>3.2. PISTIS User defines conditions (including attributes-based policies) for the selected scope.<br>4. Steps 2 and 3 can be repeated until PISTIS User completes the definition of the access policy.<br>5. When PISTIS User completes access policy definition, PISTIS Platform Administrator evaluates the defined policy.<br>6. PISTIS User reviews the evaluation results on the modified PISTIS Organisation Role.<br>6.1. PISTIS User continues PISTIS Organisation Role editing (step 2 is repeated).<br>6.2. PISTIS User commands PISTIS platform to save the defined PISTIS Organisation Role. |
| **User Benefits** | • Effective access policies per PISTIS Platform feature, data stream/dataset and attribute with extended access policies to match PISTIS data trading requirements Fine-tuned access policies on nested objects/attributes within a PISTIS resource to match PISTIS data exchange requirements. |
| **Challenges** | • Definition of precise and granular access policies.<br>• Ensuring that access policies are aligned with PISTIS business and regulatory requirements.<br>• Provision of a user-friendly graphical environment to the PISTIS User (Organisation Administrator) to efficiently modify access policies, balancing security with user convenience.<br>• Provide an effective way to evaluate an access policy before applying it to the PISTIS ecosystem. |
| **Security Issues** | • Protect the integrity of access control mechanisms.<br>• Ensure that each access policy is protected against unauthorized access.<br>• Ensure/Establish access policy immediate enforcement.<br>• Ensure that access policies do not overlap or do not conflict each other. In case of conflict, provide transparent and explanatory insights/notifications to PISTIS Administrators. |
| **Other issues** | - |

### 3.3.6 Monetization of the data

**Table 8: Monetization of the data User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.06** |
| **Story Title** | **Monetization of the data** |

| | |
|---|---|
| **Story Actors** | PISTIS User (Provider), PISTIS platform |
| **Overview** | The PISTIS Users (Data Provider) need to define the way that their dataset will be placed on the market to be able to make a profit from the foreseen transactions. |
| **Triggers** | Upon selecting a dataset to be placed on the market as an asset for sale, the Data Monetisation workflow is triggered to define the monetisation method for this exact data asset. |
| **Workflow** |  <br>**Figure 12: Monetization of the data** |
| **Story Sequence** | 1. An authenticated PISTIS User requests from the PISTIS platform the list of the owned datasets.<br>2. The PISTIS User prepares an asset that is not yet present on the market.<br>2.1. The PISTIS User selects the dataset (data and metadata) that wants to monetize through the PISTIS Data Catalogue. The Data Catalogue redirects the User to the Monetization Plan page (Monetization Plan Designer component).<br>2.2. The user selects the way that he is willing to sell the asset by opting for:<br>• a fixed monetary transaction (that could have also the value of zero) or a monthly fee as a subscription,<br>• selling the NFT of the asset,<br>• creating an investment plan for the asset |

| | |
|---|---|
| | 2.2.1.(Optional) PISTIS User can perform data valuation as assistance to his decision on the monetization option.<br>3. The user selects the access policies for the dataset, in order to make this available to only stakeholders that do possess specific attributes<br>4. The user defines the license that will accompany the dataset, alongside with various specific contract terms such as contract duration, prices, etc.<br>5. PISTIS platform registers the new asset (as metadata) to the market and a contract template is created to facilitate the smart contract creation process.<br>6. PISTIS Platform notifies PISTIS User that the asset has been published in the marketplace. |
| **User Benefits** | The user gets an interface where they can:<br>• choose the way that the asset is going to be sold,<br>• set the price of the asset alongside some descriptions of the asset,<br>• draft the transaction contract that the buyer must accept when buying the asset. |
| **Challenges** | • The definition of the pre-defined terms of the transaction contract template is a complex task as may include various and different terms for governing the data transfers. |
| **Security Issues** | - |
| **Other issues** | • Interaction with the Data Valuation components and the market analytics is foreseen to inform the user prior to setting the price.<br>• In the case of an NFT transaction, all data provider existing contracts for that specific dataset should be transferred to the new NFT owner.<br>• In the case of an investment plan transaction, all data provider existing contracts for that specific data asset, should be updated to also include the new party that co-owns the dataset. |

### 3.3.7 Data Valuation

**Table 9: Data Valuation User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.07** |
| **Story Title** | **Data Valuation** |
| **Story Actors** | PISTIS Platform, PISTIS User (Owner, Consumer) |
| **Overview** | The Owner or the Consumer want to get a quantitative and qualitative assessment of the value of a data asset. |
| **Triggers** | • Asset Description Bundler, when an Owner wants to assess the value of their data asset before prior to its publication to the PISTIS Cloud Catalogue.<br>• A Consumer who wants a custom assessment of the value of a data asset before buying it. |

**Workflow**

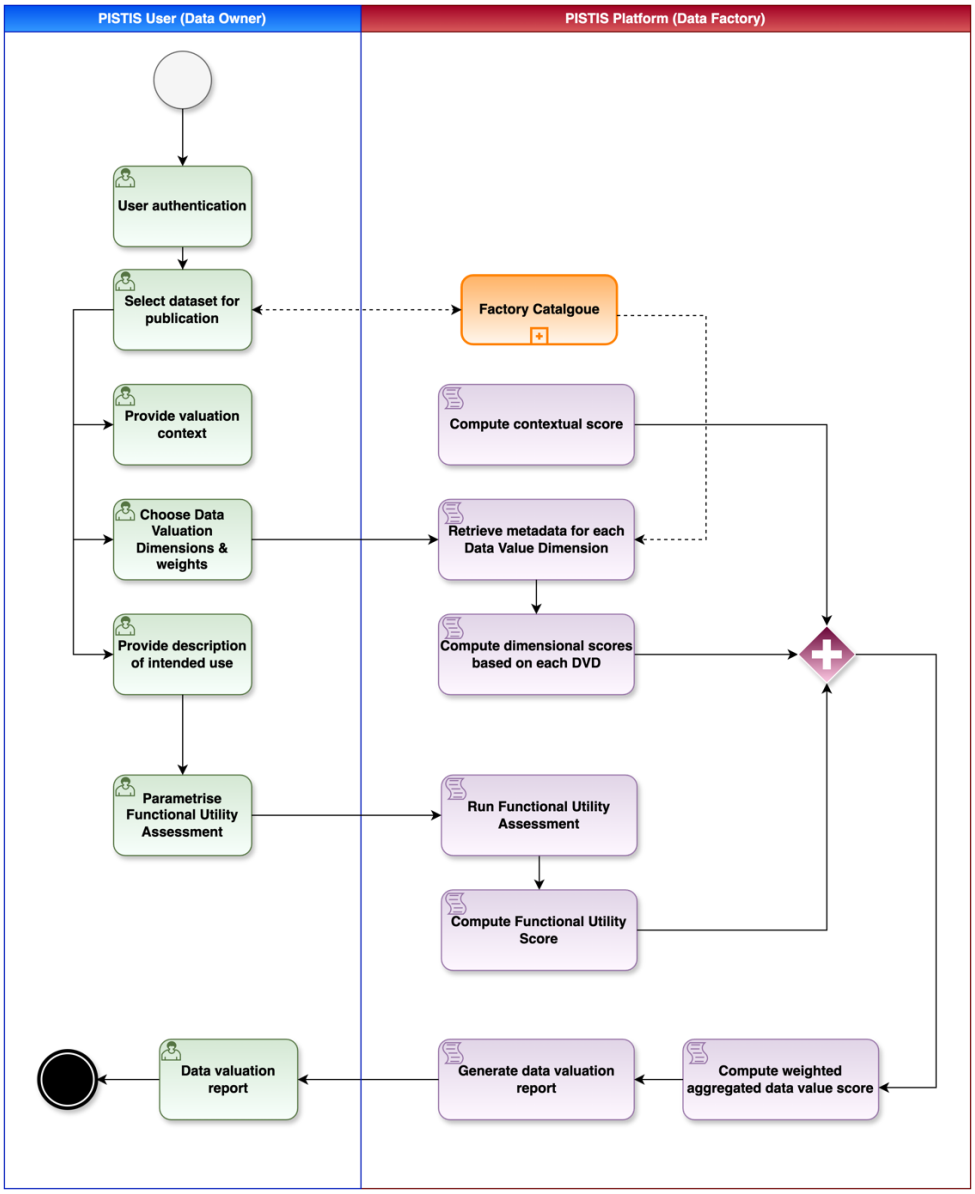



Figure 13: Data Valuation

**Story Sequence**

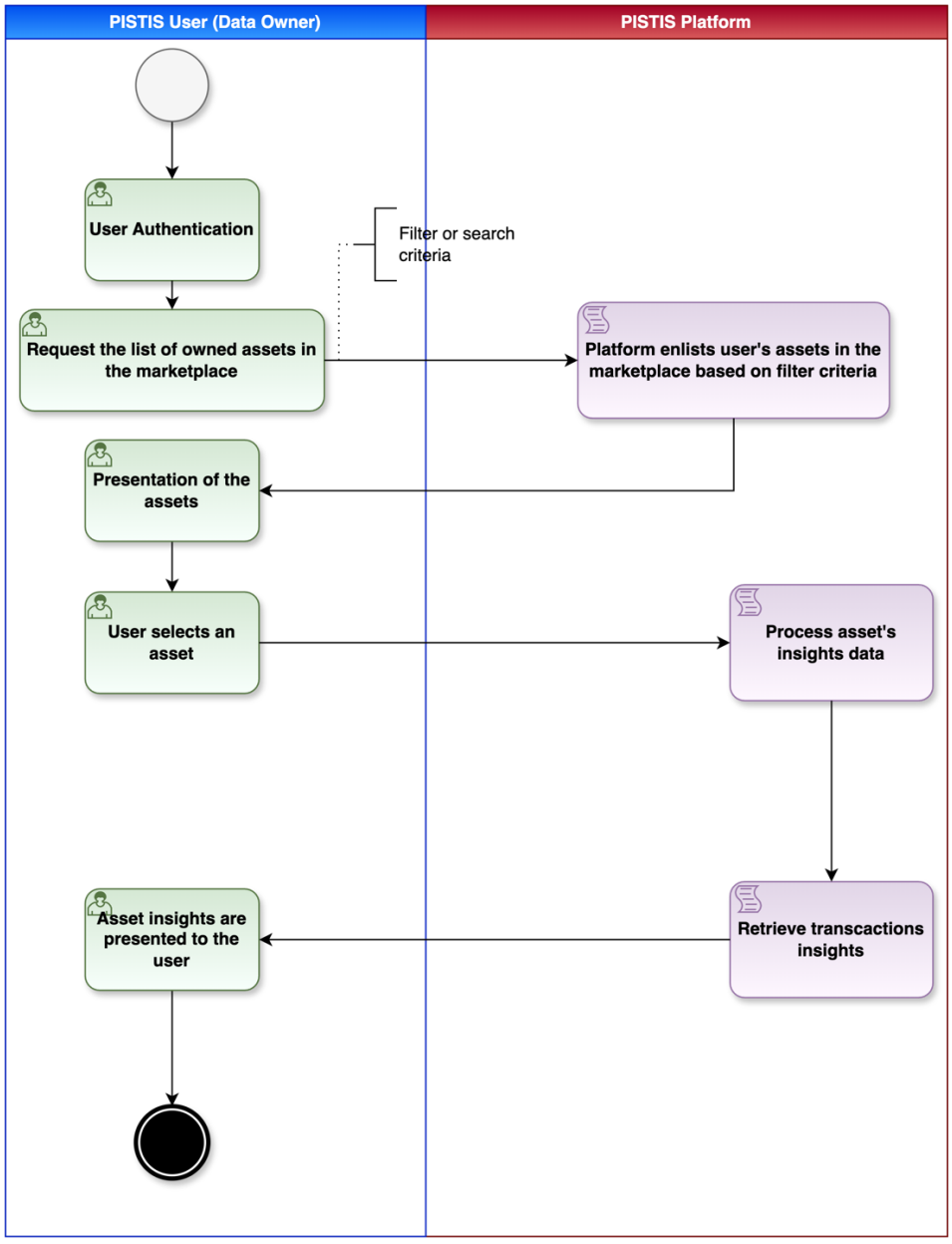
1. PISTIS Owner selects a data asset from their Factory for publication. Alternatively, a Consumer discovers a data asset through the Cloud Catalogue and request a custom valuation.
2. PISTIS User defines the valuation context, by providing the importance of the different data value dimensions (e.g., accessibility, size, format, origin, type, tools, data quality, functional utility, GDPR compliance, deanonymisation risk, business impact etc.) through a self-intuitive web page with form controls, one for each dimension. questionnaire that has been previously created by the PISTIS Administrator.
3. PISTIS User provides a description of the intended use of the data asset (e.g., data analysis, data visualisation, training algorithms), in the corresponding designated questionnaire.
   3.1. If training an algorithm (regression, classification, clustering), the PISTIS User will also define a set of metrics and minimum requirements for those metrics.

| | |
|---|---|
| | 4. Data Valuation Service retrieves the metadata relevant to each of the requested data valuation dimensions from the Factory Catalogue.<br>5. Combine the scores from the different data value, according to their importance, as previously defined by the PISTIS User.<br>6. Generate a data valuation report, including the general data value score, as well as its breakdown along the data value dimensions. Generate user-friendly explanations of the scores. |
| **User Benefits** | The user gets:<br>• A method for stating the context in which data valuation is performed. This allows for different users to perform various valuations for the same data asset. By default, non-contextual valuation involves only data quality assessment, GDPR check and deanonymisation risk analysis.<br>• The possibility to transform the data set based on the results of data quality assessment.<br>• A report of the data value score, as well as its breakdown along the defined data value dimensions. |
| **Challenges** | • From a UX perspective, the user interaction is challenged by the entire process complexity<br>• The definition of the context can be lengthy and may require a certain level of introspection (e.g., users need to carefully think of intended uses, business impact etc.)<br>• Defining the importance of each data value dimension may not be so straightforward. We will need to think what's the granularity of the tuning that will be allowed for each dimension. A continuous scale from 0-1 might pressure the user into making too fine adjustments. An ordinal scale, while losing granularity, might help at least in the beginning.<br>• Reporting is always difficult, especially for a highly dimensional process, such as data valuation. |
| **Security Issues** | *When a Consumer requests a Data Valuation for an asset in the Factory of another user, the process will be triggered using the JWT of the Consumer. Is this something that will work seamlessly? Is it a security issue?* |
| **Other issues** | *Data valuation doesn't generate a financial value of data. Instead, it is quantifying different dimensions which are associated with the value of data. Data monetisation can use these to compute a monetary value of the data asset, based on the scores obtained for each dimension.* |

### 3.3.8 Data usage and market analytics

**Table 10: Data usage and market analytics User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.08** |
| **Story Title** | **Data usage and market analytics** |
| **Story Actors** | PISTIS User (Provider), PISTIS Platform |

| | |
|---|---|
| **Overview** | The data provider uses facilities to understand how his/her data are used currently and what is the potential of them |
| **Triggers** | The data providers want to set a new data asset on the market and wants to get informed about how it could perform<br><br>The data provider wants to check how a data owned is being used by buyers or by other interested parties. |
| **Workflow** |  <br>**Figure 14: Data usage and market analytics** |
| **Story Sequence** | 1. The PISTIS User visits the market insights page where they can have a view of the overall market.<br>2. The PISTIS User filters the market insights dashboard to constrain the view of assets that are similar to the one they would like to sell (at sector level).<br>3. The PISTIS User gets some insights on how their data (which they already sold in the past) is currently used by buyers, by receiving answers from the Data |

| | |
|---|---|
| | Usage Questionnaires that capture the opinion of verified buyers, as well as from other interested users. |
| **User Benefits** | The user gets:<br>• A dashboard to view specific elements of the overall PISTIS market.<br>• Insights relevant to the data category/type the user is interested in selling.<br>• Insights on how data the user has already sold are being used by verified buyers or how they could be used by other interested stakeholders. |
| **Challenges** | Insights on the performance of data assets on the market rely heavily on the existence of a high number of transactions happening in the market, and on the availability of similar datasets to the ones under investigations.<br>Understanding usage of data from other parties via market analysis is challenging |
| **Security Issues** | N/A |
| **Other issues** | Usage Analytics should be restrained only on market transaction analyses and replies to questionnaires, as other information relevant to usage may violate users' business secrets. |

### 3.3.9 Data Query

**Table 11: Data Query User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.9** |
| **Story Title** | **Data Query** |
| **Story Actors** | PISTIS Platform, PISTIS User (Consumer) |
| **Overview** | The PISTIS User (Consumer) access the PISTIS platform to search of a dataset. Firstly, the user can browse/search the metadata catalogue for datasets. In case the user wants to proceed to more thorough search in the data, they can use the platform interface to define search criteria as a set of keywords. Upon the criteria's validation, the Platform will query the metadata for the best possible matches and in parallel forwards the search to the stored data. The results of both modules will be merged into a single list which will then be sorted and returned to the user. This data querying process will also be supported for encrypted data through the DSSE (Dynamic Symmetric Searchable Encryption) functionality exposed, which allows the characterization of encrypted data with similar metadata. |
| **Triggers** | An existing PISTIS User (Consumer) wants to find one or more Datasets that match their needs. |

**Workflow**

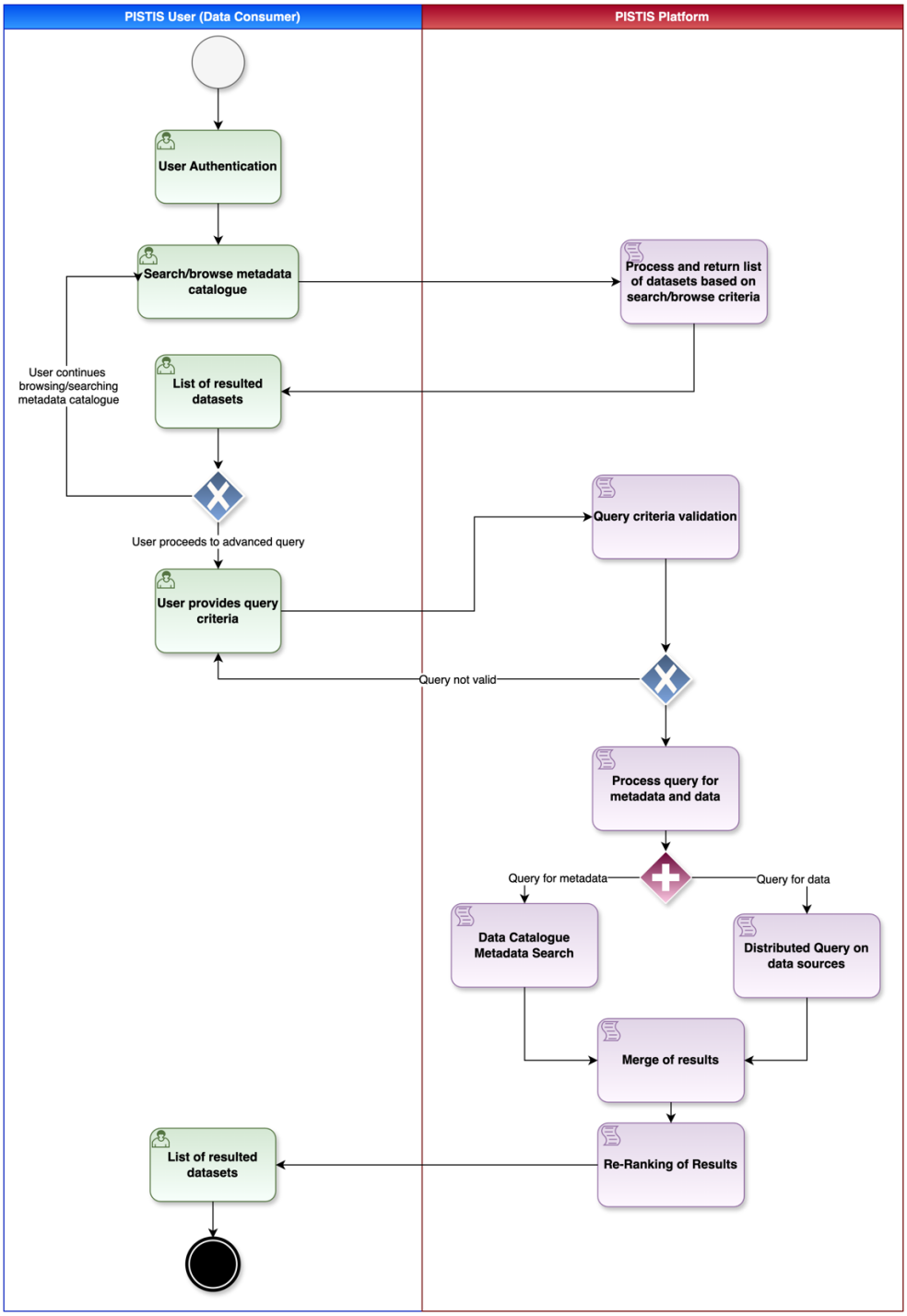



Figure 15: Data Query

**Story Sequence**

1. PISTIS User is authenticated by the PISTIS Platform.
2. PISTIS User search/browse for datasets in the metadata catalogue.
3. PISTIS User wants to proceed to more advanced search providing Query Criteria in natural language.
   3.1. PISTIS Platform validates the Query Criteria based on the grammar and language rules. If the query is not valid, the user is notified.
4. PISTIS Platform splits the Query to a part that is aimed to the Metadata and a part that is aimed to the Data.
   4.1. The Metadata part is forwarded to the PISTIS Data Catalogue.

| | |
|---|---|
| | 4.2. The Data part executes a query mechanism to the distributed data spaces.<br>5. PISTIS Platform merges the results generated by both subtasks.<br>6. PISTIS platform re-ranks the results list and sends it to PISTIS User. |
| **User Benefits** | The user will be able to:<br>• Search for a dataset using metadata.<br>• Search for a dataset<br>• Informed for possible errors in the criteria they provided for the search.<br>• Receive an ordered list of datasets that match the criteria they provided. |
| **Challenges** | • The variety of the different formats and schemas used for the storage of metadata introduces a great level of complexity.<br>• The volume of data stored in the Data Factories makes extensive searches difficult and slow.<br>• When the list of potential matches becomes large the user might not find what they are looking for if it is not placed on the top. |
| **Security Issues** | • Confidentiality and Authentication: Only the correct user, as defined by the various policies, should be able to have access to perform this task.<br>• System abuse: The user should be able to perform only a reasonable number of queries, so the system is protected against scrappers and other forms of abuse. |
| **Other issues** | - |

### 3.3.10 Data Transaction

**Table 12: Data Transaction User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.10** |
| **Story Title** | **Data Transaction** |
| **Story Actors** | PISTIS Platform, PISTIS User (Provider), PISTIS User (Consumer) |
| **Overview** | There is an intention between two users (provider and consumer) to transfer a dataset from the producer to the consumer under a specific license that defines the terms of sale. The PISTIS platform should let the two parties seamlessly transfer the data and its metadata. The data provider has listed their datasets on the PISTIS platform, and the data consumer has loaded their wallet with the necessary amount of money. The transaction is initiated by the generation, signature, and activation of a smart contract, based on the smart contract template created on data monetization process (PISTIS.OUS.06). Upon the sign of the smart contract the peer-to-peer data and metadata exchange follows between data provider and data consumer through a dedicated Data Factory Connector between the Data Factories of the two PISTIS Users. |
| **Triggers** | Once a data consumer wants to buy a data asset, the PISTIS platform holds the price that the data provider has placed on the data, and a template of the contract to which terms the two parties should agree for executing the transaction. The data consumer sends a request to the platform, using the dedicated user interface, to initiate the data exchange and the transfer of the contract value to the data provider's wallet. |

**Workflow**

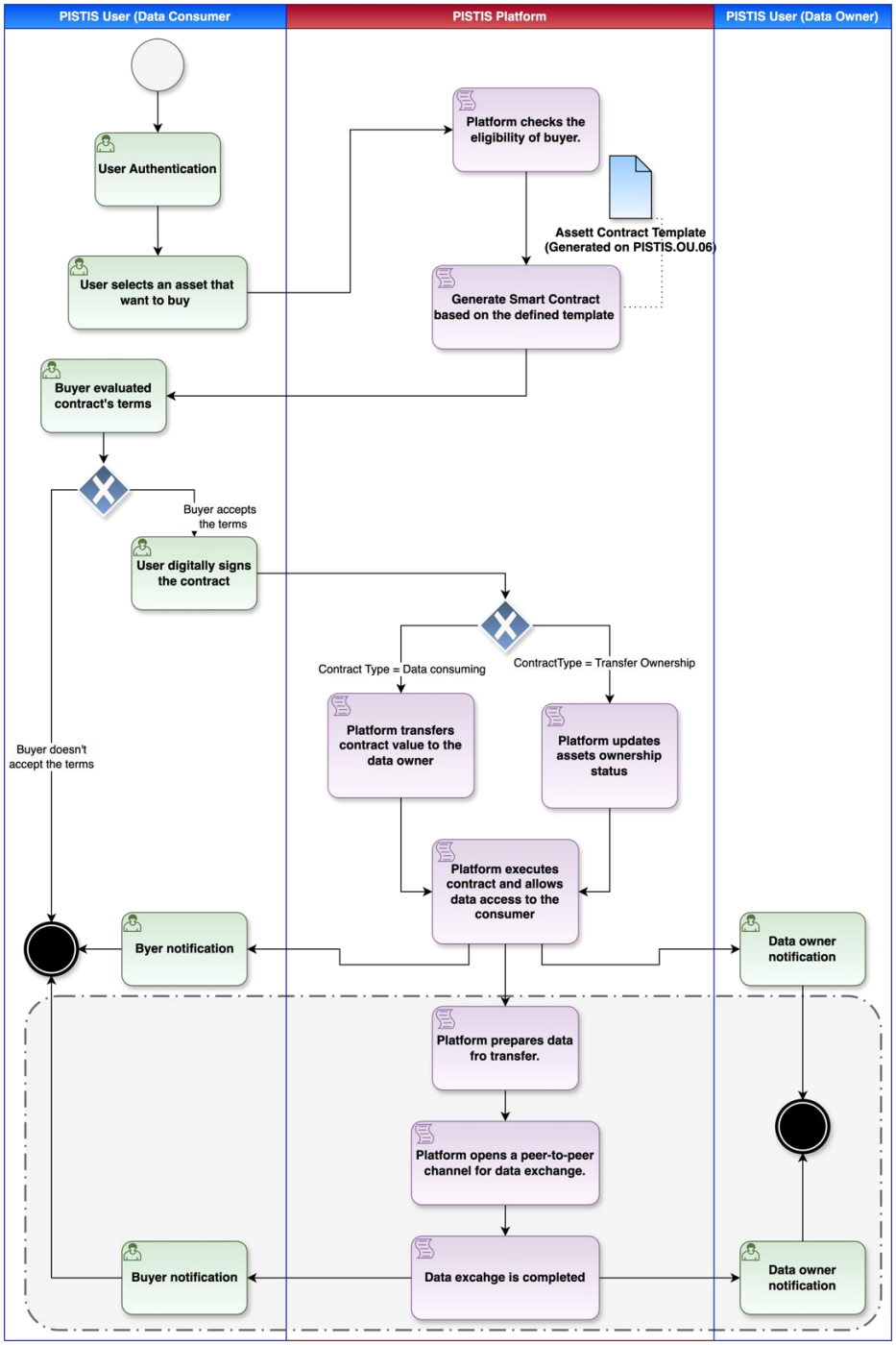



Figure 16: Data Transaction

**Story Sequence**

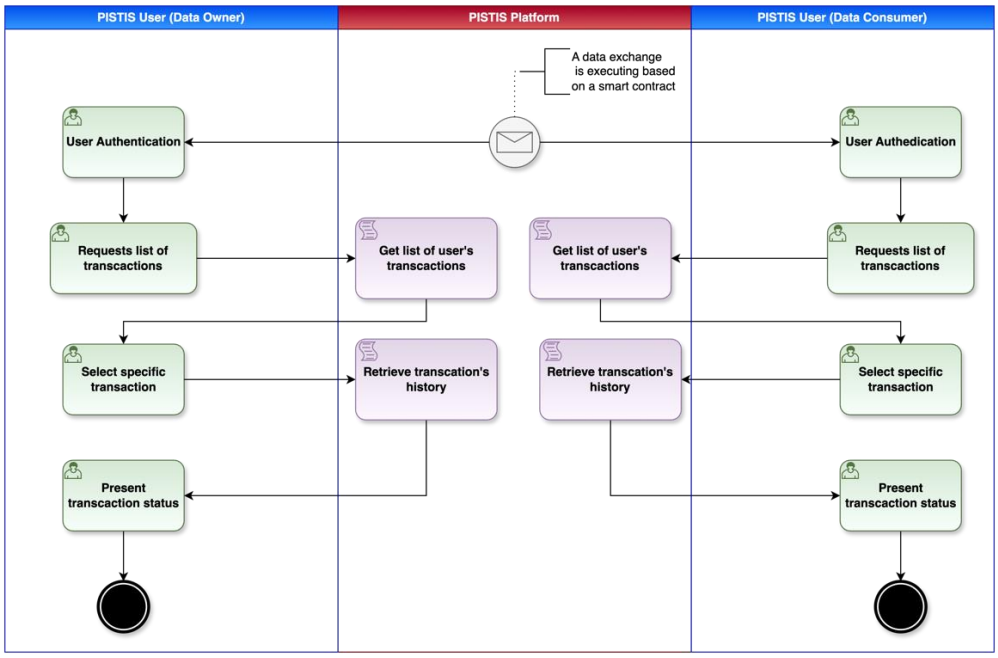
1. The PISTIS User (Consumer) views a data asset available in the market and chooses to buy it.
2. The PISTIS platform checks if the Consumer possesses the necessary amount of PISTIS coins to execute the transaction.
3. The PISTIS Platform generates a smart contract based on asset's contract template to be signed by the PISTIS User (Consumer)
4. The PISTIS User (Consumer) views the contract terms and signs it, utilising their hardware based identity key through their wallet.
5. The PISTIS Platform, depending on the type of contract,

| | |
|---|---|
| | 5.1. transfers the contract value amount from the Consumer to the Owner of the data, with a portion of the contract value allocated as fees for the platform.<br>5.2. In case of partial (investment plan) or totally (NFT) ownership change, the PISTIS Platform records in the ledger the new ownership status of the dataset.<br>6. The PISTIS Platform executes the smart contract to allow the data to be exchanged between the different PISTIS Data Factory environments.<br>7. The PISTIS platform informs both the Owner and the Consumer for the execution of the smart contract.<br>8. The PISTIS platform executes the data exchange.<br>9. The PISTIS platform encrypts the data before initialising the transfer.<br>10. The PISTIS platform activates a data exchange channel mediated through the Data Factory Connectors , based on the contract terms, through which data is transferred from one PISTIS data factory to the other and the dataset's metadata is transferred from the market to the Consumer's factory.<br>11. The PISTIS platform informs both the Owner and the Consumer for the competition of the data transfer. |
| **User Benefits** | • The PISTIS User (Consumer) gets a facility to check the terms of a contract and sign the contract.<br>• The PISTIS User (Provider) gets a facility to automate asset transactions using smart contracts.<br>• The PISTIS users can easily execute financial transactions. The process of conversion between FIAT money and Pistis stablecoin is hidden from the users. Moreover, users don't have to worry about exchange ratios.<br>• The PISTIS users do not have to hold any receipt of the purchase or manuscript of contract since everything is recorded on the blockchain. |
| **Challenges** | In case of ownership change, the platform must care to change the attributes of past but still valid contracts relevant to the same dataset, to allow the new owners to profit from new transactions.<br><br>In case of ownership change, the platform must transfer the dataset to the new owner.<br><br>In case the registration of the data transaction, especially for stream data, value transaction and smart contract on the relevant blockchain should be fast enough. This is a challenge since most of the blockchain networks require some time for status update. For example, Bitcoin has 10 minutes block confirmation time while Ethereum has on average 10-15 seconds delay. |
| **Security Issues** | N/A |
| **Other issues** | N/A |

### 3.3.11 Data exchange monitor/audit

**Table 13: Data exchange monitor/audit User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.OUS.12** |
| **Story Title** | **Data exchange monitor/audit** |

| | |
|---|---|
| **Story Actors** | PISTIS Platform, PISTIS User (Provider), PISTIS User (Consumer), PISTIS User (Auditor, Certification Body) |
| **Overview** | Authorised PISTIS Users will be able to get real time information about the progress of the data exchanges and will be notified accordingly. |
| **Triggers** | **Monitor Specific Dataset Exchange**<br><br>A PISTIS User (Consumer) has already bought a data asset through the PISTIS Platform. Thus, the contract is already verified, the dataset is encrypted, the money has already been transferred successfully, and the dataset is scheduled to be transferred peer-to-peer to the PISTIS User (Consumer). During this procedure an auditing mechanism is enabled to notify the PISTIS User (Owner and Consumer) about the data transaction status.<br><br>**Auditing History**<br><br>A PISTIS User (Auditor, Certification Body) can search the history of transactions. |
| **Workflow** | **Monitor Specific Dataset Exchange**<br><br>  <br><br>Figure 17: Monitor Specific Dataset Exchange process.<br><br>**Auditing History** |

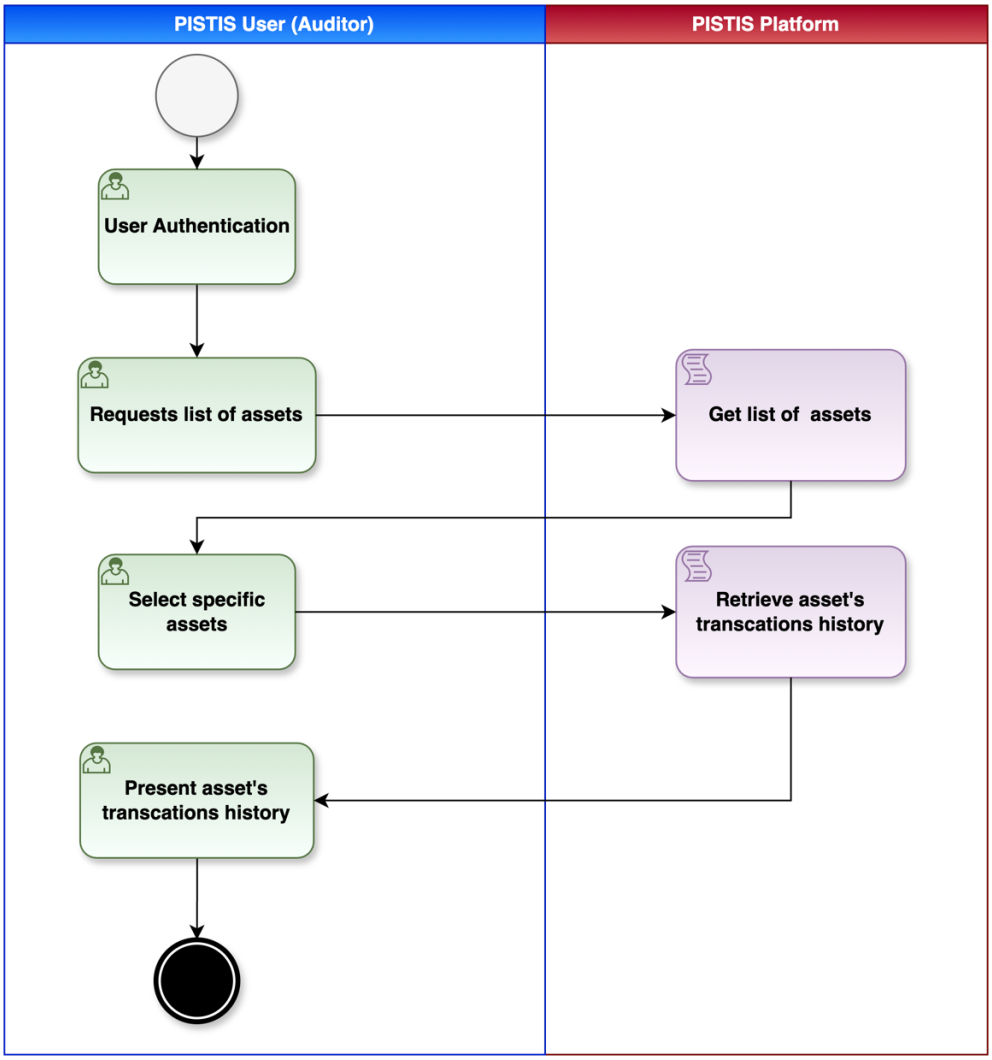



Figure 18: Auditing History process.

**Story Sequence**

**Monitor Specific Dataset Exchange**

1. PISTIS User (Consumer and Owner) successfully authenticated and authorized the Platform.
2. PISTIS User requests the list of transactions and selects one of them to view the status.
3. The PISTIS platform stores the hashes of the value and data transactions along with the dataset information, a pointer to the off-chain storage and a timestamp signed by the PISTIS Platform and generates a certificate that the transaction was successful. Further, PISTIS Platform updates the information on the lineage tracker.
4. PISTIS User can view the status and all the details of the selected transaction.
   4.1. PISTIS User (Provider) is notified by the PISTIS Platform that the transaction is completed.
   4.2. PISTIS User (Consumer) is notified by the PISTIS Platform that the transaction is completed.

**Auditing History**

| | |
|---|---|
| | 1. PISTIS User (Auditor, Certification Body) successfully authenticated and authorized.<br>2. PISTIS User searches in the PISTIS Platform for a specific dataset of interest to access the transaction history.<br>3. PISTIS User requests from the PISTIS Platform the history of the specific dataset.<br>4. PISTIS platform searches all the smart contract instances for this specific dataset and provides the history of transactions of the specific dataset to the PISTIS User (Auditor, Certification Body). |
| **User Benefits** | PISTIS User (Consumer) will be able to:<br>• Get notifications about the data transaction status.<br>PISTIS User (Provider) will be able to:<br>• Get notifications about the data transaction status.<br>PISTIS User (Auditor, Certification Body) will be able to:<br>• Access the data transaction history |
| **Challenges** | PISTIS User is successfully authenticated and authorized to the PISTIS Platform. |
| **Security Issues** | • Confidentiality & Authentication & Authorization: The transactions monitored data should be encrypted and only authorised PISTIS Users should be able to access them.<br>• Integrity: The transactions monitored data should not be altered.<br>• Availability: The transactions monitored should be available to the authorised PISTIS Users or PISTIS Administrator.<br>• Privacy: The identity of the involved users of the actual data transaction and the corresponding payment value should not be revealed. |
| **Other issues** | N/A |

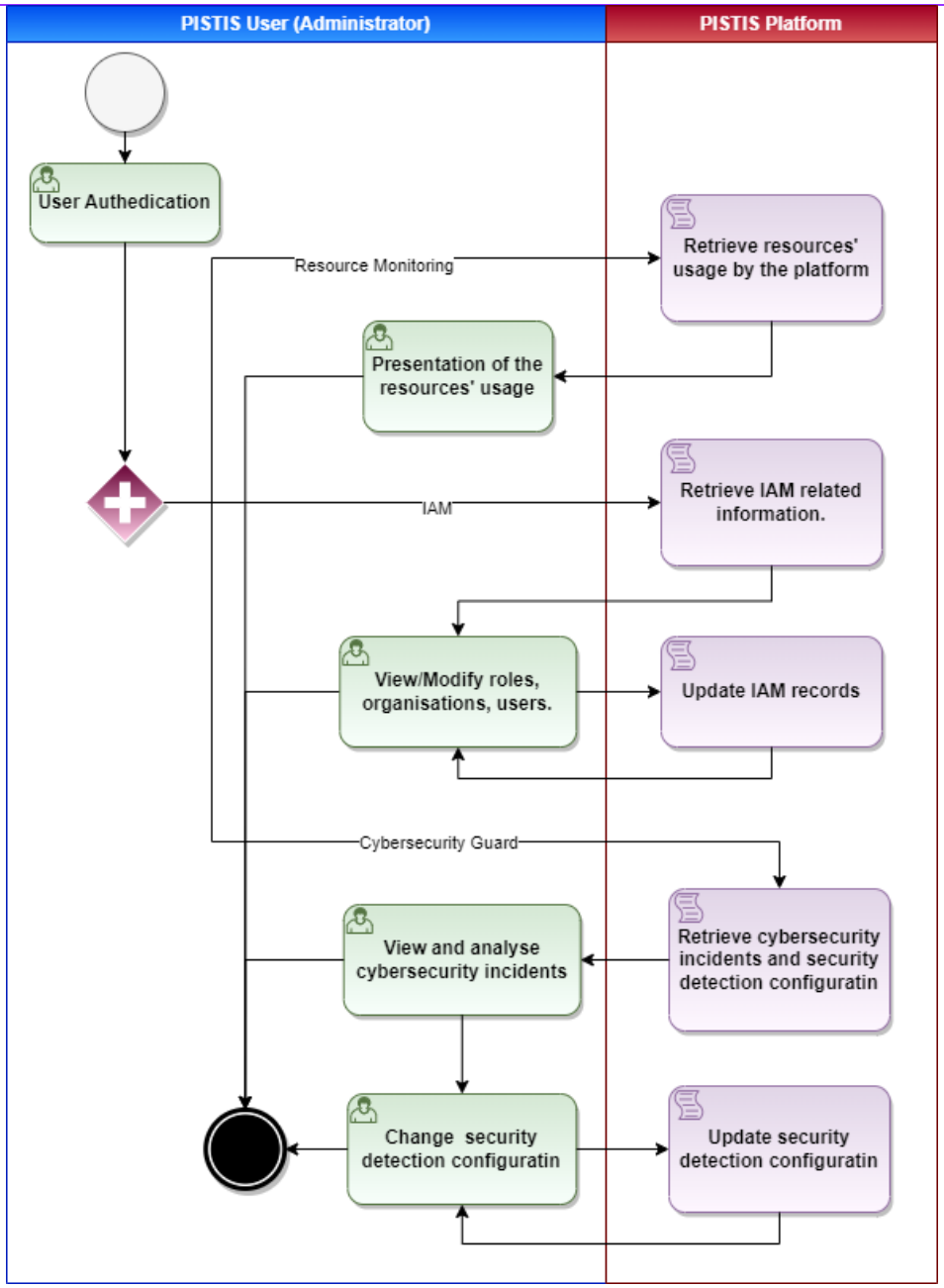
## 3.4 Support Operations User Stories

The following sections present the three user stories that support the system operations, explaining the interaction between users and PISTIS platform and the executed processes.

### 3.4.1 PISTIS platform monitor/audit

**Table 14: PISTIS platform monitor/audit User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.SOUS.01** |
| **Story Title** | **PISTIS platform monitor/audit** |
| **Story Actors** | PISTIS User (Administrator), PISTIS Platform |
| **Overview** | The operations relevant to the monitoring of the PISTIS platform status and health by the PISTIS Administrator |

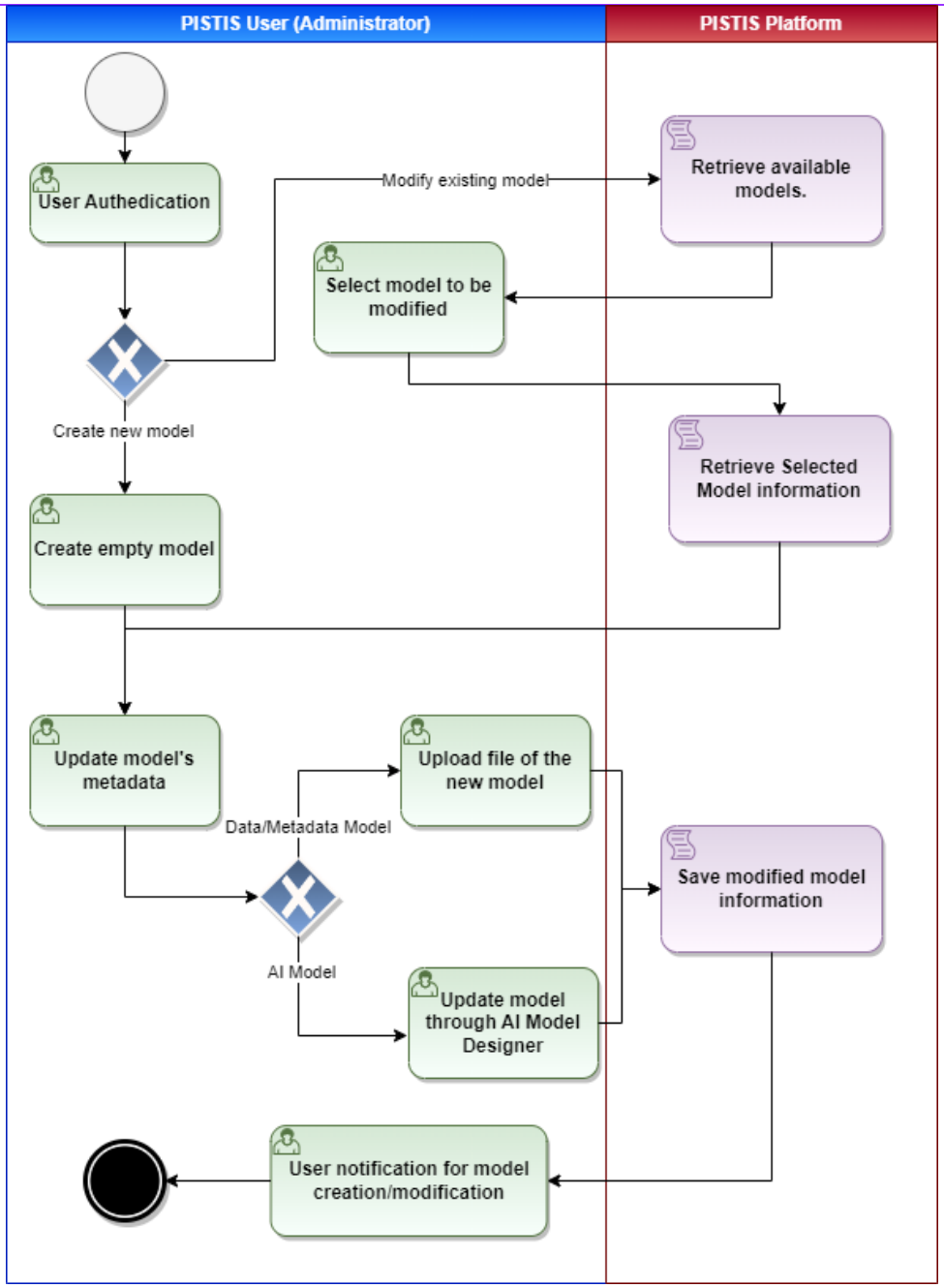
| | |
|---|---|
| **Triggers** | N/A |
| **Workflow** |  <br>**Figure 19: PISTIS platform monitor/audit process** |
| **Story Sequence** | 1. The PISTIS Administrator logs into the PISTIS Cloud Platform using his credentials and lands on the administration home page.<br>2. The PISTIS Administrator can view the dedicated dashboards to monitor the resources of the PISTIS platform, visit the page to administer the organisations/data factories and users of the PISTIS platform, and to check for cybersecurity incidents and manage the rules for detection.<br>2.1. Alt flow #1 (Resource Monitoring)<br>2.1.1. By visiting the resources page the PISTIS Administrator can view the resources used in PISTIS and the operational status of PISTIS components.<br>2.2. Alt flow #2 (IAM) |

| | |
|---|---|
| | 2.2.1. By visiting the IAM service, the PISTIS Administrator can view a list of all organisations signed up in PISTIS.<br>2.2.2. By visiting the view page of the organisation in the IAM the PISTIS Administrator can see the organisation details and the list of users of that organisation<br>2.2.3. The PISTIS Administrator can enable/disable/update/delete or view an organisation and assign attributes to the organisation.<br>2.2.4. By visiting the users page within an organisation, the PISTIS Administrator can enable/disable/update/delete or view details for a user and assign roles.<br>2.3. Alt flow #3 (Cybersecurity Guard)<br>2.3.1. The PISTIS Administrator navigates to the CyberSecurity Guard page to monitor security incidents.<br>2.3.2. The PISTIS Administrator Guard can upload a new security detection configuration file to update the existing detection rules<br>2.3.3. The PISTIS Administrator can view the error logs of the PISTIS Cloud Platform<br>2.3.4. The PISTIS Administrator can view the error logs of the connected Data Factories |
| **User Benefits** | The user will get:<br>• Graphs showcasing the resources utilisation and other related metrics (such as no of transactions, no of datasets, etc.) of the PISTIS platform.<br>• An IAM service to administer organisations and users over the whole PISTIS platform.<br>• A cybersecurity guard interface to insert detection rules and monitor incidents |
| **Challenges** | • Off-line validation checks (such as company and signatory user validation, etc.) might be necessary to enable/update the organisations onboarding PISTIS, to offer a trusted environment.<br>• Need to perform deep checks when changing user roles or removing users, to transfer ownership of operations and privileges of such users (such as wallet's access) to other users |
| **Other issues** | Information about users stored in the IAM should follow the GDPR provisions and a DPO should be assigned for any requests |

### 3.4.2 PISTIS platform configuration

**Table 15: PISTIS platform configuration User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.SOUS.02** |
| **Story Title** | **PISTIS platform configuration** |
| **Story Actors** | PISTIS Administrator |
| **Overview** | Enrichment/Population of the different global repositories (AI pre-trained models, Data, Metadata) that will be used by the different modules of the PISTIS Data Factories |

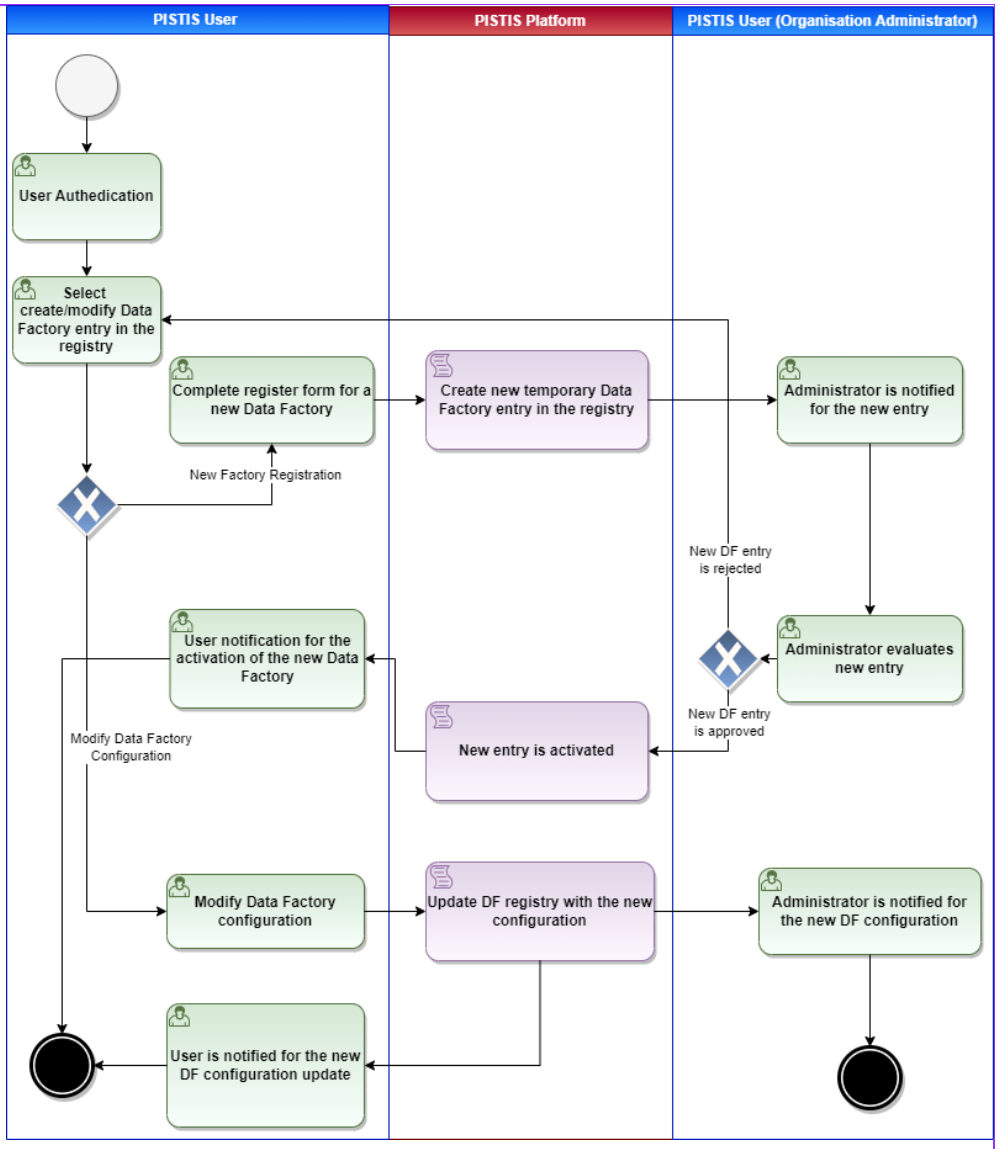
| | |
|---|---|
| **Triggers** | Introductions of changes in data models (for example when PISTIS is to support a new industry domain) or new requests for AI models will trigger this action |
| **Workflow** |  <br>**Figure 20: PISTIS platform configuration** |
| **Story Sequence** | 1. The PISTIS Administrator visits the Models Manager and sees the contents of the repositories.<br>2. The PISTIS Administrator selects the type of model they want to upload/update.<br>3. The PISTIS Administrator provides/edits the model's metadata accompanying the model they want to upload/update.<br>3.1. The PISTIS Administrator chooses the file that holds the model and uploads/updates the selected model.<br>3.2. In case of an AI model, the PISTIS Administrator logs into the AI Model Designer and designs the model. |

| | |
|---|---|
| **User Benefits** | The user gets:<br>• A facility to view the content of the Repositories for the a) Pretrained AI models, b) the Data Models, c) the Semantic Models<br>• A service to perform CRUD operation in the different reports |
| **Challenges** | Need to make all models backwards compatible when updating in order not to disrupt operations that are based on previous models |
| **Security Issues** | *N/A* |
| **Other issues** | *N/A* |

### 3.4.3 Data Factory Configuration/Setup

**Table 16: Data Factory Configuration/Setup User Story**

| | |
|---|---|
| **Story ID** | **PISTIS.SOUS.03** |
| **Story Title** | **Data Factory Configuration/Setup** |
| **Story Actors** | PISTIS User, PISTIS Administrator |
| **Overview** | Upon deploying a PISTIS Data Factory instance, this deployment should be made known via the central PISTIS Platform (PISTIS Data Factory Registry) so that it can join the overall ecosystem. |
| **Triggers** | Deployment of a new PISTIS Data Factory instance or updating of the location (e.g. IP address) of an existing PISTIS Data Factory. |

| | |
|---|---|
| **Workflow** |  <br>**Figure 21: Data Factory Configuration/Setup** |
| **Story Sequence** | 1. The PISTIS Administrator registers a new Data Factory in the PISTIS Platform.<br>2. The PISTIS User enters the IP Address of its Data Factory.<br>3. The PISTIS Administrator approves the entry of the PISTIS User.<br>4. The PISTIS User is notified of the activation of its Data Factory environment. |
| **User Benefits** | The PISTIS user gets:<br>• An interface to register the location of their PISTIS Data Factory<br>The PISTIS Administrator gets:<br>• An interface to check the status of a connection of a PISTIS Data Factory and enable/disable its connection to the network |
| **Challenges** | N/A |
| **Security Issues** | N/A |

| | |
|---|---|
| **Other issues** | N/A |

### 3.4.4 Data Pipeline Definition, Execution and Fetching

| | |
|---|---|
| **Story ID** | **PISTIS.SOUS.04** |
| **Story Title** | **Data Pipeline Definition, Execution and Fetching** |
| **Story Actors** | PISTIS User, PISTIS Administrator |
| **Overview** | The definition and subsequent execution of workflows to support data pipelines is crucial to allow the publication of data assets in the Data Catalogue located on the PISTIS Factory side. The workflow will be understood as a composition of the services offered on the factory side. |
| **Triggers** | Registering on a factory side a new PISTIS Data Asset into Factory Data Storage and Factory Data Catalogue. |

**Workflow**

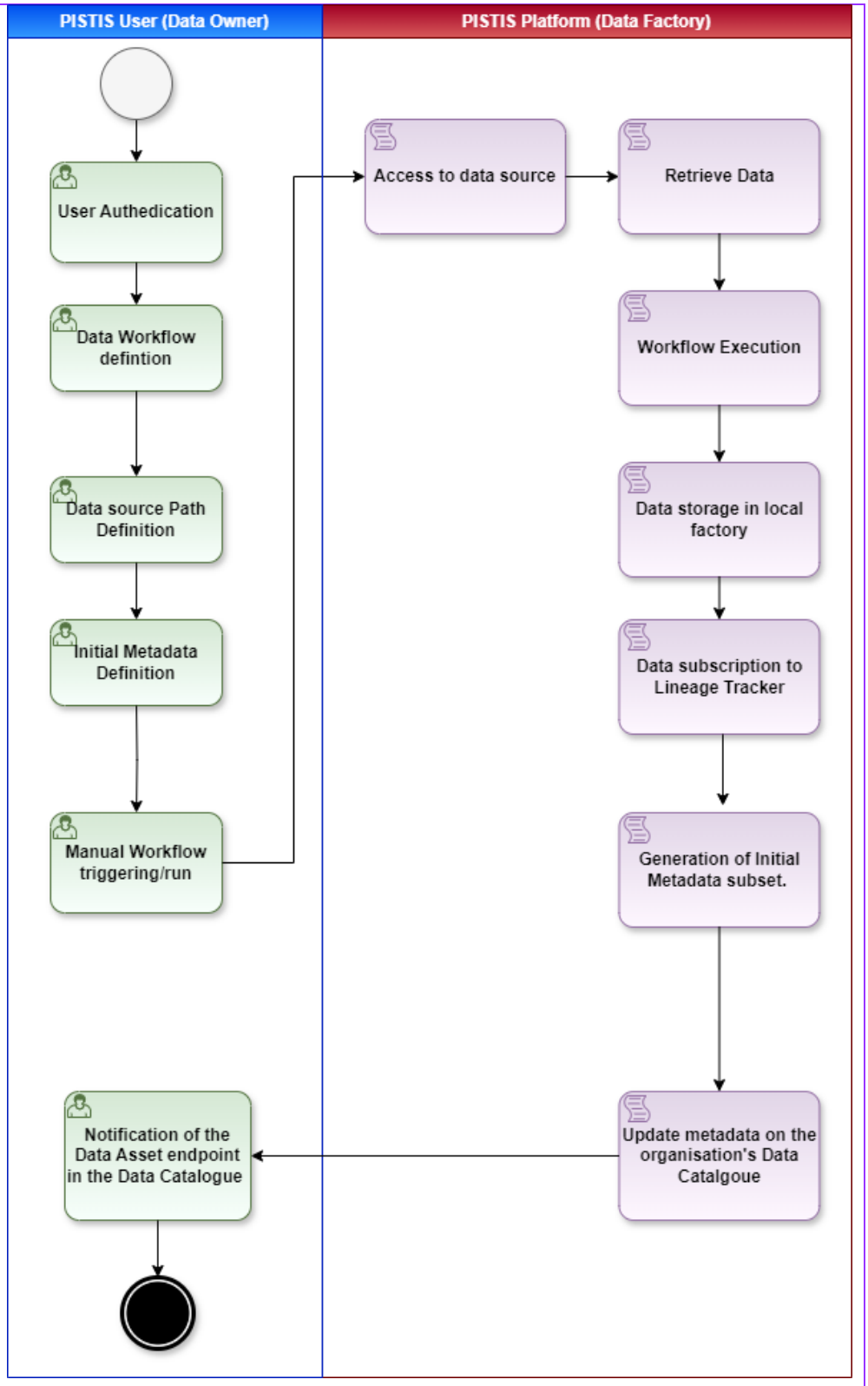



**Figure 22. PISTIS Data pipeline definition and execution.**

**Story Sequence**

1. By composing certain services or jobs, the PISTIS User creates a new workflow via the GUI.
2. PISTIS User executes the workflow defined in previous step.
3. PISTIS User fetches the execution status on the workflow on execution.
4. The GUI provides a Data Catalogue URL where the Data Asset is published if the execution was successful.

| | |
|---|---|
| **User Benefits** | The PISTIS user gets:<br>• An interface to define and run a data pipeline. In addition to this, the interface enables fetching the execution status of a concrete data pipeline.<br>The PISTIS Administrator gets:<br>• An interface to monitor the execution of a data pipeline in real time, including logs a low-level tracing. |
| **Challenges** | • Error managing during workflow execution. |
| **Security Issues** | • The workflow orchestrator should check that PISTIS User could execute each one of the services validating the access token provided by the IAM component. |
| **Other issues** | N/A |

# 4 TECHNICAL REQUIREMENTS

User Stories initially authored in D1.2, provide a high-level description of the process of the main functionalities that PISTIS will provide to its end users combining business and technology perspective. Through a comprehensive analysis of each revised user story, focusing particularly on users' benefits, challenges, issues, and leveraging technical solutions derived from state-of-the-art practices, a compilation of technical requirements has been generated and updated for D1.3. As already presented in D1.2, each technical requirement for PISTIS represents a technical-oriented feasible approach to address the business requirements expressed in the user stories.

The technical requirements for the user stories have undergone several updates and refinements in D1.3 based on feedback from the alpha implementation. In D1.3, several new requirements were introduced to address the needs of the PISTIS platform. The focus on scalability resulted in requirements aimed at handling larger datasets, accommodating higher transaction volumes, and supporting distributed environments. Additionally, requirements for interoperability were added to facilitate integration with external systems, federated data sources, and standards-based protocols like distributed ledgers. The marketplace was enhanced with new requirements for advanced pricing models, support for data bundling, and transaction templates to make the platform more user-friendly and flexible.

Some existing requirements were modified to better reflect the insights gained during the alpha phase. The requirements for access policies were updated to include more granular controls, such as attribute-level access and dynamic policy enforcement. Data transformation requirements were refined to incorporate schema evolution, ensuring compatibility with diverse and changing data structures. For user authentication and management, enhancements include the addition of two-factor authentication and improved workflows for onboarding users, ensuring a more secure and seamless experience.

Certain requirements were removed or consolidated in D1.3 to streamline the development process. Overlapping technical aspects from D1.2, such as those related to baseline security and accessibility, were merged into broader categories.

Table 17 enlists the refined requirements as identified in each user story along with additional details for each one.

**Table 17: PISTIS Technical Requirements**

| ID | Description | Explanation |
|---|---|---|
| **PISTIS.OUS.01** | **Data check-in** | |
| REQ_1 | PISTIS supports the registration of a new dataset to the ecosystem. | PISTIS platform must provide an interface or a mechanism to the user to define the new data and its metadata. |
| REQ_2 | PISTIS supports data registration from different data source type. | PISTIS platform must support multiple type of data sources such as FTP, HTTP APIs, SFTP, DB connections, etc. |
| REQ_3 | PISTIS ensures that only authorised users can register datasets. | Only authenticated and authorised users can check-in their data to the PISTIS platform. |
| REQ_4 | PISTIS stores the new dataset in an internal storage located on the organisation's site (Data Factory). | The new dataset will be stored into the PISTIS platform internal storage to be used for all the process that are defined into the PISTIS data pipeline. |

| | | |
|---|---|---|
| REQ_5 | PISTIS encrypts the dataset to ensure data confidentiality throughout its lifecycle. | The encryption process is based in a key management system linked the user's HW-based identity key managed by encryption/decryption engine (equipped with a root of trust) |
| **PISTIS.OUS.02** | **User registration and profile management** | |
| REQ_6 | PISTIS platform supports new user registration for an organisation. | A User must be able to apply for an account within PISTIS ecosystem by providing his/her necessary information and the related organisation. |
| REQ_7 | PISTIS authentication mechanisms supports integration with external Identity Providers. | A User must be able to utilize their existing account(s) in an external Identity Provider. |
| REQ_8 | PISTIS supports account management. | A User must be able to manage their account (preferences, details, and keys/certificates). |
| REQ_9 | PISTIS supports PISTIS Account binding with a Trusted CA (i.e. eIDAS). | A User's account must be able to be bind their verifiable credentials with a Trusted CA (eIDAS) |
| REQ_10 | PISTIS Platform user registration mechanism is integrated and aligned with eIDAS Trusted CA | Registration process must be integrated and aligned with eIDAS requirements |
| REQ_11 | PISTIS allows Organisation Administrator to modify a user's profile. | PISTIS Organization Administrator must be able to manage a PISTIS User profile |
| REQ_12 | PISTIS platform provides secure storage of user's account assets. | User credentials, personal details and keys/certificates must be securely stored and handled in the user's identity wallet. |
| REQ_13 | PISTIS supports 2FA for user authentication. | A PISTIS User account must offer configurable 2FA option |
| REQ_14 | PISTIS supports user's account data awareness. | A PISTIS User must be fully informed about what data is being collected during registration, how it will be used, and whether it will be shared with third parties |
| REQ_15 | PISTIS establishes secure connection for data exchange with third-party systems (i.e. external Identity Providers). | User credentials and any sensitive data must be securely stored and exchanged with the external Identity Providers |
| REQ_16 | PISTIS supports immediate role adjustment after profile update. | Any role change must be immediately effective. |
| REQ_17 | PISTIS supports monitor of account's activities and identification against security vulnerabilities. | PISTIS accounts must be protected against unauthorized access and data breaches, via continuous monitoring and on time identification of potential security vulnerabilities |
| **PISTIS.OUS.03** | **Data transformation and analysis** | |
| REQ_18 | PISTIS Platform supports user to select or define data transformations on a dataset. | PISTIS Platform should provide an interface to select or define the transformations to be applied over a dataset. A |

| | | |
|---|---|---|
| | | data transformation workflow is defined including all the appropriate information for the data transformation execution. |
| REQ_19 | PISTIS Platform supports multiple transformations on a dataset. | PISTIS Platform should allow the application of different transformations over a dataset. |
| REQ_20 | PISTIS Platform data transformation supports skip mechanism for non-applicable transformations. | PISTIS Platform should avoid apply transformations when these are not applicable for the structure of the applied dataset (e.g. numerical operations over string values, etc.) |
| REQ_21 | PISTIS Platform supports user-defined transformation jobs. | PISTIS Users can define their own transformation jobs when they built a data transformation workflow. |
| REQ_22 | PISTIS Platform supports predefined transformation job templates. | PISTS Users can define a transformation job by selecting from a list of pre-defined transformation job templates and parameterized them according to their needs. |
| **PISTIS.OUS.04** | **Data quality check** | |
| REQ_23 | PISTIS exposes a DQA User Interface, to support User interaction with the data quality process before and after running it. | The DQA UI supports the input of data quality requirements from Users (accept/modify/reject inferred data quality rules, parametrise custom data quality rules), as well as the output of data quality reports. |
| REQ_24 | PISTIS supports analysis of data and metadata quality based on standard quality criteria. | Analysis of the quality of data and metadata stored in PISTIS according to the predefined criteria and procedures. |
| REQ_25 | PISTIS data's metadata is enhanced with quality analysis results. | The quality analysis results must be stored as (enhanced) part of metadata |
| REQ_26 | PISTIS supports periodic quality checks. | It should be possible to check the quality of data on a regular basis, especially in the case of dynamically changed data (i.e. data stream or data coming from APIs) |
| REQ_27 | PISTIS platform supports user-managed data transformations based on data quality assessment recommendations. | Data quality assessment should also return suggestions for improving the quality of a dataset. A user can accept or decline them. If the user accepts them, these need to be passed to the Data Transformation module, which runs them to generate a new data asset, based on the data quality rules. Following this, the Data Quality Assessment should be run again on the (new) transformed data asset. |
| **PISTIS.OUS.05** | **Definition of access policies** | |

| | | |
|---|---|---|
| REQ_28 | PISTIS supports the enforcement of already defined GDPR checking rules. | PISTIS Platform should support the automatic extraction of such rules from GDPR related obligations written in open digital right language (ODRL) |
| REQ_29 | PISTIS Platform supports access policies editor | PISTIS platform must offer a GUI-like editor for the definition and editing of certain access rights to a PISTIS Organization Role or the creation of new policies to match PISTIS data exchange requirements |
| REQ_30 | PISTIS Platform supports high availability for Access policies editor. | Access policy editor must be available any time to all PISTIS Organization Administrators |
| REQ_31 | PISTIS Platform provides effective access policies as extension of CRUD model. | Access policies must extend CRUD model with PISTIS-specific data trading requirements (i.e., Trading, Transformation, Pricing, etc.); applicable to any PISTIS Platform feature, data stream/dataset and attribute |
| REQ_32 | PISTIS Platform allows user to define access policies precisely and granularly. | Access policies must be able to be fine-tuned on nested objects/attributes within a PISTIS resource to match PISTIS data exchange requirements |
| REQ_33 | PISTIS platform supports access policy evaluation. | Access Policy Editor must provide an effective way to evaluate an access policy before applying it to the PISTIS ecosystem. i.e. Access Policy definition must ensure that generated access policies do not overlap or do not conflict with each other. |
| REQ_34 | PISTIS platform supports Immediate enforcement of access policies modifications. | Access Policy definition must ensure immediate enforcement upon creation or modification |
| **PISTIS.OUS.06** | **Monetization of the data** | |
| REQ_35 | PISTIS allows user to create data assets for publishing in the marketplace. | The data asset shall be described using metadata and other information prior to being placed on the market |
| REQ_36 | PISTIS requires each data asset to have a monetary value. | The Data Asset should be accompanied by a PISTIS coin value. This could be also zero for freely provided data assets. |
| REQ_37 | PISTIS supports various data monetisation options. | Three different monetisation options should be offered to a data asset, and only one should be selected: Fixed/Monthly price, NFT trade, Investment Plan |
| REQ_38 | PISTIS supports the creation of user-defined data transaction contract templates. | The user must specify specific contract terms to accompany each asset sold, indicating issues such as the duration of the contract, the times a buyer can download the data, etc. |

| | | |
|---|---|---|
| REQ_39 | PISTIS supports asset's NFT Generation. | The user shall be able to generate an NFT from a data asset and have it ready for trade |
| REQ_40 | PISTIS supports asset's NFT Trade. | The user shall be able to sell the NFT of the data for the dataset to have a new owner. |
| REQ_41 | PISTIS supports asset's Investment Plan Activation. | The user shall be able to set an investment plan on a dataset, indicating the equity they are willing to offer to the interested investors, and the price they are asking for it. |
| **PISTIS.OUS.07** | **Data Valuation** | |
| REQ_42 | PISTIS platform stores all the information on actions regarding data and provides access to its components. | Metadata related to ownership, lineage, transformations, terms of use and license shall be available and stored in the PISITS Data Catalogue and the PISTIS Lineage Tracker. |
| REQ_43 | PISTIS platform stores the data valuation results of a dataset along with its metadata. | Metadata about the results of data valuation should be created or updated. This will include data value dimensions, their respective scores, the aggregate scores, and the aggregation metrics. |
| REQ_44 | PISTIS platform's data valuation considers the dataset's usage analytics. | Data valuation will include insights generated from usage analytics of a data asset. These will be retrieved from the Data Usage and Intentions Analytics, based on the answers of PISTIS Users in the dedicated dataset questionnaires. |
| REQ_45 | PISTIS platform supports deanonymisation risk analysis. | Check the risk of deanonymising a previously anonymised data asset. The user will need to define the quasi-identifiers inside the data set, run possible anonymisation and attempt deanonymisation. This will involve the Anonymisation module. |
| REQ_46 | PISTIS platform's data valuation considers GDPR compliance checker for the datasets. | Include a GDPR compliance score in the Data Valuation process. This should be retrieved from its corresponding metadata. |
| REQ_47 | PISTIS platform's data valuation supports functional utility checks on a data asset based on user-defined metrics. | To compute the functional utility of a data asset, ML models and evaluation metrics shall be made available by the PISTIS User. |
| REQ_48 | PISTIS platform provides an interpretable data valuation report. | The data value score should be easy to understand. It should be easy to deconstruct it in its composing parts, down to the lowest level of data value dimensions. Further, the PISTIS user will visualise scores at different aggregation levels, to better interpret the drivers of the value of their data. |

| | | |
|---|---|---|
| REQ_49 | PISTIS records data valuation outcomes in a secure, immutable manner to ensure traceability and auditability of valuation activities. | Provides an immutable record of data valuation results, enhancing transparency and trust in data transactions. This should also enable runtime auditing of all transactions mediated by the PISTIS platform |
| **PISTIS.OUS.08** | **Data usage and market analytics** | |
| REQ_50 | PISTIS provides detailed market insights related to data assets. | The user shall be able to have an overview of the PISTIS market, by viewing indexes and metrics that are related to specific data, filtered by attributes such as sector, data type, metadata, etc. |
| REQ_51 | PISTIS provides data usage insights for each data asset. | The user shall be able to see analytics relevant to the usage (by combining lineage and market transaction information) of a specific dataset, or relevant to a group of similar datasets. |
| REQ_52 | PISTIS provides a detailed view of the lineage of the data published in the platform. | The user shall be able to see the lineage of a dataset owned, including both a log of their actions and those of the actions of other stakeholders that have bought his asset |
| **PISTIS.OUS.09** | **Data Query** | |
| REQ_53 | PISTIS provides a UI for advanced search mechanism on the published data assets. | PISTIS platform must provide a UI that enables users to create their queries on the data assets. |
| REQ_54 | PISTIS search mechanism supports query criteria validation. | For the submission of a query, the PISTIS Platform must validate that the criteria provided follow certain language grammar and syntactic rules to avoid non-understandable queries. |
| REQ_55 | PISTIS search mechanism supports quey splitting for metadata and data. | PISTIS platform must identify the parts of a query that refer to the metadata and those that refer to the data and split them accordingly |
| REQ_56 | PISTIS search mechanism searches datasets through their metadata | PISTIS platform must offer a mechanism that is able to discover datasets using their metadata. |
| REQ_57 | PISTIS search mechanism searches datasets on their data. | PISTIS platform must offer a mechanism that is able to discover datasets by matching the stored data. |
| REQ_58 | PISTIS search mechanism supports metadata and data search results merging. | PISTIS platform must have a way to combine the results returned by two different search mechanisms in a unified list. |
| REQ_59 | PISTIS search mechanism supports search results re-ranking | PISTIS platform must sort the results returned by the search mechanisms in a descending order of relevance. |
| REQ_60 | PISTIS performs data queries to allow secure access to | PISTIS platform must offer a process that enables users to query encrypted data while ensuring that both the query |

| | | |
|---|---|---|
| | metadata without exposing sensitive information. | and metadata remain secure, maintaining privacy during the data retrieval process. |
| **PISTIS.OUS.10** | **Data Transaction** | |
| REQ_61 | PISTIS generates smart contract for each data transaction. | The PISTIS Platform shall create a smart contract based on the terms chosen by the PISTIS User (Provider) before allowing any data transaction. |
| REQ_62 | PISTIS validates data transaction policies and conditions to ensure compliance before executing data exchanges. | Ensures that all data transactions adhere to pre-defined policies, protecting both parties involved and preventing unauthorized or incorrect exchanges. |
| REQ_63 | PISTIS smart contract mechanism requires the digital sign of the smart contract from the buyer. | The PISTIS User (Consumer) shall view, accept, and sign the smart contract to enable the transfer of the data |
| REQ_64 | PISTIS supports Peer-to-Peer Data Exchange | PISTIS platform should enable direct peer-to-peer data exchange between Data Providers and Data Consumers through a dedicated Data Factory Connector. |
| REQ_65 | PISTIS supports Secure Data Encryption on data transactions. | The platform should employ robust encryption algorithms to protect the data during transfer. Encryption keys should be securely managed and accessible only to authorized parties. |
| REQ_66 | PISTIS supports digital wallet for financial transactions. | Users should be able to manage their wallets, including cash in, balance checking, transaction history and cash out. |
| REQ_67 | PISTIS supports multi-signatures on a smart contract. | Multi-signature technology should be supported allowing multiple users of the same organization to collectively authorize a transaction by requiring multiple signatures to approve and execute a transfer. |
| REQ_68 | PISTIS verifies compliance with established data trading policies and successful value transaction during data exchanges to ensure conditions are met before proceeding. | The verification step guarantees that all required conditions are satisfied before executing data exchanges. |
| REQ_69 | PISTIS automates the execution of smart contracts during data transactions to ensure proper enforcement of trading agreements. | Automating the execution of smart contracts ensures that all predefined conditions are met without manual intervention, guaranteeing compliance, and building trust between parties. |
| **PISTIS.OUS.11** | **Data exchange monitor/audit** | |

| | | |
|---|---|---|
| REQ_70 | PISTIS supports data transaction status notification. | PISTIS User (Owner/Consumer) must be notified about the status of the transaction. |
| REQ_71 | PISTIS ensures the only authorised users are notified for a data transaction. | Only authenticated and authorised PISTIS Users (Owner/Consumer) should be notified about the data transaction status. |
| REQ_72 | PISTIS stores data transaction activities for auditing purpose. | Logging transactions ensures that all automated processes related to data exchanges are traceable, providing an audit trail that guarantees compliance. |
| REQ_73 | PISTIS provides detailed history of a data transaction. | Only authenticated and authorised PISTIS Users (Auditor/Certification Body) should have access to the transaction history of a specific dataset. |
| REQ_74 | PISTIS records details of each data and value transaction in a secure and immutable manner to guarantee accountability and traceability. | Ensures that all data transactions are permanently recorded, providing transparency, accountability, and the ability to audit data transactions. |
| **PISTIS.SOUS.01** | **PISTIS platform monitor/audit** | |
| REQ_75 | PISTIS supports Resource Monitoring System | The PISTIS platform must provide the PISTIS Administrator with a view to monitoring all resources relevant to the infrastructure and to the operation of the system. |
| REQ_76 | PISTIS supports organisation management via IAM | The PISTIS platform must provide the PISTIS Administrator with an interface for the IAM, to manage organisations and their attributes, and users and their roles. |
| REQ_77 | PISTIS supports users' management via IAM. | The PISTIS platform must provide the PISTIS Administrator with an interface for the IAM, to users and their roles. |
| REQ_78 | PISTIS supports cybersecurity Monitoring. | The PISTIS platform must provide the PISTIS Administrator with an interface for the monitoring of possible cybersecurity incidents |
| REQ_79 | PISTIS supports user-defined cybersecurity Policies Updates. | The PISTIS platform must provide the PISTIS Administrator with an interface for the updating of the platform cybersecurity policies |
| REQ_80 | PISTIS ensures that audit logs related to any operation being performed by PISTIS platform is managed with a high degree of integrity. . | Secure storage of audit logs, exposed only to authorized users ensures confidentiality and high granularity access controls that safeguarding data, metadata, audit logs integrity. |
| **PISTIS.SOUS.02** | **PISTIS platform configuration** | |
| REQ_81 | PISTIS supports CRUD Operations on data/metadata/AI models repositories. | The PISTIS Platform must provide the PISTIS Administrator and with an |

| | | |
|---|---|---|
| | | interface to perform CRUD operations in the different Global Repos |
| **PISTIS.SOUS.03** | **Data Factory Configuration/Setup** | |
| REQ_82 | PISTIS provides an interface to organisation's administrators to Register Data Factory location to PISTIS ecosystem. | The PISTIS Platform must allow users to specify the location and the configuration of their Data Space Factory through the registration of its IP Address. |
| REQ_83 | PISTIS notifies platform administration for the request of new PISTIS Data Factory connection entries | The PISTIS Platform shall notify the PISTIS Administrator for a new connection entry request made in the central PISTIS Platform. |
| REQ_84 | PISTIS allows platform administration to evaluate and activate a new PISTIS Data Factory connection request. | The PISTIS Platform must ask the PISTIS Administrator to accept/activate a new PISTIS Data Factory connection registration. This step is mandatory for the activation of the Data Factory. |
| REQ_85 | PISTIS notifies organisation administrator when a new PISTIS Data Factory connection was activated. | The PISTIS Platform shall notify the organisation's administrator for the activation of the requested new PISTIS Data Factory connection. |
| **PISTIS.SOUS.04** | **Data Pipeline Definition, Execution and Fetching** | |
| REQ_86 | PISTIS Data Factory enables definition and execution of jobs templates to support data pipeline related tasks. | PISTIS Data Factory supports data pipeline job definitions and execution as part of a PISTIS workflow. |
| REQ 87 | PISTIS Data Factory facilitates job orchestration to support data pipelines/workflows. | PISTIS Data Factory supports data pipeline job orchestration. |
| REQ 88 | PISTIS Data Factory supports workflows/data pipelines fetching. | PISTIS Data Factory provides a fetching mechanism to validate the execution status of workflows/data pipelines. |

# 5 MVP DESIGN

As already presented in D1.2, the MVP represents the fundamental version of a product that effectively addresses the core problem or fulfils the essential needs of the users. The principles of the methodology remain unaltered compared to D1.2 but the final outcome is reevaluated based on the updated requirements as they are extracted from the refined user stories in D1.3. In favour of completeness and clarity we retain the unaltered sections of the section.

The idea of MVP in PISTIS is to release a version of the envisioned PISTIS framework that is both desirable and operational for the end users and can be completed during the project implementation's period. **The PISTIS MVP design, in the context of WP1, is a set of requirements that will be addressed in the final delivered version of the PISTIS platform.**

The process of determining the requirements defining the PISTIS MVP involves the employment of the MoSCoW method[4] . This method is used to set priorities on the list of requirements derived from the analysis of User Stories. MoSCoW is an acronym representing a prioritization technique used in project management and software development. It can be used to categorise and prioritise requirements or tasks based on their importance. According to the MoSCoW method each requirement is characterised by one of the following priorities:

- **Must-haves (M)**: These are the non-negotiables, the core features or requirements that must be included in the final product. Without these, the project would not be considered a success.
- **Should-haves (S)**: These are important but not critical. They add significant value to the project, and ideally, they should be included unless there are constraints that prevent it.
- **Could-haves (C)**: These are nice-to-have features. They are desirable but not crucial for the project's success. If there's time and resources, they can be included, but they can also be pushed to future releases.
- **Won't-haves (W)**: These are deliberately excluded from the current scope of work. They might be considered for future phases or versions, but they are not part of the immediate plan.

To prioritise the updated PISTIS requirements for MVP, the followed methodology included the evaluation of each requirement by two groups:

- **End-users Group**: This group includes the partners that will operate the final PISTIS framework during the pilots. This group expresses the opinion of the end-users, and the evaluation is based on the business and operational perspective.
- **Technical Group:** This group comprises partners who hold the essential technical expertise necessary for implementing the PISTIS MVP. Their primary role involves evaluating and expressing the technical feasibility of meeting each requirement outlined for the final delivered product.

The next step of PISTIS MVP definition methodology was the presentation and the discussion of the requirements to the two groups. Subsequently a prioritisation voting survey was carried out. Each participant assigned an integer to each requirement according to a voting scale 1-4. Each number represented one of the priorities of the MoSCoW method. 1=Won't Have, 2=Could Have, 3=Should Have, 4=Must Have.

[4] Kravchenko, Tatiana, Tatiana Bogdanova, and Timofey Shevgunov. "Ranking requirements using MoSCoW methodology in practice." Computer Science On-line Conference. Cham: Springer International Publishing, 2022.

The evaluation score for prioritizing each requirement was derived through the use of a weighted mean for aggregating the voting results. The calculation of the prioritization evaluation score for each requirement follows a specific equation.

$$Priority = Round\left(\frac{a*\sum_{i=0}^{N} VE_i + b*\sum_{j=0}^{M} VT_j}{a*N+b*M}\right)$$

Where:

$N$ = number of participants from End-Users Group

$VE_i$= the vote of the participant $i$ of the End-Users Group

$M$ = number of participants from Technical Group

$VT_j$= the vote of the participant $j$ of the Technical Group

$a$, $b$ = the weight of the vote from each group

The application of weights in the votes of each group was decided so that the final score will emphasize more to the needs of end-users of the product. Finally, the calculated mean was rounded so the final score will represent the closest integer number to the MoSCoW numerical scale. For the weight variables, the balanced value of a=1.2 and b=0.8 was used and the results are presented in the following table.

**Table 18: Requirements Prioritising Vote Results**

| Requirement | Evaluation | Importance | Requirement | Evaluation | Importance | Requirement | Evaluation | Importance | Requirement | Evaluation | Importance |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| REQ_1 | 4 | Must Have | **REQ_23** | 3 | Should Have | **REQ_45** | 3 | Should Have | **REQ_67** | 2 | Could Have |
| REQ_2 | 4 | Must Have | **REQ_24** | 3 | Should Have | **REQ_46** | 3 | Should Have | **REQ_68** | 3 | Should Have |
| REQ_3 | 4 | Must Have | **REQ_25** | 3 | Should Have | **REQ_47** | 3 | Should Have | **REQ_69** | 3 | Should Have |
| REQ_4 | 2 | Could Have | **REQ_26** | 2 | Could Have | **REQ_48** | 3 | Should Have | **REQ_70** | 3 | Should Have |
| REQ_5 | 3 | Should Have | **REQ_27** | 2 | Could Have | **REQ_49** | 3 | Should Have | **REQ_71** | 4 | Must Have |
| REQ_6 | 4 | Must Have | **REQ_28** | 2 | Could Have | **REQ_50** | 3 | Should Have | **REQ_72** | 4 | Must Have |
| REQ_7 | 2 | Could Have | **REQ_29** | 2 | Could Have | **REQ_51** | 3 | Should Have | **REQ_73** | 4 | Must Have |
| REQ_8 | 3 | Should Have | **REQ_30** | 2 | Could Have | **REQ_52** | 3 | Should Have | **REQ_74** | 3 | Should Have |
| REQ_9 | 3 | Should Have | **REQ_31** | 2 | Could Have | **REQ_53** | 4 | Must Have | **REQ_75** | 3 | Should Have |
| REQ_10 | 3 | Should Have | **REQ_32** | 2 | Could Have | **REQ_54** | 3 | Should Have | **REQ_76** | 3 | Should Have |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| REQ_11 | 3 | Should Have | **REQ_33** | 2 | Could Have | **REQ_55** | 3 | Should Have | **REQ_77** | 3 | Should Have |
| REQ_12 | 4 | Must Have | **REQ_34** | 3 | Should Have | **REQ_56** | 3 | Should Have | **REQ_78** | 3 | Should Have |
| REQ_13 | 2 | Could Have | **REQ_35** | 3 | Should Have | **REQ_57** | 3 | Should Have | **REQ_79** | 2 | Could Have |
| REQ_14 | 3 | Should Have | **REQ_36** | 3 | Should Have | **REQ_58** | 3 | Should Have | **REQ_80** | 2 | Could Have |
| REQ_15 | 3 | Should Have | **REQ_37** | 3 | Should Have | **REQ_59** | 3 | Should Have | **REQ_81** | 3 | Should Have |
| REQ_16 | 3 | Should Have | **REQ_38** | 3 | Should Have | **REQ_60** | 3 | Should Have | **REQ_82** | 3 | Should Have |
| REQ_17 | 4 | Must Have | **REQ_39** | 2 | Could Have | **REQ_61** | 3 | Should Have | **REQ_83** | 3 | Should Have |
| REQ_18 | 3 | Should Have | **REQ_40** | 2 | Could Have | **REQ_62** | 3 | Should Have | **REQ_84** | 3 | Should Have |
| **REQ_19** | 3 | Should Have | **REQ_41** | 2 | Could Have | **REQ_63** | 3 | Should Have | **REQ_85** | 3 | Should Have |

According to the results, twenty-five (11) requirements must be addressed in the final product, forty-four (60) are important but not crucial and seven (17) are not necessary.

The PISTIS MVP will be a platform that addresses, in its final version, sixty-nine (71) technical requirements with Must-Have and Should-Have priority.

The following table enlists the final set of the PISTIS MVP requirements sorted by their prioritisation.

**Table 19: PISTIS MVP Requirements**

| ID | Description | Priority |
|---|---|---|
| **REQ_1** | PISTIS supports the registration of a new dataset to the ecosystem. | Must Have |
| **REQ_2** | PISTIS supports data registration from different data source type. | Must Have |
| **REQ_3** | PISTIS ensures that only authorised users can register datasets.Only authenticated and authorised users can check-in their data to the PISTIS platform. | Must Have |
| **REQ_6** | PISTIS platform supports new user registration for an organisation. | Must Have |
| **REQ_12** | PISTIS platform provides secure storage of user's account assets. | Must Have |
| **REQ_17** | PISTIS supports monitor of account's activities and identification against security flaws. | Must Have |
| **REQ_53** | PISTIS supports monitor of account's activities and identification against security flaws. | Must Have |
| **REQ_64** | PISTIS supports Peer-to-Peer Data Exchange | Must Have |
| **REQ_71** | PISTIS ensures the only authorised users are notified for a data transaction. | Must Have |

| | | |
|---|---|---|
| **REQ_72** | PISTIS stores data transaction activities for auditing purpose. | Must Have |
| **REQ_73** | PISTIS provides detailed history of a data transaction. | Must Have |
| **REQ_5** | PISTIS encrypts the dataset to ensure data confidentiality throughout its lifecycle. | Should Have |
| **REQ_8** | PISTIS supports account management. | Should Have |
| **REQ_9** | PISTIS supports PISTIS Account binding with a Trusted CA (i.e. eIDAS). | Should Have |
| **REQ_10** | PISTIS Platform user registration mechanism is integrated and aligned with eIDAS Trusted CA | Should Have |
| **REQ_11** | PISTIS allows Organisation Administrator to modify a user's profile. | Should Have |
| **REQ_14** | PISTIS supports user's account data awareness. | Should Have |
| **REQ_15** | PISTIS establishes secure connection for data exchange with third-party systems (i.e. external Identity Providers). | Should Have |
| **REQ_16** | PISTIS supports immediate role adjustment after profile update. | Should Have |
| **REQ_18** | PISTIS Platform supports user to select or define data transformations on a dataset. | Should Have |
| **REQ_19** | PISTIS Platform supports multiple transformations on a dataset. | Should Have |
| **REQ_20** | PISTIS Platform data transformation supports skip mechanism for non-applicable transformations. | Should Have |
| **REQ_21** | PISTIS Platform supports user-defined transformation jobs. | Should Have |
| **REQ_22** | PISTIS Platform supports pre-defined transformation job templates. | Should Have |
| **REQ_23** | PISTIS exposes a DQA User Interface, to support User interaction with the data quality process before and after running it . | Should Have |
| **REQ_24** | PISTIS supports analysis of data and metadata quality based on standard quality criteria. | Should Have |
| **REQ_25** | PISTIS data's metadata is enhanced with quality analysis results. | Should Have |

| | | |
|---|---|---|
| **REQ_34** | PISTIS platform supports Immediate enforcement of access policies modifications. | Should Have |
| **REQ_35** | PISTIS allows user to create data assets for publishing in the marketplace. | Should Have |
| **REQ_36** | PISTIS requires each data asset to have a monetary value. | Should Have |
| **REQ_37** | PISTIS supports various data monetisation options. | Should Have |
| **REQ_38** | PISTIS supports the creation of user-defined data transaction contract templates. | Should Have |
| **REQ_42** | . PISTIS platform stores all the information on actions regarding data and provides access to its components. | Should Have |
| **REQ_43** | PISTIS platform stores the data valuation results of a dataset along with its metadata. | Should Have |
| **REQ_44** | PISTIS platform's data valuation considers the dataset's usage analytics. | Should Have |
| **REQ_45** | PISTIS platform supports deanonymisation risk analysis. | Should Have |
| **REQ_46** | PISTIS platform's data valuation considers GDPR compliance checker for the datasets. | Should Have |
| **REQ_47** | PISTIS platform's data valuation supports functional utility checks on a data asset based on user-defined metrics. | Should Have |
| **REQ_48** | PISTIS platform provides an interpretable data valuation report. | Should Have |
| **REQ_49** | PISTIS records data valuation outcomes in a secure, immutable manner to ensure traceability and auditability of valuation activities. | Should Have |
| **REQ_50** | PISTIS provides detailed market insights related to data assets. | Should Have |
| **REQ_51** | PISTIS provides data usage insights for each data asset. | Should Have |
| **REQ_52** | PISTIS provides a detailed view of the lineage of the data published in the platform. | Should Have |
| **REQ_54** | PISTIS search mechanism supports query criteria validation. | Should Have |
| **REQ_55** | PISTIS search mechanism supports quey splitting for metadata and data. | Should Have |

| | | |
|---|---|---|
| **REQ_56** | PISTIS search mechanism searches datasets through their metadata | Should Have |
| **REQ_57** | PISTIS search mechanism searches datasets on their data. | Should Have |
| **REQ_58** | PISTIS search mechanism supports metadata and data search results merging. | Should Have |
| **REQ_59** | PISTIS search mechanism supports search results re-ranking | Should Have |
| **REQ_60** | PISTIS performs data queries to allow secure access to metadata without exposing sensitive information. | Should Have |
| **REQ_61** | PISTIS generates smart contract for each data transaction. | Should Have |
| **REQ_62** | PISTIS validates data transaction policies and conditions to ensure compliance before executing data exchanges. | Should Have |
| **REQ_63** | PISTIS smart contract mechanism requires the digital sign of the smart contract from the buyer. | Should Have |
| **REQ_65** | PISTIS supports Secure Data Encryption on data transactions. | Should Have |
| **REQ_66** | PISTIS supports digital wallet for financial transactions. | Should Have |
| **REQ_68** | PISTIS verifies compliance with established data trading policies and successful value transaction during data exchanges to ensure conditions are met before proceeding. | Should Have |
| **REQ_69** | PISTIS automates the execution of smart contracts during data transactions to ensure proper enforcement of trading agreements. | Should Have |
| **REQ_70** | PISTIS supports data transaction status notification. | Should Have |
| **REQ_74** | PISTIS records details of each data and value transaction in a secure and immutable manner to guarantee accountability and traceability. | Should Have |
| **REQ_75** | PISTIS supports Resource Monitoring System | Should Have |
| **REQ_76** | PISTIS supports organisation management via IAM | Should Have |

| | | |
|---|---|---|
| **REQ_77** | PISTIS supports users' management via IAM. | Should Have |
| **REQ_78** | PISTIS supports cybersecurity Monitoring. | Should Have |
| **REQ_81** | PISTIS supports CRUD Operations on data/metadata/AI models repositories. | Should Have |
| **REQ_82** | PISTIS provides an interface to organisation's administrators to Register Data Factory location to PISTIS ecosystem. | Should Have |
| **REQ_83** | PISTIS notifies platform administration for the request of new PISTIS Data Factory connection entries | Should Have |
| **REQ_84** | PISTIS allows platform administration to evaluate and activate a new PISTIS Data Factory connection request. | Should Have |
| **REQ_85** | PISTIS notifies organisation administrator when a new PISTIS Data Factory connection was activated. | Should Have |
| **REQ_86** | PISTIS Data Factory enables definition and execution of jobs templates to support data pipeline related tasks. | Should Have |
| **REQ_87** | PISTIS Data Factory facilitates job orchestration to support data pipelines/workflows. | Should Have |
| **REQ_88** | PISTIS Data Factory supports workflows/data pipelines fetching. | Should Have |

# 6 GENERIC TERMS OF PISTIS DATA SHARING CONTRACTS

Having identified the relevant technical requirements for PISTIS and having defined the MVP, it is necessary to consider the contractual framework that practically organises or governs (1) the exchange or (2) collaboration and co-creation of datasets on the PISTIS data marketplace. For this purpose, two sets of non-exhaustive model generic terms have been drafted, each comprising examples of clauses that the partners can adopt, tailor, amend, or further negotiate during the demonstrations. Making use of and relying on certain model generic terms can have a number of advantages, including legal certainty, fairness and security, all constituting essential components in building users' *trust* in the platform.

The contents of this section has been prepared and is closely related to the discussion concerning legal issues related to smart contracts in section 4.3 of Deliverable D1.1 – PISTIS Operation Principles and Context Detailing, as well as to activities carried out in the context of Task T1.3 – GDPR, Cross-Border Legal Aspects and Contracts Definition for Data Sharing and Trading. As the model generic terms have been prepared to aid the Smart Contract Checker tool, within the context of PISTIS architecture, the work carried out under this section can fall under the *Data Peer-to-Peer Transfer Gateway*, as shown below.

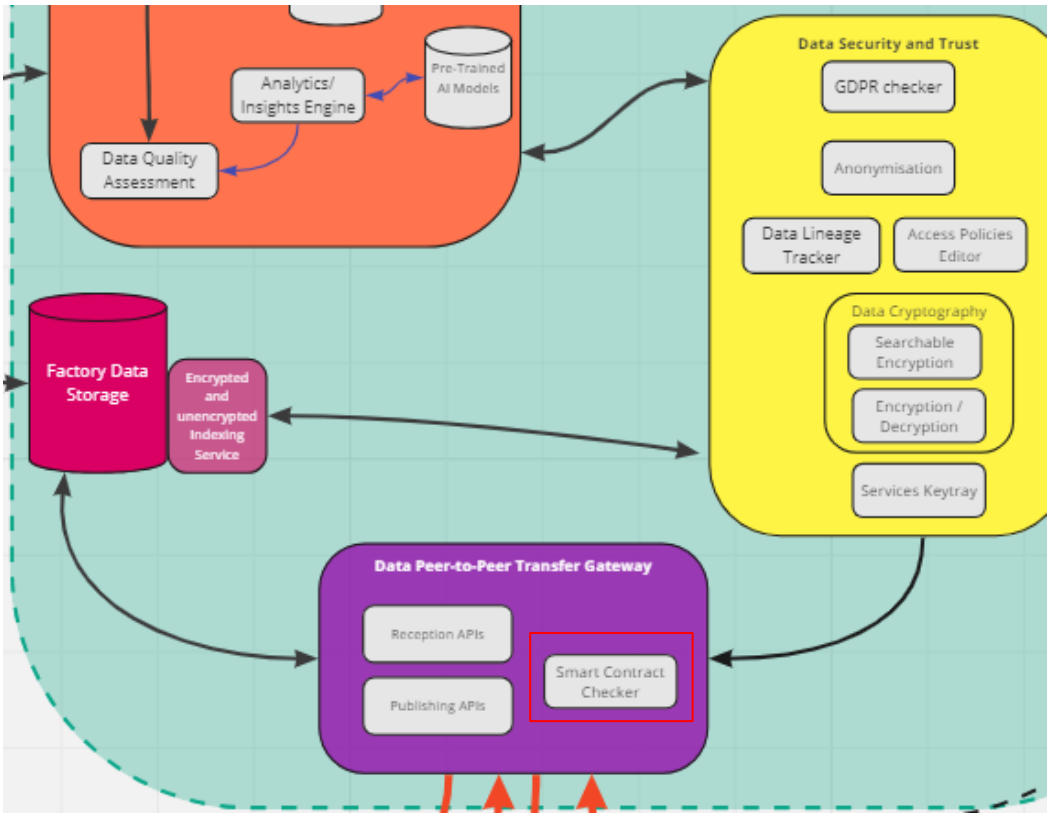



**Figure 23: Positioning the Smart Contract Checker within PISTIS architecture**

More specifically, the following two sets of model generic terms have been prepared:

1. *PISTIS Generic Terms of Data Sharing* set out in Annex 1, intended for cases of data sharing between a data provider on one hand and a data recipient on the other; and
2. *PISTIS Generic Terms of Data Sharing and Joint Dataset Creation Agreement* set out in Annex 2, intended for cases of multiple parties' collaboration on the creation of a common dataset, by each party contributing their respective data and resources.

The model generic terms have been prepared for the PISTIS demonstrator partners only and solely to be used for the demonstration of business-to-business data transactions on the BETA version of the PISTIS data marketplace, which has been under development and testing at the time of writing. Their readiness and appropriateness for use in any other context falls outside the scope of this deliverable and the related activities.

They are part of the seven elements of the PISTIS data sharing contract framework, namely (1) matchmaking of users, (2) data check-in and asset description, (3) detailed overview of dataset, (4) analysis of data usage and license, (5) data valuation, (6) determining data-sharing terms including selection of pre-defined terms, and (7) execution of transactions including provision of payment and delivery of dataset. This means that these model terms can be used and become part of any data-sharing arrangements and agreements that will be formed on the PISTIS data marketplace between data providers and data recipients.

The model generic terms of data sharing provide an aid and explanations, and are for guidance, inspiration and information purposes only. PISTIS demonstrator partners are free to change the terms as they deem appropriate for their respective use cases. If the model generic terms are taken in, used and applied, it is recommended that they are to be amended in accordance with the context of each relevant context. The demonstrator partners are not obliged to use these model terms for their data sharing and collaboration arrangements; they merely aim to enable and facilitate concrete data-sharing and collaboration arrangements and operations between the demonstrator partners in their various roles, such as data provider, data recipient, or otherwise.

Importantly, the terms are *generic*. This means that each comprises a set of non-exhaustive, proposed model terms to be used for data sharing transactions or joint dataset collaboration, but neither considers any specific arrangements that may be relevant between certain specific data providers and data recipients. Given that the data sharing transactions will be carried out in different sectors, such as mobility, traffic safety, and energy, and become subject to locally applicable rules, the demonstrator partners are advised to always consider any sector-specific legal requirements applicable to their use cases in the relevant jurisdiction. The relevant contracting partners shall always remain responsible for applying these model terms as well as adding any additional provisions to them.

In any case, any use of or reliance on either model document provided in Annex 1 and Annex 2 should not be without obtaining qualified legal advice and assistance before executing a data-sharing transaction on the PISTIS data marketplace or entering a joint dataset creation arrangement to ensure that the terms relied upon are appropriate for the intended purpose. Neither the PISTIS consortium nor any individual partner shall be liable for any loss, whether direct, indirect, special, or consequential, suffered by any party because of or in connection with the use of or reliance on these model generic terms.

Finally, these generic terms have been drafted with the aim of being as balanced and fair as possible, i.e. without favouring either contracting party. Whilst the partners may tailor, amend and negotiate the generic terms, doing so may give rise to a contractual imbalance or unfairness. It is worth noting that that the European Commission's *Study on model contract terms and fairness control in data sharing and in cloud contracts and on data access rights*[5] has found that unfair model contract terms

[5] European Commission. "Study on model contract terms and fairness control in data sharing and in cloud contracts and on data access rights." 2022. Available at https://op.europa.eu/o/opportal-service/download-handler?identifier=dfb3a486-e6d4-11ec-a534-01aa75ed71a1&format=PDF&language=en&productionSystem=cellar&documentIdentifier=dfb3a486-e6d4-11ec-a534-01aa75ed71a1.0001, accessed on 3 December 2024

(often caused by an imbalance of the parties' negotiating power or lack of competition) are considered a major challenge in arrangements concerning access to and use of data, which, in turn, may lead to loss to society not just in terms of economic benefits, but also regarding innovation and societal gains. From an economic perspective itself, for example, the study has found current levels of data sharing as "suboptimal," with "optimal" policies having the potential to increase expected profits across the bloc between 2021 and 2030 by approx. 75%.[6] Therefore, whilst departing from the generic terms remains each partner's discretion, partners are advised to consider the potential impact of any such step carefully and holistically

[6] Ibid., pp 8

# 7 TECHNOLOGY RADAR

This section contains a list of key external initiatives and projects that are relevant to the core functionalities of the PISTIS project and provides details about how each external initiative is aligned with the work of PISTIS. Specifically, it highlights whether a given initiative is aligned with PISTIS in some way (and if so, how it is aligned); or not (and if so, why not). The Technology Radar contributes to state of the art analysis delivered in WP2 and WP3, as well as highlighting how core functionality in the PISTIS platform is leveraging, aligning with, or contributing to external initiatives.

## 7.1 DATA MANAGEMENT

### 7.1.1 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, and using specialized repositories for storing data models, metadata, data transformations. The section also addresses initiatives related to data quality management and data lineage tracking, which allows organizations to track how data is used, providing valuable insights into data flow and usage patterns.

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.

**Table 20: Initiatives related to ETL**

| # | INITIATIVE/ PROJECT NAME | SHORT DESCRIPTION AND RELEVANCE FOR PISTIS |
|---|---|---|
| 1 | Apache NiFi [7] | 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.<br>NiFi was evaluated in the first iteration of PISTIS to enable data-driven workflows. However, it was not finally used to implement the data management building block, as the nature of the processes carried out in PISTIS, building pipelines of components rather than pure data flows, are more suitable for tools for workflow management such as Apache Airflow. |
| 2 | Apache Airflow | Airflow is a platform for generating, schedule and manage workflows by the Apache Foundation. Airflow is scalable (using queues and multiple worker nodes), programmable (Python) and extensible () framework, enabling the definition and scheduling of complex and lean pipelines.<br><br>PISTIS is using Airflow to define the workflows associated to the data ingestion, transformation, enrichment and storage. Components such as the Job Configurator or the insights Generator, just to name a few, make use or are invoked using pipelines defined Airflow in the background. Airflow gives flexibility and robustness to PISTIS workflows, as well as extensibility for future enhancements. |

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

| | | |
|---|---|---|
| 3 | DataVaults [8] | 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.<br>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, has been extended and used in PISTIS in the analytics engine. |
| 4 | Glass [9] | 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 [10] | 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.<br>The strategy for data transformation and enrichment described in i3-MARKET may be applicable to the PISTIS project. Of particular interest is work relating to 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 assist the development and implementation of (meta)data management systems and registries / catalogues for managing information and metadata descriptions.<br>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 [11] | 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 |

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

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

| | | |
|---|---|---|
| | | 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.<br>This project is developing a framework for 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. However, NiFi was not used in PISTIS, therefore the relevance of MARVEL for PISTIS does no longer stand. |
| 7 | PharmaLedger [12] | 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 [13] | Piveau Consus is a subproject of the Fraunhofer FOKUS open-source project piveau. Piveau Consus provides you a highly performant and 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 [14] | 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 policy makers, by means of several ways to visualize the resultant data has been provided by the project.<br>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. However, there is no direct translation of PolicyCloud technical features to PISTIS. |
| 10 | TANGO [15] | 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.<br>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 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. The MLOps component of TANGO uses similar technology as the one used in PISTIS (MLFlow). |
| 11 | W3C [16] | 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 [17] | 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 [18] | OpenLineage is an open-source lineage tracking project, providing an open standard for meta data- and lineage collection. It is integrable flexibly while offering its services via various interfaces, depending on the needs of the user. |

### 7.1.2 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 21: Initiatives related to privacy, security and trust**

| # | INITIATIVE/ PROJECT NAME | SHORT DESCRIPTION AND RELEVANCE FOR PISTIS |
|---|---|---|
| 1 | DataVaults [19] | 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 individuals based on the shared data.<br>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 [20] [21] | 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 [22] | 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 [23] | 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. |
| 5 | EU Project/ PRISMACLOUD[24] | 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. |
| 6 | EU Project/ LIGHTest [25] | 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. |
| 7 | EU Project / ARIES [26] | 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. |
| 8 | EU Project / CREDENTIAL[27] | 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. |
| 9 | Initiative / OpenDP [28] | 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. |
| 10 | Open Source Project / ABY Framework [29] | The 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. |
| 11 | Initiative / Confidential Computing Initiative [30] | The Confidential Computing Consortium is an industry collaboration that aims to promote the adoption of confidential computing technologies, including secure multi-party computation. The 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. |
| 12 | Initiative / ZKProof Standards [31] | Several initiatives and organizations, such as the Zero-Knowledge (ZK) 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 trustful data sharing and exchange interoperability. |
| 13 | Authority /ENISA [32] | ENISA (European Union Agency for Cybersecurity) 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. |
| 14 | EU Project / pdp4e [33] | PDP4E's focus is on spreading the creation of products, systems and services that better protect the privacy and personal data of EU citizens. To achieve this, PDP4E focuses 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. |
| 15 | InterPlanetary File System[34] | IPFS is a distributed file storage protocol designed to create a peer-to-peer network for storing and sharing data. Unlike traditional centralized systems, IPFS employs content-addressed storage, which ensures that files are retrieved based on their cryptographic hash rather than a specific location. This enables secure, resilient, and decentralized data sharing across the network.<br>For PISTIS, IPFS could offer a foundational infrastructure for secure and decentralized data storage and sharing, aligning with its goals of trusted and federated data ecosystems. We IPFS, PISTIS could enhance the scalability and robustness of its data trading and management functionalities. |

## 7.2 DATA TRADING

### 7.2.1 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 22: Initiatives related to distributed ledgers**

| # | INITIATIVE/ PROJECT NAME | SHORT DESCRIPTION AND RELEVANCE FOR PISTIS |
|---|---|---|
| 1 | Hyperledger BESU[35] | Hyperledger Besu is an enterprise-grade, open-source Ethereum client designed for both public and private blockchain networks. It supports a variety of consensus mechanisms, including Proof of Stake (PoS), Proof of Authority (PoA), and IBFT, making it a flexible solution for diverse blockchain deployments. Hyperledger Besu provides robust features like permissioning, privacy groups, and advanced monitoring and metrics, ensuring secure and scalable blockchain operations.<br><br>Besu's compatibility with the Ethereum Virtual Machine (EVM) enables the deployment and execution of smart contracts, achieving interoperability with the broader Ethereum ecosystem. Its advanced support for private transactions and privacy management can serve as an important framework |

[34] https://ipfs.tech/

| | | |
|---|---|---|
| | | for PISTIS, enabling secure data sharing and decentralized identity verification. |
| 2 | OpenDSU and Sovrin [36] | OpenDSU is a decentralized data sharing framework that emphasizes privacy, security, and interoperability. It provides tools for managing and sharing data assets in a distributed environment without requiring a central authority. By offering features such as decentralized identity (DID) integration, encrypted data sharing, and secure communication channels, OpenDSU ensures that sensitive information can be stored and transmitted in a tamper-proof and privacy-preserving manner.<br><br>Sovrin is a decentralized identity network built specifically for self-sovereign identity (SSI). It leverages blockchain technology to enable the secure storage and sharing of verifiable credentials, empowering individuals to have control over their digital identities. Sovrin's trust framework and governance model provide a robust infrastructure for establishing and verifying identities in a decentralized manner. With its support for standards such as DIDs and verifiable credentials, Sovrin ensures interoperability across various ecosystems. Its focus on privacy, security, and user control makes it a strong candidate for providing the identity verification and credential management capabilities needed for PISTIS.<br><br>While OpenDSU and Sovrin each offer unique strengths regarding the decentralized data sharing and self-sovereign identity, it is not yet decided which technology will be adopted for PISTIS. The motivation behind the integration of one of these technologies in PISTIS ecosystem (the technology that will be chosen is still under discussion), is its capability to link a user's identity key with the data encrypted payload in a privacy preserving manner. This essentially allows the secure execution of complex data trading agreements without the need to update user's identities keys (based on the use of re-encryption schemes). |
| 3 | Filecoin37 | Filecoin is a decentralized peer-to-peer storage network that combines the content-addressing capabilities of IPFS with blockchain-powered storage insurances. It implments an incentive layer for reliable and cost-effective file storage, ensuring data persistence through economic incentives and cryptographic proofs. Filecoin adds an incentive mechanism on top of IPFS that promotes reliable storage and consistent data access. Storage deals are recorded on-chain, and providers must submit cryptographic proofs to verify data integrity over time. This system enables Filecoin to integrate seamlessly with smart contract platforms like Ethereum, Polygon, and Solana, offering cost-effective off-chain storage with verifiable content identifiers (CIDs) that are recorded on the Filecoin blockchain. Filecoin supports a wide range of use cases, from Web3 applications such as NFT storage.<br><br>Both Filecoin and the PISTIS project aim to enhance data sharing and storage through decentralized, secure, and incentivized frameworks. Both projects are utilizing technologies like distribution ledger technologies (DLTs) and data non-fungible tokes (NFTs), smart contracts to build trust among stakeholders. Both initiatives aim to create decentralized, secure, and |

[37] https://filecoin.io/

| | | |
|---|---|---|
| | | economically viable ecosystems for data storage and sharing, leveraging technologies to foster trust and efficiency in data management. |
| 4 | Datum.org[38] | Datum is a blockchain-based data marketplace that enables data sovereignty, ensures privacy and security. It leverages distributed ledger technology for transparent and immutable data transactions, decentralized storage systems for secure and distributed data hosting. Smart contracts automate data-sharing agreements and payments using the DAT token, which drives the platform's tokenized economy. Decentralized applications (DApps) provide interfaces for managing data and permissions, while identity and access management systems ensure only authorized parties can access data.<br><br>Datum and PISTIS both aim to bring innovation in data management and monetization through secure and transparent systems. They provide a marketplace for personal data ownership, decentralized storage, and monetization using smart contracts and tokens. Both utilize distributed ledger technologies (DLTs) to ensure trust, immutability, and decentralized control over data exchanges. Additionally, both projects integrate technologies to enhance privacy, security, and efficiency in data sharing, promoting trust and transparency in the evolving data economy. |
| 5 | Corda[39] | Corda is a blockchain based platform designed to manage transactions with privacy, scalability, and legal enforceability. It operates in a permissioned environment, using a unique consensus model where only relevant parties validate transactions, ensuring that data and transactions are visible only to involved parties. Key technologies include a distributed ledger for recording immutable transactions, smart contracts for automating agreements, and cryptographic techniques to secure data.<br><br>Corda and PISTIS both focus on secure, private, and efficient data management for organizations, leveraging distributed ledger technology and NFTs. They support interoperability, allowing integration with existing systems, and use smart contracts to automate agreements and facilitate seamless data transactions. |

### 7.2.2 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. In the current version of the deliverable (D1.3), the focus is on a subgroup of the previously selected initiatives and projects presented in D1.2 (M9), that has been found to be most strongly related with PISTIS as the project's requirements have matured since M9 of the project.

[38] https://datum.org/

[39] https://corda.net/

**Table 23: Initiatives related to data monetization**

| # | INITIATIVE/ PROJECT NAME | SHORT DESCRIPTION AND RELEVANCE FOR PISTIS |
|---|---|---|
| 1 | 5GMETA [40] | The EU-funded 5GMETA project (H2020 ID: 957360) developed an open platform to leverage car-captured data to stimulate and facilitate innovative products and services. It empowered 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 created 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.<br>PISTIS has taken into consideration the car data monetization services of 5GMETA for the design and development of its Market Exchange Platform. |
| 2 | CitizenMe [41] [42] | CitizenMe is a 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. PISTIS is leveraging the ethical dimension of this marketplace on its developments. |
| 3 | Datacoup [43] [44] | Datacoup provides blockchain-enabled apps and tools for monetization of personal data.<br>It is now part of another company, called ODE. PISTIS is monitoring this monetization application to get any ideas on the relevant domain for potential development. |
| 4 | Datum [45] [46] [47] | "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."<br>Datum is relevant to PISTIS because it developed a data market (personal data only!) that functions over the blockchain. PISTIS is monitoring this monetization application to get any ideas on the relevant domain for potential development. |

| | | |
|---|---|---|
| 5 | DATAMITE [48] | 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.<br>PISTIS has established a synergy with DATAMITE towards exchanging ideas in the data trading and monetization services. |
| 6 | Eurecat Data Valuation Platform [49] | 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 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). PISTIS leverages the capabilities of this platform for its own data valuation developments. |
| 7 | FAME [50] | 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.<br>PISTIS has established a synergy with FAME towards exchanging ideas in the data trading and monetization services. |
| 8 | Oracle Enterprise Data Quality [51] | "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." PISTIS is leveraging the capabilities of this platform on the above services towards delivering corresponding innovative solutions in its delivered artefacts. |
| 9 | Permission.io [52] | 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. PISTIS leverages the capabilities of this cryptocurrency payments for its own monetization services. |

| | | |
|---|---|---|
| 10 | SAP Master Data Quality Management [53] | The Master Data Quality Management suite is part of the SAP Master Data Governance on SAP S/4HANA.<br>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. PISTIS leverages the capabilities of this suite for its own data management developments. |
| 11 | Snowflake Market-place [54] | A solution for discovery, evaluation and purchase of data, data services and applications.<br>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.<br>An embedded demo video nicely explains the capabilities of the plat-form. PISTIS is using the features of this product as examples for the de-velopment of the purchasing mechanisms in its own artefacts. |
| 12 | Talend [55] [56] | Talend offers solutions for Data Integration (transformation, mapping, en-richment 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.<br>The company seems to be interested in the topic of data monetization and their tools appear to be valuable components in such a process. PISTIS leverages these capabilities for its data management and monetization purposes. |
| 13 | Unidata [57] | Unidata is a multifunctional platform for building corporate data manage-ment systems, providing centralized data collection (inventory and ac-counting resources), standardization of information (normalization and enrichment), accounting current and historical information (control of record versions, periods data relevance), data quality and statistics. PISTIS applies these concepts in its own data quality management develop-ments. |
| 14 | UPCAST [58] | 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.<br>PISTIS has established a synergy with UPCAST towards exchanging ideas in the data trading and monetization services. |

### 7.2.3 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

includes 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 24: Initiatives related to data sources and services**

| # | INITIATIVE/ PROJECT NAME | SHORT DESCRIPTION AND RELEVANCE FOR PISTIS |
|---|---|---|
| 1 | BDVA[59] | The Big Data Value Association (BDVA) is an industry-driven organisation with a mission to develop an innovation ecosystem that enables the data-driven digital transformation of the economy and society in Europe.<br>Several members of the PISTIS consortium are members within the BDVA, and PISTIS is integrated into the BDVA ecosystem. The project aligns its outcomes with the BDVA Data Spaces Blueprint Architecture, contributes to analytics reports, and actively supports the activities of the BDVA Task Forces on "Data Spaces" and "Standards." Additionally, PISTIS maintains active participation in BDVA events. BDVA serves as a vital platform for communication and collaboration, fostering connections with other data spaces projects and initiatives. |
| 2 | Catena-X / Tractus-X [60] [61] | 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.<br>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 [62] | The official portal for European data contains open data from the European public sector. PISTIS will use some of the data to implement the planned demonstrators' scenarios and for evaluation purposes. |
| 4 | DSBA [63] | 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.<br>PISTIS may benefit from DSBA for the design and development of its Data Management Platform services. |

| | | |
|---|---|---|
| 5 | DATAMITE [64] | 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. The project aims to complete by the end of 2025. DATAMITE delivers a modular, open-source and multi-domain framework to enable data monetising, interoperability, trading, and exchange for European companies and institutions. The project passed its half-way review point in October 2024, where a working prototype framework was successfully demonstrated. The second half of the project focuses on completion of a fully working framework, as well as delivery of training materials for companies to improve their data literacy and culture. The project will also test framework deployments for intra-company sharing; for sharing of energy data in data spaces; and for sharing of data in other specific environments such as AI On Demand and edWIN.<br>The project will provide contributions relevant for PISTIS:<br>• Technological stack to exchange quality data in data spaces<br>• Integration of several IDSA (or alternative) tools, including connectors in the Data Sharing Module<br>• Easy-to-use interoperability and sovereignty tools for data exchange<br>DATAMITE outcomes are relevant for many PISTIS tasks. |
| 6 | DataBri-X [65] | Data Process & Technological Bricks for expanding digital value creation in European Data Spaces (DataBri-X) is a Horizon 2023 project started in October 2022, aiming to complete by Q4 2025. Its ambition is to enable data sharing through cross-border, cross-sectoral data spaces and enable platforms to process proprietary, personal, and open public data with a toolbox of robust and scalable bricks ("Bri-X") that assist with ensuring data and metadata interoperability, usability, discoverability, quality, and integrity. The project aims to provide tools to support governance of the data lifecycle in compliance with FAIR (Findable, Accessible, Interoperable, and Reusable) principles.<br><br>The foundation is a standards-based toolbox of technologies focused on the creation and maintenance of common ontologies, vocabularies and data models, as well as automated authoring, co-creation, curation, annotation and labelling of data. DataBri-X is deeply integrated with the International Data Spaces (IDS) framework, and the IDS Connector and Information Model standards are used extensively. Metadata for each use case is stored in a knowledge graph (DataBri-X KG). On top of this DataBri-X is implementing a data governance process based on the JenPlane project[66].<br><br>The project will address issues, such as data ownership, data provenance veracity and verification, decentralised data sharing, confidentiality, and digital rights management. |

[66] https://github.com/javadch/jenplane

| | | |
|---|---|---|
| | | There are three core DataBri-X use cases being explored in 2024: an energy industry data space; a telecoms market data space; and a legal industry data space.<br><br>All these topics are highly relevant for PISTIS. Consequently, a scientific and technical exchange between DataBri-X and PISTIS could be mutually beneficial for both projects. |
| 7 | DataHub by LinkedIn [67] | 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.<br>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 [68] | The Dat Ecosystem is a global community of many projects, most of which are self-funded. All are built on top of Hypercore protocol, a secure transport protocol that makes it easy to build fast and scalable peer-to-peer applications. Project teams 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.<br>Alongside the foundational Hypercore project, key projects include:<br>• Agregore - a browser for the distributed web that facilitates peer-to-peer data sharing without central servers, supporting protocols like BitTorrent and IPFS for direct loading and sharing of content.<br>• Ara - a content protocol with decentralized rewards and distribution. Ara can be used to package up NFTs and blockchain based content for any application that requires the sale of media. Some of the use cases of Ara are direct distribution of content for games studios, film or video, music, and individual creators.<br>• Cabal - an experimental P2P community chat platform where servers are unnecessary, everything runs locally, and each community is identified by a secret key, offering both internet and local network connectivity.<br>• Earthstar - a small and resilient distributed storage protocol designed with a strong focus on simplicity. Earthstar stores all data directly on devices, so that reading and writing works without an internet connection. Data is versioned, |

| | | |
|---|---|---|
| | | mutable, and can sync across devices, but databases are private by default.<br>This community is relevant to PISTIS since it promotes secure peer-to-peer applications and data sharing. |
| 9 | DIF [69] | DIF is an alliance of organizations that 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.<br>DIF's work on harmonized protocols, digital identifiers, and secure wallet architecture offers foundational tools that PISTIS can integrate to ensure seamless and secure identity management. By considering DIF's outputs, PISTIS can enhance its compliance with decentralized identity frameworks, improve interoperability across systems, and ensure secure and user-centric identity solutions within its platform. |
| 10 | DWeb [70] | DWeb is an umbrella organization that connects people, projects, and organizations that are active in the domain of the decentralized Web. DWeb organizes regular events and meetups to exchange ideas and advocate for decentralised web concepts, initiatives, and implementations.<br>DWeb network has Nodes, or DWeb Meetup groups, based in cities around the world. These Nodes organize local events for people to meet others, exchange knowledge, and deepen connections across the Dweb community. DWeb is a project of the nonprofit Gatherings for Good, and is sponsored by organisations including the Ethereum Foundation, Protocol Labs, Hyperledger, the Web3 Foundation and the Internet Archive.<br>For PISTIS the initiative may be relevant as a channel for dissemination and gathering feedback. |
| 11 | EDC [71] | 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.<br>Similar with the works by IDSA, EDC is also highly relevant for PISTIS, since it provides ideas on the technical components that could be used for data transfer to develop the PISTIS data ecosystem. |
| 12 | EBSI [72] | 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.<br>EBSI is a relevant blueprint, how to govern and deploy a cross-border Blockchain network. PISTIS can adopt relevant processes and artefacts. |
| 13 | FAME [73] | 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.<br>PISTIS has established a synergy with FAME towards exchanging ideas in the data trading and monetization services. |
| 14 | FIWARE [74] | 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.<br>The semantic data models provided by FIWARE represent one of the most valuable outcomes of this initiative. PISTIS could leverage these models in defining its domain-specific data models, ensuring interoperability with the FIWARE ecosystem. |
| 15 | Gaia-X [75] | 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.<br>Gaia-X is highly relevant for PISTIS, as its focus on secure data sharing and trust aligns closely with PISTIS's core values. PISTIS plans to leverage the Gaia-X Trust Framework for the design and governance of its Data Management Platform services. Specifically, PISTIS plans to implement a GAIA-X conformed Data Catalogue and align dataset descriptions with Gaia-X specifications. This alignment will enhance interoperability with other projects adhering to Gaia-X standards. |
| 16 | GXFS [76] | 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. |

| | | |
|---|---|---|
| | | PISTIS, with its native RDF-based data catalogue, aims to offer an advanced alternative to the Gaia-X catalogue developed within GXFS. Collaboration and knowledge exchange with GXFS will enhance the visibility of PISTIS and increase opportunities for its adoption and utilization. |
| 17 | Hyperledger [77] | 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.<br>The Hyperledger project is highly relevant for PISTIS, since it can be a critical 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.<br>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.<br>The works that has been done by this association are highly relevant to PISTIS due to their vision. While PISTIS does not utilize technologies developed by IDSA, it can collaborate with the organization on standardization and interoperability efforts. |
| 19 | LinkedDataHub [78] | 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 open data or knowledge graphs to set up their own open data environment. By default, the LinkedDataHub platform includes a SPARQL endpoint and a Linked Data API which do not require yet another development efforts establish SPARQL Endpoint to query knowledge graphs. 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 processes. 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 [79] | 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.<br>Discoverability and Likability of data entities, are two key features of this project that PISTIS may benefit from for the design and development of its data matchmaking services. |
| 21 | Ocean Protocol [80] | 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.<br>The Ocean Protocol is highly relevant for PISTIS, since it employs a Blockchain with Ethereum as backend. Methodologies and standards can be transferred to PISTIS. |
| 22 | piveau [81] | 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.<br>Piveau serves as a valuable foundation for building metadata and schema repositories within PISTIS. Its native Linked Data support enables seamless integration with specifications from the broader Semantic Web ecosystem. PISTIS is enhancing the Piveau Catalogue to meet identified requirements and evolve it into a marketplace. The outcomes of PISTIS will, in turn, contribute back to the Piveau open-source project, enriching its capabilities. |
| 23 | Solid [82] | 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 which are accessible through the Web. It also supports storage of Linked data. 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. Several reference implementations are available as open-source software. Solid can be used in the context of 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 to store industrial data.<br>So far in PISTIS, each factory hosts a storage in its own premises to support decentralization of the data and the metadata of this data is handled as linked data and is made available in a centralized manner. To expand PISTIS, Solid pods can replace the data storage in factories with minimal on-premises storage capacities |
| 24 | TANGO [83] | 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 develops technologies for industrial data sharing:<br>• ensuring privacy, security & trustworthiness<br>• data & AI toolkit<br>• legal and ethical compliance.<br>To enable data sharing TANGO is following the IDS approach following the recommendations of the DSSC and DSBA. In particular, TANGO technologies related to privacy, security and trust are being packaged extending one of the existing IDS Connectors (the FIWARE Data Space Connector), enabling data exchange.<br>TANGO outcomes are not directly related to the PISTIS technical components, except for the Analytics engine, which is related to the MLOps solution in TANGO. PISTIS may benefit of the mutual relation with the DSSC. |
| 25 | Mobility Data Spaces[84] | Mobility data spaces represent interconnected ecosystems where mobility-related data—such as vehicle telemetry, traffic patterns, public transportation schedules, and environmental factors—are securely shared among diverse stakeholders. These data spaces leverage federated architectures and interoperability frameworks to enable seamless data exchange while maintaining data sovereignty. They aim to address challenges like fragmented data silos, inconsistent data standards, and trust issues between stakeholders.<br>PISTIS could benefit from Mobility Data Spaces by leveraging their standards and frameworks for interoperability and secure data exchange within the mobility sector. These data spaces provide validated use cases and practical methodologies that PISTIS could align with, ensuring its solutions are grounded in proven approaches. |
| 26 | deployEMDS[85] | The deployEMDS project, co-funded under the EU Digital Europe Programme, aims to establish a common European Mobility Data Space (EMDS) by fostering a broad ecosystem of data providers and users. It focuses on facilitating standardized data sharing, ensuring data sovereignty and trust, and enhancing the discoverability and |

[84] https://mobility-dataspace.eu/

[85] https://deployemds.eu/

| | | |
|---|---|---|
| | | availability of mobility data. The project encompasses 16 use cases across nine EU countries, contributing to the development of innovative services and applications in the mobility sector.<br>The objectives of deployEMDS align closely with those of the PISTIS project, which aims to develop a reference platform for the secure, trusted, and controlled exchange and monetization of proprietary data. PISTIS could gain advantages by aligning with its work on standardization and harmonization of data sharing protocols across the European mobility landscape. |

## 7.3 TECHNOLOGY RADAR SUMMARY

In preparing for and delivering the PISTIS MVP, PISTIS project partners have worked extensively to align PISTIS requirements, architecture, design and implementation with existing internationally-recognized initiatives and standards in the fields of data extraction, transformation and loading (ETL); privacy, security and trust; distributed ledgers; data monetization; and data discovery, connectivity and integration. A summary of alignment work is as follows:

- In relation to data ETL, 13 initiatives and standards have been explored, and PISTIS is aligning to 4 of these: with PISTIS partners continuing to evaluate the benefits of alignment to a further 9.
- In relation to data privacy, security and trust, 14 initiatives and standards have been explored, and PISTIS is aligning to 4 of these: with PISTIS partners continuing to evaluate the benefits of alignment to a further 10.
- In relation to distributed ledgers, 5 initiatives and standards have been explored, and PISTIS is aligning to both.
- In relation to data monetization, 14 initiatives and standards have been explored, and PISTIS is aligning to 12 of these: with PISTIS partners continuing to evaluate the benefits of alignment to a further 2.
- In relation to data discovery, connectivity and integration, 24 initiatives and standards have been explored, and PISTIS is aligning to 8 of these: with PISTIS partners continuing to evaluate the benefits of alignment to the others.

While the Technology Radar section highlights promising initiatives aligned with the goals of the PISTIS platform, several considerations must be addressed to ensure successful implementation. Interoperability challenges may arise, as many technologies, such as distributed ledgers and privacy-preserving computation frameworks, operate on distinct protocols and standards. Integrating these into a unified platform requires efforts to bridge compatibility gaps while maintaining performance and reliability. Scalability is another concern, particularly for decentralized storage systems and blockchain-based platforms, which may face limitations when handling large-scale data or high transaction volumes. These issues necessitate the adoption of robust scalability solutions to ensure their practicality in real-world scenarios.

Regulatory and compliance risks also pose significant challenges, especially as data sharing and monetization activities increasingly span multiple jurisdictions. Technologies must align with complex regulatory frameworks, such as GDPR and data sovereignty laws, which may introduce delays and complexity in implementation. The technical maturity of some identified solutions, particularly those involving advanced cryptographic techniques or AI-driven data valuation, is another factor. These technologies are often in their early stages of adoption, leading to potential concerns regarding their stability, reliability, and long-term support.

Additionally, the cost of implementation is a critical factor. Advanced technologies often require substantial initial investments in licensing, integration, and training, which could hinder adoption, especially for smaller organizations or pilot projects. Security vulnerabilities must also be addressed, as even technologies designed to enhance security, such as encryption and privacy-preserving computation, may introduce new attack vectors if not implemented correctly. Ensuring robust security practices throughout deployment and operation is vital.

Finally, user adoption and usability remain essential considerations. Technologies with complex or non-intuitive interfaces may deter users, necessitating additional resources for training and user experience improvements

# 8 CONCLUSION

This deliverable is based on D1.2 and presents the methodology and all the activities undertaken to define the PISTIS MVP as this evolved during the alpha release of the project. Building on business requirements, technological advancements, and the initial concept of PISTIS as outlined in D1.1, the primary objective was to update the primarily defined in D1.2 technical requirements that are essential to support the PISTIS functional framework.

For the definition of the user stories, a structured methodology was employed that encapsulate the needs and benefits for PISTIS platform end-users. The approach combined traditional narrative-based user story definitions with formal business process representations, fostering a shared understanding between end-users and developers. Following an examination of the business requirements and data lifecycle of PISTIS, fourteen (14) user stories were identified and categorized into two groups: business operations and system support operations.

This analysis resulted in ninety (88) technical requirements aligned with the functional needs of end-users and emphasizing the unique value provided by PISTIS. These requirements excluded standard information systems technical aspects, focusing instead on features specific to PISTIS. The PISTIS MVP was subsequently defined through a prioritization process using the MoSCoW method, resulting in seventy-one (71) technical requirements categorized into must-have and should-have priorities. These requirements guide the design and implementation of the PISTIS framework and its components across work packages WP1, WP2, and WP3.

In D1.3, additional sections have been introduced to address the Technology Radar and Generic Terms of PISTIS Data Sharing Contracts. The Technology Radar highlights emerging technological advancements critical to enhancing PISTIS's capabilities in areas such as privacy, security, and distributed ledger systems, ensuring the platform remains aligned with cutting-edge developments. The Generic Terms of PISTIS Data Sharing Contracts provide a standardized framework for data-sharing activities, offering robust legal and operational clarity to stakeholders. These additions further reinforce the platform's adaptability and trustworthiness, laying the groundwork for a secure and collaborative data-sharing ecosystem.

The work presented in this deliverable establishes a comprehensive roadmap for the continued development and refinement of the PISTIS platform, ensuring alignment with stakeholders' expectations, technological trends, and regulatory requirements.

# ANNEX 1 – PISTIS GENERIC TERMS OF DATA SHARING

## PISTIS GENERIC TERMS OF DATA SHARING

**Subject Matter.** The subject matter of this data sharing between the Data Provider and the Data Recipient concerns a particular dataset and the data that it contains, being [**Describe*] (hereinafter collectively: '*Data Set*').

**Scope and Purpose(s) of Sharing of the Data Set.** Subject to the data-sharing arrangement formed on the PISTIS Data Marketplace between the Data Provider and the Data Recipient, the Data Provider hereby grants the Data Recipient an [**exclusive/non-exclusive*], [**worldwide/Region/Country*], [**transferable/non-transferable/sub-licensable*], [**perpetual/ for the duration of (term)*] license to access, copy and process the Data Set for the following purpose(s) (hereinafter '*Permitted Purposes*'):

- internally, including but not limited to, for conducting demos, research and development purposes;
- to develop software, data analytics or other forms and applications of automated processes, or machine learning or artificial intelligence;
- to generate derived works, recommendations and analyses, including by using models and algorithms, the results;
- to incorporate insubstantial portions, extracts, abstracts or summaries of the Data Set into analyses, presentations or tools used for commercial purposes;
- to store the Data Set in databases hosted internally or on third party hosted platforms;
- [**Describe other purpose(s) of the use of the Data Set*]

The Data Set shall not be processed for purposes that have not explicitly specified as Permitted Purposes. Any other rights of the Data Provider in or to the Data Set not granted to the Data Recipient are expressly reserved by the Data Provider.

**Derivative Works**. The Data Provider acknowledges and agrees that the Data Recipient shall retain ownership of all intellectual property right in the derivative works of the Data Set. For the avoidance of doubt, in case the Data Set is modified only in minor ways and used for substituting the original Data Set, it shall not be regarded as derived material or work, and remains under the restrictions set out for the Data Set.

**Representations and Warranties.** The Data Set is provided as checked-in by the Data Provider on the PISTIS Data Marketplace, without any warranty of any type, either express or implied, including without limitation, any warranty of merchantability or fitness for a particular purpose or use, title or otherwise, and without any warranty regarding the suitability of the Data Set, whether it operates uninterrupted or error-free, or whether errors or other defects, if any, shall be corrected.

**Effective Date**. The terms, including the license and right to use the Data Set granted under these terms, shall become effective and applicable to the data-sharing arrangement between the Data Provider and the Data Recipient upon the Data Recipient's acceptance of these terms and conditions. The Data Recipient's acceptance of the terms shall be digitally exercised and recorded on the PISTIS Data Marketplace, and the date of the acceptance shall be deemed as the effective date of these terms and conditions (hereinafter 'Effective Date').

**Delivery**. The access to the Data Set will be granted to the Data Recipient upon the receipt of the payment of the Fee as agreed herein.

**Term**. These terms and the license granted herein, shall be valid and [*applicable for a term of *Specify the term / *perpetual*] and shall be automatically renewed for additional term of [**Specify the term*] unless either Party provides the other with written notice not to renew at least [**Specify the term*] days prior to the expiration date of the current term.

**Termination.** Either Party may terminate their data-sharing arrangement immediate upon written notice if other Party is in material breach of these terms and if such breach is that is not cured within [**Specify the term*] days after being notified of the breach. Furthermore, the Data Provider may terminate the data-sharing arrangement or provision of the Data Set on the PISTIS Data Marketplace upon reasonable prior written notice to the Data Recipient if the Data Provider's rights to material portions of the Data Set or data involved in the Data Set becomes unavailable provided that the Data Provider shall provide a pro-rated refund to Data Recipient for the remainder of the then-current term.

**Consequences of Termination**. In case of termination, the access to, provision and availability of the Data Set on the PISTIS Data Marketplace will automatically terminate without further action by either party. Upon termination, the Data Recipient shall destroy all copies of the Data Set within the Data Recipient's possession or control and all of Data Recipient's rights in and the Data Set. Termination shall not affect Data Recipient's obligation to pay all fees due prior to termination, and termination shall not relieve Data Recipient of any liability for breach of these terms.

**Fees and Payment Terms.** OPTION 1: The Parties agree on that the Data Recipient shall pay the Data Provider [**X amount PISTIS Cryptocurrency*] ('Fee') for the license to access, copy and process the Data Set as defined herein on the Effective date. The Fee is exclusive of VAT or local sales tax or any other applicable taxes. OPTION 2: The Parties agree on that Data Recipient shall pay Data Provider [**X amount PISTIS Cryptocurrency*] ('Subscription Fee') on monthly basis, in advance, starting from the Effective Date and recurring on the same date of every following month, for the license to access, copy and process the Data Set as defined herein during the term of the data-sharing agreement. In case when the Effective Date corresponds 31$^{st}$ or 29$^{th}$ day of a month, the Subscription Fee becomes due and payable on the first day of every following month. The access to the Data Set will be granted to Data Recipient upon the receipt of the first payment on the PISTIS Data Marketplace. The Subscription Fee is exclusive of VAT or local sales tax or any other applicable taxes.

**Protection of Personal Data.** OPTION 1: The Data Set does not include personal data. OPTION 2: The Data Set includes personal data, and the related data processing is subject to the following: [**In the event that the dataset contains personal data including pseudonymised personal data, the Data Provider is advised to consider defining the terms and conditions for the transfer and processing of personal data*]. OPTION 3: [*].

**Miscellaneous.** For the avoidance of doubt and given that this data sharing transaction takes place between PISTIS project partners as part of the PISTIS project, each Party acknowledges that the PISTIS Grant Agreement ('GA') and the PISTIS Consortium Agreement ('CA') is applicable to this data-sharing arrangement in general, and the following rights and obligations therein in particular, without limitation:

- Proper implementation of the Action;
- Use restrictions, confidentiality, non-disclosure, data protection and security;
- Ethics and values;
- Results, (joint) ownership and dissemination;
- Information obligations, and;
- Applicable law and forum,

related to which Parties herewith reconfirm to each having put and for the term of this data-sharing arrangement, the CA and GA remaining in place proportionate and otherwise appropriate organisational, operational and technical measures. In case of any inconsistency or conflict between these terms and the GA or CA, the GA and CA shall prevail.

# ANNEX 2 – PISTIS GENERIC TERMS OF DATA SHARING AND JOINT DATASET CREATION AGREEMENT

## PISTIS GENERIC TERMS OF DATA SHARING AND JOINT DATASET CREATION AGREEMENT

**Parties and Purpose.** Partner A, and Partner B, agree to collaborate on the creation of a common dataset (hereinafter collectively: '*Data Set*'), by contributing their respective data and resources in accordance with the terms and conditions outlined in this Agreement. Hereinafter the Partner A and the Partner B collectively referred to as the "Parties."

**Scope of Contributions.** The Partner A agrees to contribute the following data: [**specify type of data, format, upload frequency etc.*], as detailed in [*Exhibit B*]. The Partner B agrees to contribute the following data: [**specify type of data, format, upload frequency etc.*], as detailed in [*Exhibit B*]. The Parties agree to collaborate in refining, cleaning, or expanding the dataset as needed to ensure the Dataset meets the specifications described in [*Exhibit A*]. Each Party is responsible for the quality, accuracy, and legality of the data it contributes.

**Joint Ownership of Rights.** The Data Set is a collective work created by both Parties and thus the Parties will jointly own and claim any copyrights and sui generis rights over the Data Set. Each Party will hold [*an equal, undivided*] interest in the Data Set.

**Permitted Use.** The Parties agree to use the Data Set for the following purpose(s):

- internally, including but not limited to, for conducting demos, research and development purposes;
- to develop software, data analytics or other forms and applications of automated processes, or machine learning or artificial intelligence;
- to generate derived works, recommendations and analyses, including by using models and algorithms, the results;
- to incorporate insubstantial portions, extracts, abstracts or summaries of the Data Set into analyses, presentations or tools used for commercial purposes;
- to store the Data Set in databases hosted internally or on third party hosted platforms;
- [**Describe other purpose(s) of the use of the Data Set*]

Neither Party may license, sell, or otherwise transfer the Data Set to a third party without the prior written consent of the other Party, except where such use is explicitly permitted under this section.

**Derivative Works**. Any derivative datasets or analyses created using the common Data Set will be owned by the Party creating such derivative works, unless otherwise agreed in writing.

**Representations and Warranties.** Each Party represents and warrants that the data it contributes to the Data Set, (hereinafter "Contributed Data"), is accurate, lawful, and free of third-party claims, and that it has the right to contribute such data. Each Party warrants that its contributed data does not infringe any third-party intellectual property rights or violate any laws or regulations.

**Effective Date**. The terms, including the license and right to use the Data Set granted under these terms, shall become effective and applicable to the data sharing and joint dataset creation arrangement between the parties upon the mutual acceptance of either party of these terms and conditions. The Parties' acceptance of the terms shall be digitally exercised and recorded on the PISTIS Data Marketplace, and the date of the acceptance shall be deemed as the effective date of these terms and conditions (hereinafter 'Effective Date').

**Costs and Expenses.** Each Party shall bear its own costs and expenses associated with the collection, preparation, and contribution of its data to the Data Set unless otherwise agreed upon.

**Revenue Sharing.** If the Data Set generates revenue through [*monetization/commercial use*], the Parties shall share the proceeds as follows:

- [**Specify percentage*] to Partner A
- [**Specify percentage*] to Partner B

Optional Clause: Details on revenue sharing will be further outlined in Exhibit C.

**Term**. These terms and the license granted herein, shall be valid and [*applicable for a term of *Specify the term / *perpetual*] and shall be automatically renewed for additional term of [**Specify the term*] unless either Party provides the other with written notice not to renew at least [**Specify the term*] days prior to the expiration date of the current term.

**Termination.** Either Party may terminate this arrangement immediate upon written notice if other Party is in material breach of these terms and if such breach is that is not cured within [**Specify the term*] days after being notified of the breach.

**Consequences of Termination**. In case of termination, each Party shall retain rights to any derivative works or analyses it has created using the Data Set up to the date of termination. The jointly owned rights to the Data Set will remain unless otherwise mutually agreed upon.

**Indemnification.** The Parties agree to indemnify and hold harmless each other from any claims or liabilities arising from the use of their respective contributed data, including any breach of its representations or warranties.

**Protection of Personal Data.** Each Party agrees to comply with all applicable data protection and privacy laws when providing its contributed data.

**Miscellaneous.** For the avoidance of doubt and given that this data sharing transaction takes place between PISTIS project partners as part of the PISTIS project, each Party acknowledges that the PISTIS Grant Agreement ('GA') and the PISTIS Consortium Agreement ('CA') is applicable to this data sharing and joint dataset creation arrangement in general, and the following rights and obligations therein in particular, without limitation:

- Proper implementation of the Action;
- Use restrictions, confidentiality, non-disclosure, data protection and security;
- Ethics and values;
- Results, (joint)ownership and dissemination;
- Information obligations, and;
- Applicable law and forum,

related to which Parties herewith reconfirm to each having put and for the term of this data-sharing arrangement, the CA and GA remaining in place proportionate and otherwise appropriate organisational, operational and technical measures. In case of any inconsistency or conflict between these terms and the GA or CA, the GA and CA shall prevail.