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

D3.1
Data Valuation, Sharing and Trading Framework

Editor(s)	Konstantinos Latanis, Sotiris Koussouris
Lead Beneficiary	SUITE5
Status	Final
Version	1.00
Due Date	30/09/2023
Delivery Date	21/10/2023
Dissemination Level	PU

Funded by the European Union under Grant Agreement 101093016. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Project	PISTIS – 101093016
Work Package	WP3 - PISTIS FAIR Data Trading and Value Exchange/Monetisation Platform Services
Deliverable	D3.1 - Data Valuation, Sharing and Trading Framework
Contributor(s)	SUITE5, ICCS, UBITECH, IMPERIAL, UMALTA, CUERVA, ISI/ATHENA
Reviewer(s)	ATOS, BAMBOO
Abstract	The deliverable D3.1 "Data Valuation, Sharing and Trading Framework" comprises a direct outcome of the initial phase of Tasks of WP3, documenting the candidate methodologies and technologies that are considered for the design and implementation of the different services of the PISTIS FAIR Data Trading and Value Exchange/Monetisation Platform.

Executive Summary

The deliverable D3.1 "Data Valuation, Sharing and Trading Framework" comprises a direct outcome of Task 3.1 "Multidimensional Data Valuation Assessment and Dynamic FAIR Value Pricing Methodology ", Task 3.2 "Data Financial Market Core features, Data Investments, NFTs and StableCoin Design ", Task 3.3 "Data Valuation Assessment and Dynamic Pricing Services ", Task 3.4 "XAI driven Monetisation and Matchmaking Services", Task 3.5 "PISTIS Data Exchange Market, Crypto Operations and Wallets" and Task 3.6 "Immutable Blockchain Network for Data Trading Management and Operations Auditing", documenting the candidate methodologies and technologies that are considered for the design and implementation of the different services of the PISTIS FAIR Data Trading and Value Exchange/Monetisation Platform (hereafter called "PISTIS Market Exchange Platform" for brevity reasons).

The PISTIS Market Exchange Platform will enable the valuation of datasets and provide meaningful insights to data providers towards sharing their datasets through the creation of blockchain-enabled contracts to interested stakeholders by exploiting NFTs while maintaining the privacy and security of data and transactions. Eventually, it will constitute an exchange market for datasets that will bring together different organisations without the necessity for storing their data outside their premises.

The PISTIS Market Exchange Platform consists of 5 different core bundles of services:

a) The Data Value Contract Composer: Facilitates the description of datasets, generates corresponding NFTs and designs the data subscription processes.
b) The Monetisation XAI Engine: Provides insights for the relevant data markets and usage analytics for the FAIR valuation of datasets.
c) The Data Exchange Preparation: Allows the preparation and execution of data contracts and manages user private credentials.
d) The PISTIS Data Exchange Market: Enables the trading of datasets through monetary transactions.
e) The Data Exchange Governance: Facilitates the design and execution of data contracts and monitors the privacy and security of datasets during their whole sharing lifecycle.

Deliverable D3.1 comprises the stepping stone where the design and development activities of WP3 "PISTIS FAIR Data Trading and Value Exchange/Monetisation Platform Services" will be based on, driving the process for the final selection of the methodologies and technologies that will be devised for the implementation of the PISTIS Market Exchange Platform. This deliverable has received input from deliverable D1.1 "PISTIS Operation Principles and Context Detailing" regarding the high-level conceptualization of PISTIS Market Exchange Platform and is interconnected with deliverable D2.1 "Data Interoperability, Management and Protection Framework" for the interrelation among components of WP2 "PISTIS Data Spaces Factory and Trusted Data Management Services" and it will provide feedback to deliverable D4.1 "PISTIS Reference Architecture and API Documentation" towards the finalization of the overall PISTIS Architecture and the planning of integration processes. It will also constitute the basis for the deliverable D3.2 "Monetization and Trading services – Alpha version" for the design and development of the first release of the PISTIS Market Exchange Platform.

Table of Contents

List of Figures

Terms and Abbreviations

AEM	Automated Exchange marketplace
AISP	Account Information Service Provider
AMM	Automated Market Maker
API	Application Programming Interface
BDVC	Business Data Value Canvas
CDP	Collateralized Debt Position
CW-PoW	Compute and Wait Proof of Work
DAG	Directed Acyclic Graph
DAML	Digital Asset Modeling Language
DAO	Decentralized Autonomous Organization
DaVe	Data Value Vocabulary
Dapp	Distributed appilcation
DCAT	Data Catalog Vocabulary
DEX	Data Exchange
DHT	Distributed Hash Tables
DLT	Distributed Ledger Technology
DNL	Data Nutrition Label
DPoS	Delegated Proof of Stake
DRM	Digital Rights Management
DVM	Data Value Map
DVN	Data Value Networks
DVP	Data Valuation Process
DQV	Data Quality Vocabulary
ECB	European Central Bank
EVM	Ethereum Virtual Machine
EU	European Union
EUT	EURECAT (PISTIS partner)
FAIR	Findable Accessible Interoperable Reusable
FIAT	FIAT money
GDPR	General Data Protection Regulation
HTML	HyperText Markup Language
INATBA	International Association of Trusted Blockchain Applications
IoT	Internet of Things
JSON	JavaScript Object Notation
KCDVC	KNOW Center Data Value Check
KPI	Key Performance Indicator
LIME	Local Interpretable Model-agnostic Explanations
LLC	Limited Liability Company
MAE	Mean Absolute Error
ML/AI	Machine Learning/Artificial Intelligence
MFA	Multi-factor Authentication
MSE	Mean squared error
M2M	Machine-to-Machine
NFT	Non-Fungible Token

NLP	Natural Language Processing
NIST	National Institute of Science and Technology
OECD	Organisation for European Cooperation and Development
PBFT	Practical Byzantine Fault-Tolerance
PISP	Payment Initiation Service Provider
PSD2	Second Payment Services Directive
PoH	Proof of Humanity
PoS	Proof of Stake
PoW	Proof of Work
RDF	Resource Description Framework
RFID	Radio Frequency Identification
RSME	Round Squared Mean Error
SCA	Strong Customer Authentication
SDK	Software Development Kit
SNFTs	Sharable NFTs
TEE	Trusted Execution Environment
UX	User eXperience
W3C	World Wide Web Consortium
XAI	eXplainable Artificial Intelligence
XML	eXtensible Markup Language
YML	Yet another Markup Language
ZKP	Zero Knowledge Proof

1 INTRODUCTION

The deliverable D3.1 documents the methods and technologies that are considered for the design and development of the different components of the PISTIS FAIR Data Trading and Value Exchange/Monetisation Platform, which will be responsible for the valuation, trading and monetization of datasets in PISTIS.

All WP3 partners within the M4-M9 project period have collected and analysed valuable material that will be evaluated in the next period of the project towards selecting the most preferred methodological and technological solutions for the implementation of the components that will comprise the PISTIS Market Exchange Platform. The exact details on the internal structure and background knowledge used in each component will be documented in deliverable D3.2 and its following updates D3.3 and D3.4.

In the following image, the different highlighted bundles of the PISTIS FAIR Data Trading and Value Exchange/Monetisation Platform are depicted as part of the overall PISTIS Architecture. These bundles are based on the initial designed architecture described in PISTIS DoA [48], but may be subject to minor changes during the course of the architecture design activities of the project.

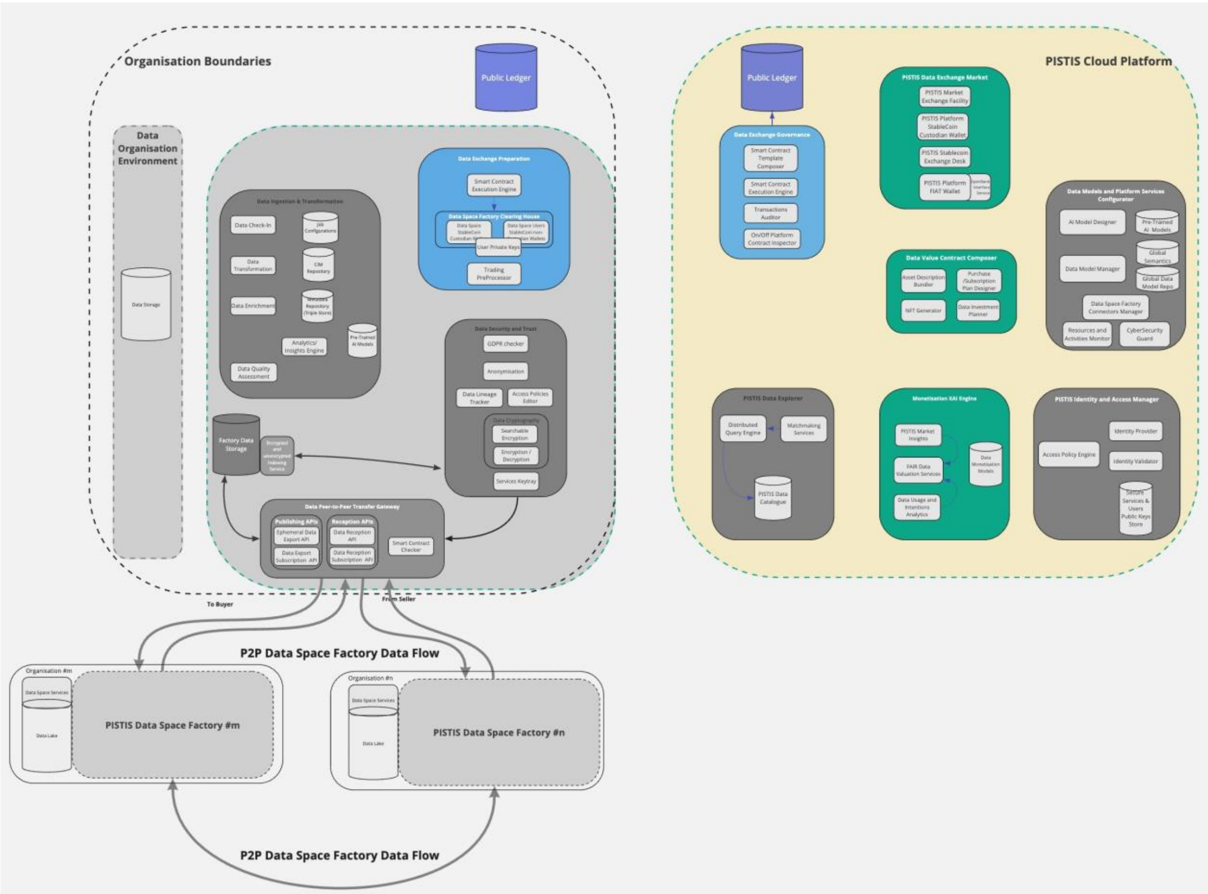

Figure 1-1: The architecture of PISTIS FAIR Data Trading and Value Exchange/Monetisation Platform within the overall PISTIS architecture

The deliverable D3.1 also has obtained input from deliverable D1.1 and will provide feedback to deliverables D3.2 and D4.1 towards supporting the development and integration activities of WP3 and WP4.

1.1 DOCUMENT STRUCTURE

The present deliverable has been structured as follows:

- Section 1 introduces the work performed under M4-M9 project period in WP3 and presents the structure of this deliverable.
- Sections 2, 3, 4, 5, 6 describe the candidate methods and technology stacks for the design and development of the different components of the PISTIS Market Exchange Platform.
- Section 7 provides the conclusions of the work done under the period M4-M9 of WP3.

2 DATA VALUE CONTRACT COMPOSER

The Data Value Contract Composer will provide to users an interface to bring dynamic pricing in the PISTIS Market Exchange Platform and assist data providers with data monetisation by introducing the components described in the following subsections.

2.1 ASSET DESCRIPTION BUNDLER

2.1.1 Overview

The Asset Description Bundler puts together a package which fully describes the asset that is placed for trading, including reference information (e.g., owner, lineage, terms of use, license, pricing, etc.) extracted from metadata.

In addition, it gathers information relevant to data valuation, including output resulting from the sub-processes that feed into data valuation, such as data quality metrics, functional data utility metrics, valuation context (including data usage and purpose), data security and trust.

We expect the Asset Description Bundler to:

- Retrieve metadata from the Metadata Repository and the Data Lineage Tracker;
- Collect the output from FAIR Data Valuation Services, which in turn, will retrieve data and metadata from several components: Data Usage and Intentions Analytics, PISTIS Market Insights, Data Quality Assessment, Repo of Pre-trained AI Models, Anonymisation, GDPR Checker;
- Support the PISTIS Matchmaking Services.

2.1.2 Methods

The goal is to develop a semantic model to describe, store and exchange information pertaining to a data asset:

- Reference metadata (owner, lineage, terms of use, license)
- Quantifications of data quality, functional utility, security, and trust
- An aggregate of these quantifications, with the possibility of tracing back the components of the aggregate
- Price and a description of the pricing method

2.1.3 Technologies

Any kind of key-value type of file format: JSON, YML, XML etc. Ideally, we should be looking to extend and aggregate existing semantic models.

Context interchange format for Data Valuation

This is a straightforward JSON file, which encodes key-value pairs representing data value dimensions. It is the format used for information interchange between components of the Data Valuation Process [60].

Data Value Networks

Data Value Networks [6] is semantic vocabulary for Data Value (DaVe) based on RDF schema, which extends W3C's Data Quality Vocabulary (DQV) and Data Catalog Vocabulary (DCAT).

2.2 NFT Generator

2.2.1 Overview

Based on the EU Blockchain Strategy[1] *"Shaping Europe's Digital Future.)"* and the Digital Europe Programme, the policy for building a data economy which supports sustainable prosperity, participatory economy and regenerative value creation models, is a priority. The strategy involves the following steps for the next 10 years:

- The European public sector is building its own DLT-based services and tokens, which should soon be interoperable with private sector platforms.
- Blockchain is seen to enable transparent accounting and traceability of carbon emissions across product value chains using immutable ledgers and tokens.
- Promotes the use of blockchain technology specifications (e.g., digital assets, token specifications).
- The Digital Europe Programme entails the provision of the European Blockchain Observatory and Forum and the International Association of Trusted Blockchain Applications (INATBA).

In line with this strategy, PISTIS promotes the provision of data (as tokens) incorporating an anti-rival nature of digital goods via its realisation through DLT technologies. By identifying the relevant policy at EU level, a roadmap to support regenerative digital economies and ecosystems will be designed, hence sustainable societies. These will include new protocols on crypto-assets, crypto-currencies and NFTs.

However, until presently, economic institutions function under the theoretical view of the narrowly self-interested maximization of rivalrous resources traded using zero-sum crypto- or fiat-driven currencies, whilst dominant competition policies and market regulation, are strongly rooted in micro-economic theory and game theory according to which all actors seek increasingly more of inherently scarce resources.

Eventually, this has led to poor policy design, and market and socio-political failures in the past recent decades. Unlike traditional monetary systems, where accounting items are standardized, and valuations are universal, anti-rival resources should maintain the same level of integrity as the global monetary economy. Ergo, an immutable system must be introduced. DLT technologies, particularly blockchains, offer the necessary foundations for decentralized and immutable community accounting. Thus, we can create secure and reliable distributed computing systems managed via peer-to-peer architecture, allowing communities to participate in governance, while enabling market trust.

The use of anti-rival goods and services, i.e., information, software etc, by one person does not prevent others from using them. The expenses of producing a "copy" of an anti-rival good is usually negligible. However, the market should find a mechanism to set marginal cost of

[1] https://digitalstrategy.ec.europa.eu/en/policies/blockchain-strategy

production and prices to zero. “Artificial scarcity” (e.g., digital rights management, DRM) could solve this problem, whereby the allocative efficiency would entail more sharing for institutional and systemic progress. This ecosystem architecture will also embed non-fungibility – beyond the common fungibility mechanism of the contemporary economic environment in trading goods and services.

By combining non-fungibility and blockchain technology, NFTs have opened the possibility to have an easily verifiable, publicly available and tamperless digital proof of title for an asset. The technical composition of NFT revolves around 3 important components, I.e., DLT technology, Smart contracts and Tokens.

The PISTIS NFT minting and provision of a data exchange “marketplace” will involve asset preparation and introduction of novel tokens, set-up and storage, pricing – also related to the PISTIS stablecoin mechanism - ownership types, wallet set-up and ultimately transaction abilities.

Furthermore, much of the communication that takes place on the IoT is made up of machine-to-machine (M2M) communications and NFTs are potentially useful because they allow machines to authenticate the data coming in from other machines. NFTs make it to mainstream IoT adoption, are linked to ownership notion.

2.2.2 Methods

In contrast to usual price mechanisms, new systems which efficiently allocate anti-rival resources, must primarily incorporate the externalities created also by resource usage. “Externalities” refer to *indirect benefits or costs which arise from the actions of others in the ecosystem*. In PISTIS, we are especially interested in *positive* externalities. This lays directly in contrast to contemporary market regulation systems, which mitigate negative externalities.

Consequently, positive externalities that will be embedded in PISTIS architecture (via new NFT and cryptocurrency protocols) will promote the use of a resource by one agent while at the same time the “well-being” of others and the “health” status of the entire ecosystem, *will increase without compensation.*

PISTIS's novel approach is analogous to the interoperability via which Bitcoin networks allow the *instantiation of the Bitcoin cryptocurrency*. Nevertheless, while Bitcoin creates artificial scarcity, the value of the new NFTs introduced in PISTIS, will not be based on scarcity but on the underlying human relations, and their “*positive utility function”*, monetised by specific value units.

Up until now, sharable goods in the form of positive externalities (quality rating) do not fall within standard market transactions, and remained un-modelled by big-databases, thus market opportunities and innovations, could be modelled via NFTs, toolkits, multi-layer cross-tokenization for ecosystem value chains. From a technical perspective, the well-known NFTs are tradeable mediums of exchange, but Sharable NFTs (SNFTs) will via new smart contracts will capture positive externalities. Herein, sNFTs serve the currency functions of being a metric of value, a medium of sharing, and a store of credit.

Specifically, goods and services are categorized based I) on their “*excludability*” and ii) on their “*subtractability*”. “Orthodox economics” imply that rival and nonrival goods/services are the basic elements of market mechanism design. Under this taxonomy, rival ones comprise excludable and non-excludable agricultural and industrial services and goods (food, appliances, fossil fuel, etc.) and common-pool goods such as fisheries etc. On the other hand, nonrival ones entail community/membership services albeit necessitating monetisation, or non-excludable ones of completely free to the public e.g., public beaches. New mechanism design theory extends the taxonomy to include one more type, namely the “anti-rival” goods and services. Aligned with contemporary state-of-the-art technological evolution, these include the excludable “network goods” I.e., “Fortnite” and the non-excludable ones such as symbiotic services on the internet and goods derived with or without (open source) property rights. PISTIS will encompass all aforementioned categories, all the more so, considering that modern economies are growingly based more and more on “anti-rival” goods and services, which are exponentially developed in alignment with the rate of technical developments in all sectors.

- Under this new framework, data -and metadata- are anti-rival symbiotic goods/services on the network, hence never before stored, transferred, processed and monetised as variant types of NFTs.
- PISTIS NFTs, is the first attempt to introduce and model data and metadata via the use of NFTs.

In particular, the impact on information exchange on the PISTIS application domains could resonate the benefits identified by Kostick-Quenet *et al.* (2022) for the health information exchange like *automating data access and control*, increased *transparency* and *efficiency.* However, although “*automating data-sharing agreements, smart contracts can address long-standing inefficiencies and the lack of transparency*” [34] data security and privacy, intellectual property rights, and guaranteeing, equity, sustainability, and trust by the different agents engaged in sharing and using the NFTs are still not solved. These issues are nonetheless addressed by the PISTIS project as shown in this and other deliverables. Taking the above issues into account, systems dynamics modelling used as our data flow mechanism – which are not bounded or linear - could include allocative (NFTs) and non-allocative (SNFTs) feedbacks, hence classical mechanism design theory might not apply, and we go beyond Ostrom’s axioms introducing rival, nonrival & anti-rival tokens in order to incorporate “*qualitative data*” with positive or negative externalities (on goods and services).

Moreover, the novel NFTs structure and functionalities shall be linked to the PISTIS monetisation mechanism, I.e., a stablecoin which might be designed to replicate the Euro, fully convertible to its fiat equivalent on a 1:1 value exchange. Once publicly launched (as planned in the post-project exploitation era), both NFTs and stablecoin, may provision the issuance of community/government token(s) to promote stability by linking the value of ePISTIS and NFTs – as well as all tokens of the ecosystem - to fiat Euro in auditable fiat reserves, that will be managed by a qualified asset manager (Automated Exchange marketplace/AEM) with a tilt towards shorter duration notes and bonds, similar to the context of the Euro CBDC by the ECB. This will mitigate NFTs’ and PISTIS stablecoin volatilities and provide transparency for fiat-to-digital currency conversion.

A novel data exchange PISTIS DEM/DEX AMM architecture and its deployed ecosystem will accommodate all NFT attributes and incorporate all crypto-operations and tokenomics for the stablecoin, NFTs and wallets. Via the proposed topology, economic/technical risks/breaches shall be deflected, while enhancing decision-making, prediction capabilities and robust NFT functionalities, efficient monetisation and integration of AI/ML interoperability with PISTIS bundles and the developed UX.

After *"demonstrator" experiments*, a set of requirements will emerge for designing the UX for community platforms using sNFTs as anti-rival tokens. The main attributes of PISTIS topology vis-a-vis the new tokens, will promote:

- *Synchronous and asynchronous Accessibility.* The PISTIS bundles and the overall architecture should be engaging for the users. It is of utmost importance, *for users to set up their own digital* wallets to be able to receive the tokens.
- Immutable Transparency: The new sNFTs should be available and easily accessible in order primarily to increase trust in ecosystems, community rules, and form the mechanisms by which the different types of tokens will be minted in an easy format.
- Scalable Impact: The PISTIS set-up will support intuitive and easy tracking of the impact produced by users when utilising the non-sharable and sharable NFTs, as this will render the basic incentive to keep using the platform.

Sharing schemes and wrapping

The new sNFTs could potentially be extended to support existing NFTs, hence to make sharing functionality available for already existing NFTs, backward compatibility will be introduced. This can be achieved by *wrapping* an existing NFT as a shareable NFT contract.
Four (4) different sharing schemes will be further evaluated:

- *Open sharing* i.e., anyone can share this NFT and view the shared piece in their wallet.
- *Permissioned sharing* i.e., the holder of the original NFT has to first offer the NFT to be shared with the receiver. The receiver can verify sharing and mint a new sNFT or do nothing.
- *Third party permissioned sharing* i.e., the users could operate as a governance mechanism to permit sharing of tokens.

Individual wallets and user interfaces sharing design, i.e., through which they can view, share, and receive sNFT rewards.

(s)NFT Smart Contract description

In particular, the NFT transactions could be implemented through the following component functionality:

SC 1: a registry system on the ledger records the essential information for all stakeholders.

SC 2: a smart contract mints the NFT digital tokens through a series of functions, i.e., i) approves the value of the content owners by verifiers, ii) makes NFT tokens transferrable, burnable and transferable (new functionality) iii) burns tokens, iv) mints NFTs as a badge of credit validation.

SC 3: a smart contract with a multi-signature allows verifiers to verify the content and the minting of the NFTs..

SC 4: a DEX/AMM smart contract allows for automated trading of the tokens with digital money, provides incentives for liquidity providers by charging transaction fees, and estimates (FAIR) the dynamic price discovery for the content in a free market.

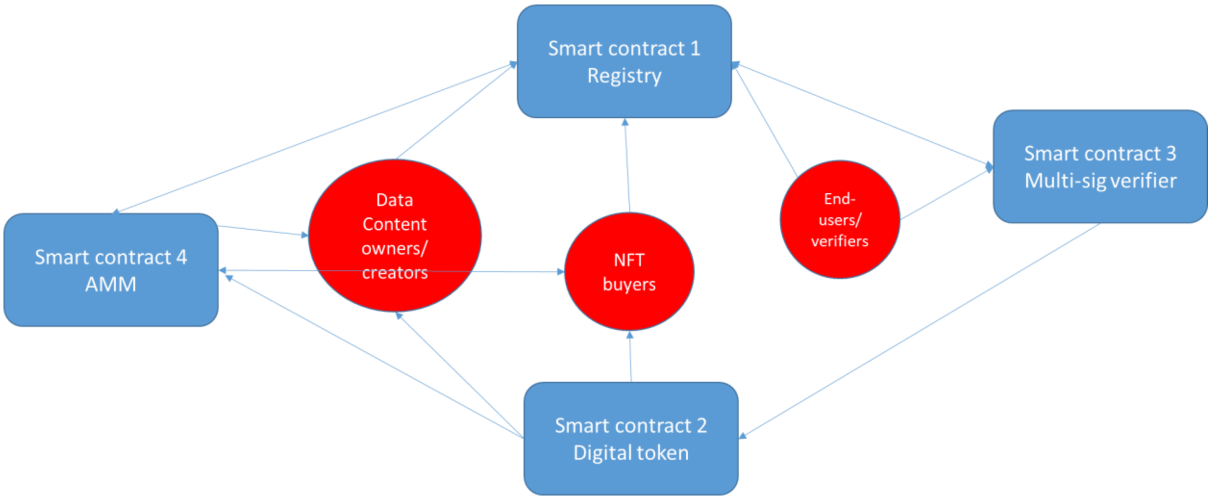

Figure 2-1: NFT smart contracts

2.2.3 Technologies

PISTIS NFT technologies will specifically focus on i) recording and collecting metadata of the type of contribution, e.g., hyperlinks, ii) details that define the agent contributed satisfaction/utility, iii) pseudonymous wallet addresses of the contributors and iv) sNFT transferability registry capabilities, to contain a reference to their origin (wallet address) or the entire originator vector information (wallet address and smart contract address). PISTIS approach will enable further development, for instance serving as an incentive inside the ecosystem to its own members so as the sNFTs could serve as bridges between other DAO communities.

Ethereum Virtual Machine ledger[2]

An EVM ledger hosts the smart contracts which will define the behaviour and characteristics of the novel sNFTs. A distributed ledger solution to handle large NFT sharing transaction volumes with minimal fees (e.g. an EVM compatible consortium blockchain), will add functionality such as a more scalable ledger solution with low transaction fees, advanced metrics and sharing operability, shall entail the option to convert already existing NFTs from other ecosystems of ERC-721[3] or ERC-1155[4] interfaces into the new sNFT protocol.

New patented protocol and token type

The new token type could be built upon an existing smart contract standard that implements non-fungible tokens, such as ERC-7216 or ERC-11557, or via a completely new *patented*

[2] https://www.ledger.com/academy/glossary/ethereum-virtual-machine
[3] https://ethereum.org/en/developers/docs/standards/tokens/erc-721/
[4] https://ethereum.org/en/developers/docs/standards/tokens/erc-1155/

protocol. The new attributes could support a low-cost shareability of the tokens while preserving the "sharing trail". The smart contracts will be deployed on any network using Solidity[5], Hardhat[6], Pinata[7], Alchemy[8] etc.

The hosting NFT and sNFT wallet will have the specific .js, plug-in installed, and an API Key will be created by Node.js. Then, it could be compiled, deployed, tested, and debugged via Hardhat as a MyNFT.sol[9]. Afterwards, the contract classes will be imported, compiled and tested on the VM ledger following a similar process. The NFTs will acquire minting properties, the metadata will be configured using IPFS, and a transaction will be signed on the network. The final collectable in the wallet, could be uploaded (besides PISTIS) on platforms such as OpenSea[10], Rarible[11].

2.3 Data Investment Planner

2.3.1 Overview

The Data Investment Planner represents an innovative approach poised to transform the way we approach the financing of data set management and trading. Its visionary concept enables participants in the supply chain to create an entirely different financing framework. This framework involves sharing capital, which could take the form of a percentage of trading profits or other pre-defined benefits with investors. This innovative approach offers data owners and providers the opportunity to raise investment by offering a share of the potential profits generated by their data. In essence, it is a way for data generators to obtain financial support. This synergy facilitates the delivery of sophisticated backend monetisation services that address the needs of those on the supply side of the data set market.

2.3.2 Methods

The operation of the Data Investment Planner could involve several possible methods, which may favour integration into the platform as well as extract the best outcome:

- Capital allocation: where the platform allows data owners to allocate a value to the data based on trading revenues or other pre-determined returns to buyers. This allocation is automated and transparent, ensuring that all parties have clear visibility of the capital allocation.
- Trade pre-processing: The Data Investment Planner interacts with other modules to facilitate the negotiation of data sets. this would ensure that data owners and data providers can monetise knowing the benefits of the transaction before it takes place, with the assurance that in this way they are transparent and secure.

[5] https://github.com/ethereum/solidity
[6] https://hardhat.org/
[7] https://www.pinata.cloud/
[8] https://www.alchemy.com/
[9] https://github.com/OpenZeppelin/openzeppelin-labs/blob/master/initializer_contracts_with_args/contracts/MyNFT.sol
[10] https://opensea.io/
[11] https://rarible.com/

- Communication with stakeholders: for an efficient and user-friendly design it is necessary to communicate with the different stakeholders (data compilers/providers) to have an estimate of the cost of the data packages.

2.3.3 Technologies

The design and implementation of the Data Investment Planner will leverage technologies and tools to ensure robust, secure, and efficient operation:

- Blockchain Technology: Utilized for secure and transparent equity distribution and trading transactions. Blockchain ensures that all transactions are immutable, transparent, and verifiable by all parties involved. Tools like *Ethereu*[12] (A decentralized platform that runs smart contracts. It is widely used for creating and managing blockchain systems) and *Hyperledger* [13] (An umbrella project of open-source blockchains and related tools started by the Linux Foundation) can be implemented.
- Smart Contracts: Employed to automate the equity distribution process, ensuring that investors receive their fair share of trading revenues or other benefits as specified by data owners/providers. One technology that provides a development environment, testing framework, and asset pipeline for Ethereum is *Truffle Suite*[14].
- API Integration: The platform will integrate with the Trading Pre-Processor through APIs, ensuring seamless communication and transaction processing between the two components. With technologies like *Postman*[15], a popular API client that makes it easy for developers to create, share, test, and document APIs that facilitates seamless communication between the different modules, or *Apigee*[16], Google Cloud's full lifecycle API management platform that enables API providers to design, secure, deploy, monitor, and scale APIs.
- Cloud Computing: Leveraged for scalable and flexible infrastructure, allowing the platform to efficiently handle varying loads and ensuring high availability and reliability, one of the best options is *Microsoft Azure*[17]that offers a cloud computing service created by Microsoft for building, testing, deploying, and managing applications and services through Microsoft-managed data centers.

2.4 Purchase/Subscription Plan Designer

2.4.1 Overview

The Purchase/Subscription Plan Designer, also forming part of the Data Value Contract Composer, is designed to facilitate and manage data exchanges between data stakeholders, particularly those on the demand side, and data providers. Its primary objective is to define and govern how an asset, (either data or data-related services), can be made available to these stakeholders through two main options: a) subscription plans, involving the provision of a data

[12] https://ethereum.org/en/
[13] https://www.hyperledger.org/
[14] https://trufflesuite.com/
[15] https://www.postman.com/
[16] https://cloud.google.com/apigee
[17] https://azure.microsoft.com/en-us

asset to stakeholders with regular and continuous access to specific data assets or services for a predefined period, typically on a recurring basis. These plans are often chosen by data consumers who require access to a continuous stream of data or data updates; and b) simple purchases, where data can be made available through simple one-time purchases. This option suits data consumers who have specific, one-time data requirements or prefer a pay-as-you-go approach rather than committing to recurring subscriptions.

Overall, by incorporating the Purchase/Subscription Plan Designer into the Data Value Contract Composer, data providers can effectively define the terms and conditions of data access, empowering data stakeholders to choose the most suitable and convenient way to access the data they require while ensuring a fair and transparent data exchange process for all parties involved.

2.4.2 Methods

Some indicative methods that the operation of the Purchase/Subscription Plan Designer will be based on are the following:

- Purchase/Subscription Plan Creation: Through a user interface, data providers can create subscription plans by defining details, such as the type (i.e. recurring subscriptions or one-time purchases), access duration, value, potential renewal options, pricing methods (e.g. tiered pricing, flat fees) and access policies.
- Plan Template Library: For the facilitation of plan creation, data providers may access a template library, which includes pre-designed subscription plan templates that can be edited to match specific needs.
- Data Preview and Sample Presentation: Data previews and samples will be available to allow potential subscribers to have a quick idea of the dataset quality prior to subscription.
- Payment Plan: When a subscription plan is selected by a data consumer, the payment process will be initiated through the integrated stablecoin mechanisms of the PISTIS Market Exchange Platform.
- Data Access Policy: Access control to datasets is enabled according to subscription plans designated by data providers, safeguarding the security and privacy of data on trade by setting which data consumers may access these datasets.
- Automated Renewal and Notification Process: In case of recurring subscriptions, the Subscription Plan Designer will send notifications to subscribers before renewal date, towards confirming or changing their subscription.

2.4.3 Technologies

The implementation of this component will not be based on any existing technology due to the scope and nature of the component. Ideas for the implementation will be borrowed by the IDS Connectors implementation, however the component in focus is rather much smaller than a trading system that can be defined in the IDS architecture. As such, the implementation will be performed using state of the are development frameworks, such as:

Frontend Technologies

- Vue.js[18]: An MIT-licensed open-source language, with a progressive, incrementally adoptable JavaScript framework for building UI on the web.

Backend Technologies

- Node.js[19]: As an asynchronous event-driven JavaScript runtime, Node.js is designed to build scalable network applications.
- Django[20]: Django is a free, open source and written in Python back-end web framework.
- PostgreSQL[21]: PostgreSQL is a free and open-source relational database management system emphasizing extensibility and SQL compliance.

[18] https://github.com/vuejs
[19] https://nodejs.org/en
[20] https://www.djangoproject.com/
[21] https://www.postgresql.org/

3 MONETISATION XAI ENGINE

The Monetisation XAI Engine is equipped with a fully-fledged bundle of data analysis services for the extraction of data usage intelligence and data utilization intentions across the whole data value chain, as described in the below subsections.

3.1 PISTIS MARKET INSIGHTS

3.1.1 Overview

Defining the value of the data is closely related and influenced by market insights, as the prices and the demand of similar assets provided by competitors can indicate up to a certain level the appropriate asset value. Furthermore, different functions within an organization have an interest in various data or data sets that may have diverse market value. Also, the value of data can be different at the time of their acquisition and of reselling or renting.

As to these issues, PISTIS places great importance on providing the stakeholders accessing the data spaces and sharing data with insights on how data and data sets are assessed and valued across markets or within one or more specific marketplaces. That's the specific goal of the PISTIS Market Insights component, made up of methods for market analysis and a set of technologies implementing them.

Consequently, in what follows, we are going to refer first to methods that can be used to extract and communicate market insights to the data providers and consumers. Then, we outline a set of technologies that implement them and are considered by PISTIS or can be integrated into it.

3.1.2 Methods

Important information like statistics, key performance indicators (KPIs), and dashboards that combine different state-of-the-art metrics can be used to gain insights into a data marketplace and help better understand the potential usefulness, purpose, and value for each dataset.

- Aggregated metrics that showcase the total amount of datasets currently existing in the market, the rate at which new datasets are shared, and their total data size and volume (e.g., number of rows). Dataset quality metrics can also be used, e.g., average percentage of completeness/uniqueness/timeliness of all datasets.
- Purchase information can also be visualized in the form of KPIs like frequency of successful purchases, price evolution over time, percentage of change, volume of purchases and total PISTIS coins exchanged over a period (previous day, week, month, etc.). Other relevant insights can be a list of top sellers/buyers by volume of assets provided/purchased either as individual organizations or as groups of entities e.g., electric power companies.
- Market trends can showcase what is the current and projected demand/popularity for specific types of datasets and their min/average/max price. This, for example, can be achieved by charts that visualize market growth over time for datasets of the same domain or industry that highlight emerging technologies. Also, a list of the top assets in sales accompanied by extra quality profiling analysis (completeness, etc.)

The previous insights can also be enhanced through the usage of interactive filters that allow the data consumers/providers to explore the available datasets according to their preferences. For example, dataset filters can include price ranges, categories, type, domain, stakeholders, intended usage, exchange type, data provider name, origin, size and format, and any other available information. Nevertheless, an analysis on asset level may be provided as well, as it can assist consumers that have more targeted preferences.

Furthermore, a side-by-side comparison between the owned/purchased datasets versus the overall exchanged in the PISTIS market may also provide a tool to determine, for example, if an asset is under-priced or overpriced.

In addition, valuable insights can be achieved through price monitoring processes, which collect and analyse data from various data sources such as competitors' websites and online marketplaces and can be leveraged to help estimate the data value. Price monitoring can be achieved either though manual processes or automated tools.

Manual processes require regularly visiting the competitors' websites and marketplaces to gather price data, however, manual price monitoring tends to be error-prone and time consuming especially when there is a high number of competitors that needs to be taken into consideration.

Automated price monitoring utilises web scraping tools to gather, analyse and present in an organised manner the outcomes of the analysis by using dedicated dashboards[22].Thus, by identifying the markets' trends, the users can achieve an optimal data value estimation in a fast and easy manner, especially for the users who are newly introduced in data markets.

An analytics model can be later formulated and built leveraging the information gathered from of the price monitoring analysis as well as consumers demand and buying patterns for similar data assets to infer the optimal value for the assets at hand. The analytics models can either be simple machine learning models, such as XGBoost[23] or more sophisticated such as neural networks [54]. Reinforcement learning is also utilised for the task of data valuation and pricing [31].

3.1.3 Technologies

Pandas[24]

Fast, powerful, flexible, and easy to use open-source data analysis and manipulation tool built on top of the Python programming language. It can be used to aggregate data, calculate statistics, extract trends, etc.

Apache Echarts[25]

[22] https://brightdata.com/products/insights?gspk=YWltdWx0aXBsZW91MjI4MA&gsxid=UBtLP3oGoRei&pscd=get.brightdata.com&sid=price-monitoring&utm_source=affiliates&utm_campaign=YWltdWx0aXBsZW91MjI4MA)-

[23] https://xgboost.readthedocs.io/en/stable/

[24] https://pandas.pydata.org/

[25] https://echarts.apache.org/en/index.html

Free, powerful charting and visualization library offering easy ways to add intuitive, interactive, and highly customizable charts.

Elasticsearch[26]

A distributed search and analytics engine. It serves as a centralized repository for your data, enabling rapid and highly relevant searches while offering robust and scalable analytics capabilities. Provided the metadata and extra information of the assets, it can be utilized to search and filter assets on the marketplace.

BeautifulSoup[27]

A Python library for parsing HTML and XML documents. It is useful for web scraping price data from external sources.

Scikit-learn[28]

A python library widely used for predictive analytics and machine learning models. It can be leveraged to build predictive models fort the optimal asset price based on the information gathered from the market insights service.

XGBoost

An optimized distributed gradient boosting library that implements machine learning algorithms under the Gradient Boosting framework. Can provide models to predict the optimal price of data based on the price monitoring analysis.

3.2 FAIR DATA VALUATION SERVICES

3.2.1 Overview

The FAIR Data Valuation Services are a set of methods that support the process of data valuation. As a reminder, data is an economic good with different properties than the traditional private goods, since it can be easily duplicated and sold multiple times [35].

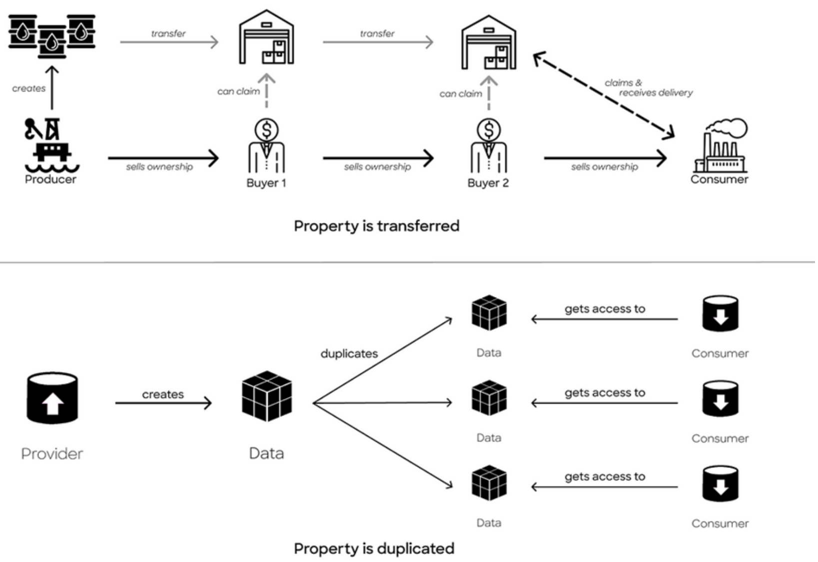

Figure 3-1: What it means for data to be a non-rival good, adapted from Ocean Protocol

[26] https://www.elastic.co/
[27] https://pypi.org/project/beautifulsoup4/
[28] https://scikit-learn.org/stable/

The value of a dataset can also depend on the price and availability of similar datasets [1], hence it is important to provide methods to find similar datasets and utilize this information. Some of these methods are also mentioned in section 3.3.2.

According to Sim et al. [55], the principles that drive the worth of a dataset and make it more desirable than others are the following:

- Monotonicity: Having more data is more valuable, especially when coming from multiple sources.
- Submodularity: Same data is less valuable when added to a larger dataset.
- Similarity to reference distribution: A dataset is more valuable when it is similar and representative of the target task's distribution.
- Intra-diversity: A dataset is more valuable if it contains more diverse data points covering a larger region of space.
- Replication robustness: Since adding the same data to an ML model does not improve its performance, then replicated data should not be as valuable.
- Cost: Data that is costly should be more valuable
- Timeliness: Data that is up to date is more valuable

Moreover, according to a blog post published in Ocean Protocol[29], one must keep in mind that data is an economic good with different properties than the traditional private goods, since it can be easily duplicated and sold multiple times.

Data valuation is the process of assigning a quantitative value to a data asset. It is a complex, context-dependent, and multi-dimensional process, built upon several other processes [61]: data and metadata quality assessment, functional data utility assessment, legal and ethical assessment, privacy analysis. The consensus in both business and academic literature is that data has value, however this is difficult to quantify [17]. Some of the reasons for the inability of unlocking the economic value of data relate to organisations' narrow thinking or the fear of breaching privacy or legal frameworks. Each of the processes previously mentioned can consist of several other sub-processes of their own (e.g., DQA may consist of computing several data quality metrics – timeliness, validity, completeness etc.), whose quantifiable results need to be aggregated over several levels to obtain an overall result – the value of the data asset. The overall value, as well as its subsumed data value dimensions will be used for:

- Reporting the value of data at different aggregation levels, thus increasing the transparency of the method and promoting the trustworthiness of the PISTIS data market.
- Support the data monetisation services, by providing the building blocks for methods that are trying to assign monetary value to data assets.
- Support the matchmaking services (Section 2.4 in Deliverable D2.1), by providing the building blocks for building recommender systems to match users and data assets.

To establish the context for data valuation, the FAIR Data Valuation Services will be using information from the PISTIS Market Insights and Data Usage and Intentions Analytics.

[29] https://oceanprotocol.com/

Information pertinent to data value will be stored and exchanged via the PISTIS Asset Description Bundler.

3.2.2 Methods

In this Section we outline the methods applied for the operation components. As to this issue we report here some of the main points advanced in the Deliverable D1.1. First, while there is still no standard way to assign value to data, Fleckenstein et al. [24] identify three main categories for classifying data valuation models:

- *market-based models*, which see data value as cost and revenue.
- *economic models*, which target financial and public benefit.
- *dimensional models*, where data value is based on different dimensions.

Before presenting a review of existing methods for data valuation, we would like to introduce a brief discussion on two key aspects that influence this process: *valuation contexts* and *data value dimensions*.

Data valuation contexts

While their roles in data value assessment is generally acknowledged, contexts raise two great challenges: i) definition and technical formalisation, and ii) a method of quantifying their impact on the value of data.

Previous work related to the definition and use of contexts originates in data quality literature. Cai and Zhu note that "*data quality depends not only on its own features but also on the business environment using the data, including business processes and business users*"[14]. Pipino et al.[46] differentiate between task-independent and task-dependent assessments of data, with the latter consisting of organization business rules, company and government regulations and technical constraints. Askham et al.[4] discuss the dependency between data quality assessment and the context in which it occurs, and mention that organisations should consider not only quality dimensions but also organisational requirements for data, as well as the impact of non-compliance. Even and Shankanarayanan[20] observe that contexts are often disregarded when designing data quality frameworks and suggest that the value of the same data may ultimately depend on "*contextual factors, such as the organisational level at which the data is used, the specific task, and/or the personal preferences of the decision makers*"[20].

A data valuation method that generalises to all types of data is an ambitious goal and attempting a holistic view of data adds technical complexity due to the heterogeneity of data. Classification of various data types should allow for a first level of contextualisation. The PwC report on data value [50] provides several classification dimensions for data: domain, services, or applications that they enable, source (authored, user provided, captured, derived), content (master data, transactional, reference data, metadata, unstructured), usage, generation method (volunteered, observed, inferred), and ownership.

Dodds[18] provides a template for classifying datasets, including information on characteristics, capabilities of the publishing organisation and its available funding, the purpose behind publishing the data, and the ethical, legal, and social contexts in which it will be used. Based on these, the author describes 10 data types: study, statistical index, sensor feed, register, database, description, personal records, social graph, observatory, forecast.

Data value dimensions

Earlier in this sub-section we have mentioned how a whole class of data valuation methods is built around the concept of data value dimensions. Indeed, while existing literature recognises their importance to data valuation, there is no consensus on what these dimensions are, their relations to each other, aggregation rules, or how they influence data asset monetisation. Moreover, most research simply enumerates dimensions, and there seems to be a gap in terms of formal definitions of dimensions and metrics, in a manner similar to data quality assessment.

Information capacity is one of those dimensions and can be defined as "*the current stock of understandings informed by a given installed base*" (Viscusi and Batini)[64] representing "*the potential of a digital information asset that can be defined and evaluated independently from the usage*." The information capacity of an organization, on the one hand, determines "*the economic utility of a digital information asset*"; on the other hand, it enables capabilities that are "*the possibility and/or right of the user or a user community to perform a set of actions on a computational object or process*" (*Ibid.*).

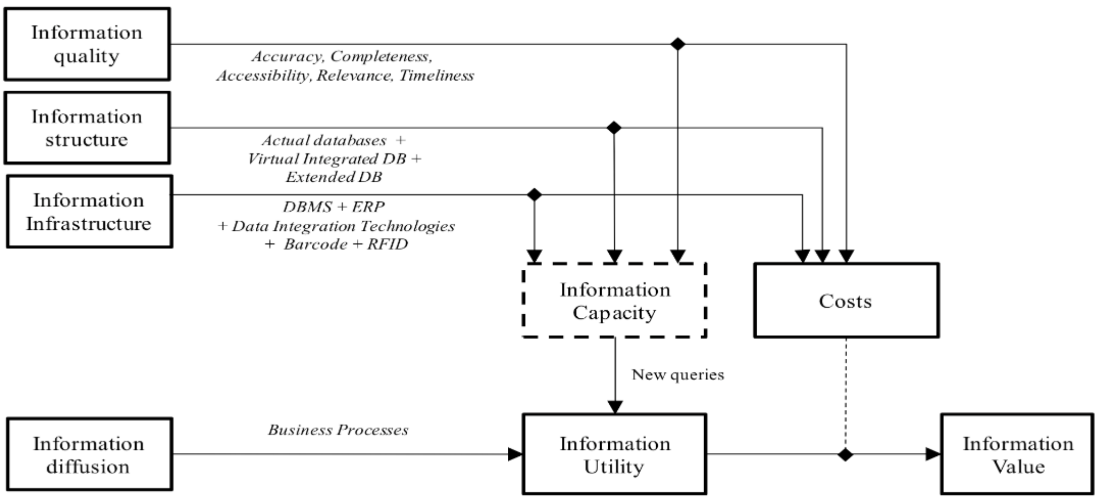

Figure 3-2. A model for the assessment of information value, adapted from Batini et al. (2018)

As shown in the model of Figure 3-2 proposed by Batini et al. [8], the information capacity of an organization is made up of different dimensions like:

- *data quality*, the degree of *integration of data*,
- *infrastructure* (Data Base Management Systems, Enterprise Systems, Data Integration technologies, Information and Communication technologies like Barcode, RFID, etc.).

Those dimensions represent the *costs* that should be considered for the data valuation together with the *information utility* coming from the diffusion on the organization process of the information associated with the data. As to the data, according to Ahituv [2] they must have relevant attributes such as *timeliness*, *contents*, and *format.*

A way to understand the information utility for organizations' processes is to identify the core activities they are part of, like a *data value chain* [21]. As to this issue, Lim et al/ [37] propose a data–value chain as a nine-factor framework for data-based value creation in information-intensive services, whose spectrum goes from the *data source* through *data collection* and *data analysis* to the *use* made by the final user.

The study by PwC [51] considers that data value is built upon a combination of data quality, interoperability, and restrictions (legal, risks). Belleflamme directly connects it to the 4 Vs of big data (volume, veracity, variety, velocity) [9], while other approaches [40],[41] link it to its position along the data value chain.

With these considerations in mind, we now resume the discussion about existing *data valuation methods*.

Brennan et al. [12] develop a simple survey-based method for data value assessment, centred around five dimensions adapted from Sajko et al. [52]: operational impact (utility), dataset replacement costs, competitive advantage, regulatory risk, timeliness. Initial assessments of 34 datasets (from various domains), by a group of 20 data experts from both industry and academia, have revealed operational impact as the most important data value dimension.

Kannan et al. posit that the value of data is better quantified when used in an application and similarly to Mawer [41] they conclude that data increases in value as it advances through a processing pipeline [30]. In their framework, data is characterized by multiple intrinsic and extrinsic facets, with some gaining precedence depending on the context. They propose a questionnaire-based method to gather objective responses (binary, quantitative, categorical) that characterise a data set. In the simplest of cases, all facets are treated equally; then, to account for the many possible applications, the composition of the questionnaire and the relative values of the responses could be adapted to the context or the role of the respondent.

Data sheets for data sets

A similar approach is that of Gebru et al [26]., who promote the creation of data sheets for data sets, with the main goal of increasing the communication between creators and consumers. Their work derives from the IEC Datasheets, common in the electronics industry, and is somewhat akin to the documentation that accompanies clinical trials in the United States. The creation of datasheets is the responsibility of the data set creator, through a process of reflection on the creation, distribution, and use of the data set, including assumptions, potential risks, and implications of use. This should equip consumers with the tool they need for making an informed decision about using the data set, considering its content, collection process, recommended uses, restrictions and including the stated assumptions, risks, and implications. As opposed to Kannan et al., data sheets do not offer a quantitative valuation; on the contrary, the authors are trying to avoid binary responses and encourage creators to be as detailed as possible. The process focuses on details about motivation, composition, collection, pre-processing-cleaning-labelling, uses, distribution, maintenance.

Data Nutrition Label (DNL)

DNL [29] is a diagnostic framework which provides a concise, robust, and standardised view of the core components of a data set. The goal of the DNL is to inform and improve the selection and interrogation of data sets and is primarily aimed at data specialists. The solution is composed of several loosely coupled modules, covering various facets of a data set, some with a degree of context-dependency: metadata, provenance, variables, statistics, pair-plots, probabilistic models, ground truth correlations.

Business Model Maturity Index

According to the business model maturity index, the data value is estimated based on its contribution to an outcome. This involves identifying various use cases to reach a desired outcome (e.g., sales increase), and all the data required for each use case. Following, the cost and impact of each use case are calculated (used to estimate the value of each use case) along with the relative significance of each piece of data for each use case. The data that are most impactful are those with the highest value. This technique heavily relies on business expertise, as the computation of use case costs and impacts is rooted in the organization's operational processes and framework. Expert judgment is also required in assigning the relative importance of each dataset in this process[30].

Decision-Based Valuation

The Decision-Based Valuation follows a similar approach to the business model maturity index, but it also adjusts the value of the data based on how the data value may change over time, the quality of the data relative to its end use, and the degree of effort needed to convert data into usable information. Various use cases are identified along with the required data and their cost and impact are calculated to estimate the value of each use case. Following, the accuracy of the data and the collecting frequency are assessed based on the needs of the use cases, and a quality score is extracted used to adjust the use cases value. Next, the data acquisition and preparation cost are estimated, and the final data value is computed by taking into consideration both the relative contribution of the data and the data acquisition costs[31].

Supervised machine learning framework

This approach can be used to predict the approximate value of a dataset based on market sales history. A simplified approach of how this method works is:

1. **Data collection**: Information from historical sales is collected and stored including metadata, quality assurance metrics, price, date, and other relevant features.
2. **Feature engineering**: The pre-processing of the dataset includes finding the feature set that is most relevant to predicting a dataset's price, filling any missing values, discarding outliers and erroneous records, and encoding the selected variables for training using scalers and one-hot encoders. Then, splitting the pre-processed dataset into training and test sets to evaluate the performance of the model accurately.
3. **Model training**: Training of multiple models using the training set and evaluation on the test set. Then, selection of the model with the best performance given specific regression metrics (e.g., MSE, MAE, RMSE).

[30] https://internetofwater.org/valuing-data/business-model-maturity-index-method/
[31] https://internetofwater.org/valuing-data/decision-based-valuation-method/

4. **Model deployment**: The best trained model is now deployed on the platform to predict a dataset's price. The model receives the input of a feature set and outputs the predicted price.
5. **Adjustment on model drift**: The continuous monitoring of the performance of the model. If the predefined metrics heap away from a specific threshold, then the appropriate alerts are triggered, and model retraining/refinement takes place using newly acquired data points, if necessary.
6. **Explainability**: Techniques to explain the model predictions and what are the important factors that led to output include model agnostic explanations (e.g., LIME), Shapley values, Anchors, etc.

Using this framework, the model can provide a prediction for the value/price of a dataset given closely related sales of similar datasets.

Cost-based approaches

According to [56], the value is given by the full cost of collecting, storing, analysing the data and releasing data products. This method requires clear records of the finances allocated for each of these activities. Even though such information should exist in an organisation's financial records, it may be difficult to extract and compile to obtain faithful estimates. Another drawback of this method is that it doesn't it restricts the quantification to the production of data assets, which doesn't tell anything about further benefits that these may generate.

Market-based approaches

The value is given by the market price of equivalent products or by the willingness of participants to pay. *Market-equivalent pricing* is establishing the value by looking at similar products in the market. In this setting, issues may arise when comparing with data assets that have gone through transformations that increased their value or with the undervaluation of open ("free") data. *Stated preferences* simply ask users how much they are willing to pay or receive in exchange for a data asset; this relies on the assumption that these are expert users, who understand very well the properties of the asset they are trading. *Revealed preferences* aims to understand the value of an asset, when compared to other assets, under a limited budget by the interested user. This method is used in cost-benefit analysis and appear to be credible, however can be limited in purpose.

Income-based approaches

"Value is defined by the productivity improvements and future cash flows that can be derived from data". These methods are exposed to inaccurate estimations, by either overestimating the value of data (if alternate data sources are not known at the time of the valuation), or by underestimating (by missing positive externalities).

Benefit monetisation

The value is estimated by looking at the benefits resulting from using a data asset and assigning them a monetary value. This method can be applied in the case of open data (e.g., census), whose benefits can reflect into education, healthcare, service planning etc. The main drawback is that missing to consider benefits immediately leads to underestimations.

Impact-based approaches

"Value is determined by assessing the causal effect of data availability on economic and social outcomes, or the costs in terms of inefficiencies or poor decisions due to limited or poor-quality data". Such an approach seems to be able to underline the "business" value of data and can be very useful for explaining data value to non-technical stakeholders (e.g., organisation executives, policy makers). Its main drawback appears to be its context specificity.

Data Value Map (DVM)

This methodology [53] is built on the premise that seemingly difficult data problems are in fact insufficiently understood business problems, and that the value derived from data initiatives is enabled by shared understanding. The DVM is a framework that promotes a shared understanding between the stakeholders involved in data initiatives. It consists of:

- six components (4 related to data processing – Acquisition, Integration, Analysis, Delivery and 2 transversal – Business Value and Data Governance) and
- two sets of guiding questions:
    - the first concerning data, people, processes, and technology, applicable to the 4 processing components and
    - a second set, concerning benefits, behaviours, ownership, applicable to business value and governance.

The DVM is a visual tool, whose main benefit is bridging the business-technology gap which often exists in digital initiatives, by a establishing a common language and a shared vision between stakeholders.

Business Data Value Canvas (BDVC)

BDVC [31] is a knowledge-based model which approaches big data management as a cognitive system, by seeking to align technology and business, to create value from data. The model brings together lessons learned from previous data management models: the data lifecycle model, by the NIST Big Data Public Working Group [15], the data value cycle, developed by the OECD [43] and the nine factors framework for data value by Lim et al. [37]. The BDVC however is not conceived for assessing the value of data; instead, it defines a series of business management dimensions (data preparation, data analytics, data interaction, data effectuation and data intelligence) and tries to articulate the potential business value that can be generated from each of them, and the information technology solutions that enable it. The author demonstrates its versatility by applying the methodology to three use cases from various domains (banking, insurance, energy) and suggests that the methodology can be used both for describing existing big data systems, as well as planning new big data solutions. We can consider the application of BDVC to all the use cases provided by the PISTIS partners, thus further testing the versatility of the methodology.

Mapping data properties to data value

(Kannan et al., 2018) propose a survey-based manual method, in which users answer simple questions related to a set of 65 properties related to the data set being valuated. These properties are grouped in 17 facets (e.g., data layout, composition, format, quality etc.) and try to cover as many properties as possible, even at the risk of being repetitive. Each answer is then mapped to a score between 0 and 1 (e.g., for the binary question "Is the data error free?", answering "Yes"=1, whereas answering "No"=0) and for context-free valuation, all facets are equal and the value of the dataset the sum over all its properties. The authors suggest that different facets can change weights with context; however, they do not address the issue of how to estimate these weights for each dataset, which probably means that in a first step they could be manually tuned by the user. As a potential improvement, the authors suggest the extraction or inference of some of these properties from metadata.

Perceived relative importance of data value dimensions

Brennan et al. [12] deploy an online manual survey to collect information about a set of previously researched data value dimensions: operational impact (utility), dataset replacement costs, competitive advantage, regulatory risk, and timeliness. The survey consists of 20 questions covering each of these DVDs, and 4 additional self-reflection questions on the process itself; answers were collected on a Likert scale. 34 surveys were submitted by participants from several business segments (finance, aviation, publishing, legal, ICT), concluding that operational impact has the highest relative importance, followed by timeliness and replacement costs. The authors show how DVD importance can be measured via manual surveys and suggest that the method can be used to collect baseline assessments. They also hypothesise that relative importance might differ for different business sectors and that such a method can be used to provide sector-wide estimates.

Data Valuation Process

The DVP by EUT [61], [62] defines a series of dimensions and integrates them in a multi-layered approach:

1. the context in which the data set will be used; this is provided by the user and covers the areas of System & Economics, Legal & Obligations, Data Science, Data Properties, Business uses.
2. data quality assessment, including state-of-the-art data quality dimensions and metrics: completeness, uniqueness, domain validity, format validity, timeliness.
3. data utility assessment (including model performance and chance estimators).
4. deanonymisation risk assessment (if applicable).

The reporting of the data value is done in a top-down fashion and uses a visual scoring metaphor like energy classes (A-E).

Data Shapely

Ghorbani and Zou [27] propose a method to identify the individual value of data points with regards to a given machine learning model and task. The Data Shapley algorithm computes Shapley values with regards to a triple composed of the dataset, the learning algorithm, and chosen performance score(s). In this approach, data value is tightly connected to the functional utility of the dataset, measured by the performance of training selected ML models. The authors demonstrate the utility of Data Shapley for identifying outliers or misclassifications, improving performance by removing low-value training data or identifying high-value external data, and improving domain adaptation of pretrained models.

3.2.3 Technologies

EUT Data Valuation Component[32]

The Data Valuation Component by EUT considers that the value of data is generated from two main areas: data quality and data usability, which are assessed through the lens of the context in which a dataset will be used. The context is set by the user, during a context gathering procedure, based on which the relevant components of data quality and data usability are established. The tool is trying to maximise the automation degree of all these processes, thus allowing for more in-depth analyses to support the value of data and a reduction of the time dedicated to the data valuation process. It is the only solution for data valuation currently offering the possibility to perform various anonymisations of the dataset and compare their impact on the updated value.

Data quality assessment implements the metrics of domain validity, format validity, completeness, uniqueness, and timeliness. Data utility implements standard algorithms and evaluation metrics for regression, classification, and clustering. Additionally, the component includes a GDPR compliance checker. Data value is reported through scores at various levels of aggregation, explanatory reports, and data profiling.

Data Pricing Recommendation Tool, H2020 i3-Market

A tool [33] specifically aimed at SMEs' needs for assigning a monetary value on structured data sets. The valuation method that it implements includes several of the well-known data value dimensions: costs (collection, storage, analysis), estimated value for the consumer, expected customers, data quality (completeness, accuracy, validity, uniqueness, scarceness) and the type of applicable license. These are weighted by pre-set weights and are finally modulated by the credibility of the data seller (established upon previous reviews).

The tool is part of the i3-Market Backplane, an open-source software integrating data markets via APIs and allowing participants to monetise and trade their data sets.

The tool integrates some of the dimensions available in the EUT Data Valuation Component and provides a slightly more advanced aggregation formula. However, it is not clear what is the basis of this aggregation formula, and no evaluation has been performed.

Saffron

Saffron [5] is a data value assessment tool, developed as an extension of DaVe to enable the representation of the value of data assets. It integrates several of the well-known data value dimensions: usage (provenance derived metrics), quality (completeness and accuracy), trust and data management. The contextual nature of data valuation is addressed by integrating the notions of metrics settings and dimension weights, which are customisable by the user. The Saffron Dashboard presents a visualisation of the value of an entire project (seen as a

[32] https://eurecat.org/en/portfolio-items/

composition of several data assets), as well as individual valuations of each asset and a breakdown for each dimension. To improve traceability, Saffron allows for the tracing of metric values over time and includes explanations for the calculated value.

The tool was tested on a limited number of use cases, however neither a validation report, nor a public implementation are currently available.

KNOW Center Data Value Check[33]

The Data Value Check is a spreadsheet-based tool used "to generate and select promising data-drive use cases based on a cost-benefit logic". If most of the other tools are focusing on the technical dimensions (quality, usability, provenance) that underpin the value of data, the KCDVC is oriented towards the quantification of the economic and legal factors. The methodology behind the tool is briefly explained on the project's webpage, but any other references to the tool and its availability are difficult to find.

Scikit-learn
A simple python library that provides tools for predictive analytics and machine learning models. In our case, it can be used to train a model that will predict the value of a dataset.

Tensorflow[34]
A library for machine learning and artificial intelligence methods. Can be used to train deep neural networks for value prediction.

Evidently[35]
An open-source Python library that can help evaluate, test, and monitor ML models from validation to production.

LIME[36]
A python library for machine learning model interpretation. In our case, it can be used to extract the features and the way they affect the models' predictions regarding the data value estimation.

XAI-eXplainable AI[37]
Contains various tools that enable the analysis and evaluation of data and models. Can provide important explanations for the value prediction model e.g., feature importance.

Alibi explain[38]
An open-source Python library that provides various implementations of widely used explanation methods for machine learning models. It can be used to indicate which features are the most influential and in what way for the data value estimation models.

[33] https://www.know-center.at/en/consulting/offer/ai-vision-strategy/data-value-check/
[34] https://www.tensorflow.org/
[35] https://www.evidentlyai.com/
[36] https://github.com/marcotcr/lime
[37] https://en.wikipedia.org/wiki/Explainable_artificial_intelligence
[38] https://github.com/SeldonIO/alibi

3.3 Data Usage and Intentions Analytics

3.3.1 Overview

Data Usage and Intentions Analytics focus on understanding how data may be used within an organisation based on the underlying intentions of the data owners/users. It involves collecting, analysing, and interpreting data-related activities to gain insights into how data will be accessed, processed and shared by individuals or systems [43].

The objective of Data Usage and Intentions Analytics is to provide data owners with a deeper comprehension of their data assets and the context in which they may be utilised. By analysing data usage patterns, intentions and related online data, organizations can decide the way they will use their data, not only in terms of processing but also even in terms of sharing. Such decisions result in enhanced data quality and security.

Organisations will evaluate the potential usage of their datasets based on analysis of similar online data that will reveal information about the way their own datasets may be shared and traded (e.g. such as the historical frequency of each feature being the target of a predictive task). More specifically, such similar online data will be analysed regarding querying, modifications, duration of usage and sharing in order to showcase analogous utilizations for the owned datasets. Moreover, analysis on the behaviour of the users with previous datasets will be helpful as it will discover patterns that will drive usage recommendations for current datasets. The domains that are more relevant for exploiting a dataset will also be presented to indicate potential markets where the dataset can be traded. In addition, the users may introduce the way they would like to utilize data and the Data Usage and Intentions Analytics, by processing such input, will present the most effective means of exploiting these data. The privacy and security of data can also be safeguarded by focusing on the sensitive aspects of datasets and recommending necessary actions for complying with relevant data protection policies.

Data Usage and Intentions Analytics will be facilitated with the introduction of dedicated dashboards that will illustrate various statistics along with charts and graphs about the use of datasets, enabling the users to better understand how they use their data and bring to light their value, both implicitly and financially, leading to an efficient decision making of the data owners based on the provided insights.

Data Usage and Intentions Analytics in PISTIS can comprise a valuable service for data owners that would like to obtain a comprehensive understanding of their datasets and optimize data utilisation by taking advantage of the maximum value of data while respecting privacy and regulatory issues.

3.3.2 Methods

The process of gathering information regarding usage and intentions analytics can be divided into two primary phases. During the initialization phase, which occurs when the datasets are initially uploaded to the marketplace, some insights about the datasets and their potential applications can be immediately provided. Furthermore, in a subsequent phase, when more data becomes available concerning the dataset's purchases and utilization by other partners, even more comprehensive insights can be offered.

During the initialization phase, information regarding the uploaded datasets can be collected through questionnaires completed by the data providers. These questionnaires may inquire about how the data has been or is intended to be used by the provider/owner, their expected revenue from its use, and other relevant details. Additionally, during this phase, metadata will be gathered as owners are required to provide tags, descriptions, and domain-specific information about their datasets. All this information, combined with data model information, can be utilized to identify similar datasets already available in the marketplace. Hence, the "Data Usage and Intentions Analytics" service can offer insights into how similar datasets have been used by other stakeholders, in terms of their applications and domains, their pricing, and the availability of analogous assets on the market. If external data sources are accessible, they could also be leveraged to discover similar datasets and provide insights based on them as well.

Another valuable insight that can be delivered during this phase is a metric that assesses the applicability of the datasets for machine learning tasks. By analysing the uploaded datasets and evaluating various quality metrics such as data completeness, the presence of columns with static values, and the existence of duplicate records, a health metric can be derived and presented to the users, offering a more comprehensive understanding of the provided datasets.

In a subsequent phase, once the datasets have been bought, it could be possible to gather information about their usage from the buyers. Additionally, information about the usage of datasets derived from the original ones can be obtained, thanks to the availability of lineage tracking services. By analysing this dataset usage information, valuable insights can be derived. These insights may include details about the machine learning applications in which the datasets have been utilized, the specific stakeholders who express interest in particular types of datasets, the datasets typically combined with them when they are purchased, and the pricing details for both the initial datasets and the derived ones. These insights can then be presented through statistical graphs and charts in user-friendly dashboards of the "Data Usage and Intentions Analytics" service. For example, these dashboards could report the frequency each feature in the dataset has been used as a target variable in a machine learning task.

The identification of similar datasets plays an important role in the extraction of the usage and intentions analytics, but also in the entire process of data valuation. Some approaches that could be of use towards this end are:

- Keyword and metadata analysis: utilize search engines like Elasticsearch and filters that allow end users to efficiently search for similar datasets.
- Clustering methods: use clustering algorithms (e.g., K-means) to group similar datasets together using all their available metadata.
- Fuzzy matching: find similar datasets by measuring the distance of datasets using their text descriptions.
- Collaborative filtering: Find similar datasets using the purchase patterns of other data consumers using recommender systems algorithms.

- Graph-based methods: represent datasets as nodes and their relationship (e.g., same data provider) using edges. Then use this graph to find similar datasets by analysing this structure.
- Adversarial validation: Find similar datasets by examining their performance on binary classifier. If it is able to differentiate between two datasets, then there is a dissimilarity.

The information that can be used for assessing dataset similarity includes metadata provided by the system (such as tags, descriptions, and data model information), as well as the questionnaires completed by dataset owners and buyers. Additionally, data obtained from the purchase history (including pricing, derived datasets, and usage information) and various profiling metrics (such as quality metrics) are also valuable for this task.

Usage intentions are an important part of the context definition during data valuation. Several of the methodologies for data valuation described in Section 3.2 provide the possibility to declare the intended use for the dataset being valuated.

For example, one of the “activity” nodes in **Data Value Networks (DVN)** – an alternative to data value chains, derived from the properties of data as an asset – is “data exploitation”, which allow a data stakeholder to define potential exploitation actions: question answering, decision-making, visualisation, service creation.

One of the six components of the **Data Value Map (DVM)** – a framework for promoting shared understanding between stakeholders involved in data initiatives – requires users to reflect about and declare the business value (in terms of benefits and behaviour) that may result from using a data asset. Similarly, the **Business Data Value Canvas (BDVC)** requires its users to describe the business value associated to each of its own data value chain (preparation, analytics, interaction, effectuation, intelligence).

The **Data Valuation Process (DVP)** by EUT allows the user to declare both the business-related benefits of utilising a dataset, as well as its desired application to developing technical solutions (visualisations, analytics, ML models). In the case of developing ML models, the process requires the user to select the model that they intend to train with the dataset and the type of metrics to use for its evaluation.

Survey-based approaches, such as the **mapping of properties to data value** or **data sheets for data sets** also require some form of usage intention declaration from the part of the user.

3.3.3 Technologies

User surveys

Declaring the intentions with respect to a data asset is a prerequisite for any type of further analysis. The preferred method for this seems to be through **user surveys**, during which clear questions are asked about the goal of acquiring and processing a data asset. Such an approach has already been used in the tools developed based on the methods presented in the previous sub-section: the Data Valuation Component, the DVM, DBVC, and data sheets for data sets.

DaVe[39]

Storing this information in a semantic format that facilitates exchange between different components can be achieved using **DaVe**, the vocabulary for data value, which implements

[39] https://lov.linkeddata.es/dataset/lov/vocabs/dave

the logic of Data Value Networks presented in the previous sub-section, including activity nodes concerning potential exploitation of data assets.

Pandas
Widely used python library for data science, data analysis and machine learning applications. It provides an easy way to analyse and visualise data of multi-dimensional arrays.
Scikit-learn
A simple python library that provides tools for predictive analytics and machine learning models. In our case, it can be used to provide the clustering algorithms required for similarity identification.
Tensorflow
A free and open-source software library for machine learning and artificial intelligence mainly focusing on training and inference of deep neural networks. It can provide methods for graph-based similarity models.
Elasticsearch
Can utilize description and metadata to optimally retrieve similar assets based on string matching algorithms. It provides filters that allow the uses to search through big volumes of datasets.
TheFuzz[40]
A string-matching library that uses Levenshtein distance to calculate the differences between sequences in a simple-to-use package. It can be used to find similar assets by comparing the provided descriptions.
LibRecommender[41]
An easy-to-use library that let users quickly train and deploy different kinds of recommendation models. It can be used to provide recommendation to users based on purchase patterns of other consumers.

3.4 DATA MONETISATION MODELS REPO

3.4.1 Overview

The Data Monetisation Models repository is a comprehensive resource placed within the Monetisation XAI Engine, providing to data owners an collection of available data monetisation models. These models will be continuously re-trained and optimised, towards powering the PISTIS Marker Insights (*see section 3.1*) module that will be used to reveal the trends and the dynamics of the overall data market.

Overall, the Data Monetisation Models repository will serve as the knowledge base for data valuation and dynamic pricing, empowering data owners who seek through the PISTIS Data Monetisation services, to leverage their data assets and generate revenue streams by effectively monetising their data.

[40] https://github.com/seatgeek/thefuzz
[41] https://github.com/massquantity/LibRecommender

3.4.2 Methods

Given the high-level description of Data Monetisation Models repository, possible methods that can be utilised to effectively support model management, discovery, integration, and collaboration, thus enabling users to leverage and benefit from the available models, can include:

- Model storage and organisation methods to store and organise data models within the repository; this can involve creating a structured directory or database system to categorise and store models based on different criteria such as model type, domain, or application.
- Version control techniques, to manage different versions of the data models; allowing for easy tracking, retrieval, and comparison of model versions.
- Metadata management, to store descriptive information (e.g., name, description, author, creation date, performance metrics, dependencies, licensing information, etc.) about each data model. This metadata facilitates efficient search, discovery, and understanding of the models in the repository.
- Model documentation for each data model to provide insights into its purpose, inputs, outputs, usage guidelines, and any associated documentation/ research papers, ensuring that users have clear instructions on how to use and interpret the models effectively.
- Model validation and evaluation methods to validate and evaluate the performance and reliability of the data models. This can involve using evaluation metrics, test datasets, and validation techniques specific to the domain or problem the models address, towards ensuring that the models in the repository are accurate and fit for purpose.
- Model search and retrieval methods to enable search and retrieval of models based on specific criteria (e.g., keyword-based search, filtering options based on model attributes, etc.) to enable users to find relevant models efficiently.
- Model integration and deployment methods, such as the provision of APIs to facilitate seamless integration and deployment of the data models into target systems or applications. Ensuring compatibility with different programming languages and environments can enhance usability and adoption.
- Continuous updates and maintenance of the repository by incorporating new models, addressing issues or bugs, and enhancing existing models based on user feedback This ensures that the repository remains up-to-date and valuable to its users.

3.4.3 Technologies

Taking into consideration the possible methods for the operation of the Data Monetisation Models repository; a non-exhausting list of commonly used technologies and tools that can be used for the design and implementation of a component functioning as a repository of Data Monetisation models is provided below. However, it shall be noted that the final selection of technologies and tools will depend on various factors such as the specific requirements, the scale of the repository, etc.

Programming Languages

Programming Languages, such as Python, Java, are frequently used for implementing repository components. Python is popular for its rich ecosystem of ML libraries, while Java provides robustness and scalability.

Version control systems
Version control systems, like Github are essential for managing the source code and versioning of the repository component. They allow for easy collaboration, tracking changes, and managing different versions of the codebase.

Database Management Systems
Depending on the requirements, different database management systems (DBMS) can be used to store and manage the data models. Relational databases like MySQL or PostgreSQL, NoSQL databases like MongoDB, can be employed based on the specific needs of the repository.

API Development and Documentation
Tools like Swagger or Postman can be utilised for designing, documenting, and testing APIs exposed by the repository component. These tools simplify the process of API development and enable easy integration with other systems.

4 DATA EXCHANGE PREPARATION

The Data Exchange bundle resides in every PISTIS data space factory, meaning inside the organizational boundaries of every data supplier and consumer. This bundle encompasses various services associated with configuring monetization schemes for specific data assets, as well as executing and monitoring sharing and trading contracts. More specifically, three major components are comprising the data exchange bundle, namely as trading pre-processor, smart contract execution engine and lastly the data space factory cleaning. The trading pre-processor component allows data owners/suppliers to determine the of a data asset before publishing it on the data market exchange. It takes input from the data quality assessment module and the monetization XAI engine, while provides output to the data value contract composer and security – trust component. A smart contract execution engine is responsible for processing transactions recorded in the blockchain ledger, making the designated data artefacts available to the demand side, and conducting contract observability operations to check the status of relevant contracts for trading and sharing activities. Lastly, the data space factory clearing house component assigns the transaction values to the digital wallets of each party. These wallets can be either custodial or non-custodial depending on the user preference.

4.1 DATASPACE STABLECOIN CUSTODIAL WALLETS

4.1.1 Overview

A wallet[42] is a tool that gives access to the digital tokens, by the owner, allowing him to send and receive tokens. These tokens are not physically stored inside the wallet but rather exist on the blockchain network. The only thing that is stored inside a wallet is users' secret key (private key). A private key is a randomly generated, secret key known only to the owner. It is used to sign transactions and prove ownership of digital assets. The private key must be kept secure and should never be shared with others.

There are two kinds of wallets named as software and hardware, which are then further divided into custodial and non-custodial.

- Software Wallets [57]: A software wallet is a computer program that is executed in the operating system of an electronic devices such as personal computers, smartphones, and tablets. As software it comes in different forms like Desktop app, Mobile app or as an extension of the web browser. A major drawback is the fact that software wallets are part of the operating system, so security levels are as good as the entire operating system. Consecutively the operating system should be always updated and free of malicious software, in avoidance of unwanted access to private key and the tokens.
- Hardware Wallets [49]: This type of wallet is a hardware device that is constructed to store the users' private key and connects with a computer usually via USB. They provide an extra layer of security compared to software wallets because they operate in an isolated, offline environment, making them resistant to many online threats such

[42] https://wiki.iota.org/use/wallets/what-is-a-wallet/

as hacking and malwares. In order to interact with the blockchain the hardware wallet establishes a secure connection with a software wallet which then connects with blockchain network.

Custodial wallets[43] are offered by third-party providers, such as CoinBase[44] and Kraken[45]. These providers hold the private keys and control the funds on behalf of the user, maintaining full responsibility in case of security issues. Users typically create accounts with the custodial service, which provides them with a wallet address and access to their funds through a web interface or a mobile application.

Stablecoins[46] are a type of digital asset designed to maintain a stable exchange value with a real-world asset like traditional fiat currencies (such as Euro) or commodities like gold. Even though they offer the same functionalities as any other cryptocurrency (Bitcoin, Ethereum etc), they also mitigate the high volatility issue. The main types of stablecoins are four namely as fiat-collateralized, crypto-collateralized, algorithmically managed and commodity-collateralized, differentiating mainly on the stability mechanism which keeps the exchange ratio 1:1. Further details on stability mechanisms are mentioned in section 5.3.

4.1.2 Methods

The main methods applied for the operation of a custodial cryptocurrency wallet are the following:

- *Private Key Management*: For the operation of custodial wallet robust security measures to protect users' private key should be applied. The private keys are usually stored in encrypted databases to prevent unauthorized access.
- *Cold Storage*: To protect against hacking and unauthorized access, custodial wallet owners often employ cold storage solutions. Users' funds are stored offline in cold wallets, reducing the risk of exposure to online threats.
- *Multi-factor Authentication (MFA)*: Custodial wallets often require from users to enable multi-factor authentication so to add an extra layer of security, involving a mobile device during the process of registration and login.
- *Withdrawal Controls*: In order to prevent unauthorized fund transfers withdrawal limits and approval processes for large transactions are usually applied, adding an extra level of security.
- *Customer Support*: Customer support to address any issues, inquiries, or problems users may encounter.

4.1.3 Technologies

A non-exhausting list of technologies and tools that can potentially be used to support the aforementioned methods, is provided below:

[43] https://crypto.com/university/custodial-vs-non-custodial-wallets
[44] https://www.coinbase.com/
[45] https://www.kraken.com/
[46] https://www.investopedia.com/terms/s/stablecoin.asp

Stronghold

IOTA Stronghold[47] is an open-source highly secure software library designed to provide protection for user's sensitive data and cryptographic keys. It was developed by the IOTA Foundation, leveraging advanced encryption techniques and principles to safeguard information against potential threats, including unauthorized access. Stronghold employs versioned, file-oriented backups (designated by the ".stronghold" file extension) featuring dual encryption, facilitating effortless backup and secure sharing across multiple devices. It is used for decentralized and distributed architectures, specifically tailored to address the security challenges faced by distributed ledger technologies, such as the IOTA Tangle. Lastly, since Stronghold is a software library and not an executable binary it can be utilized within the code which might also holds a different token than IOTA.

IOTA SDK library

The IOTA SDK[48] is a collection of tools, libraries, and APIs that enable developers to interact with nodes in the Shimmer Testnet[49] and IOTA networks running the Stardust protocol. It provides functions and methods to perform essential tasks, including generating IOTA addresses, creating and signing transactions, attaching transactions to the Tangle, reading Tangle data, and querying the network for specific information. It is comprised of two main modules named as client and wallet. The client module provides a set of low-level functions that grant precise command over interactions with Shimmer nodes. The wallet module delivers a suite of elevated functions, encompassing account management, address generation, transaction creation, and seamless interaction with the Shimmer network. This module operates in a stateful manner and offers the flexibility to optionally connect with IOTA Stronghold, allowing for secure seed handling, storage, and state backup.

4.2 DataSpace Users StableCoin non-Custodian Wallets

4.2.1 Overview

Non-custodial wallets are wallets where users have complete control over their private keys and consecutively their funds. Users generate their own private keys, which are stored locally on their devices or in an offline hardware wallet. On the contrary in custodial wallets a third-party service provider holds the private keys and controls the funds on behalf of the user. Custodial wallet providers are also offering a web interface where the user can handle his address and funds.

4.2.2 Methods

The main methods applied for the operation of a non-custodial cryptocurrency wallet are the following:

- *Private Key Management:* One of the most crucial functionally that should be supported is the storage of the private key so the user can sign transactions and also the storage of the public key so the user can receive payments.

[47] https://wiki.iota.org/stronghold.rs/welcome/
[48] https://github.com/iotaledger/iota-sdk
[49] https://explorer.shimmer.network/testnet

- *Multi-Signature (Multi-sig) Wallets:* Multi-signatures should be offered for types of transactions where more than one parties of the same organization should sign before the execution. This adds an extra layer of security and reduces the risk of a single point of failure.
- *Hardware Wallets:* The non-custodial wallet should offer the possibility to the owner to store the private key not on the local disk but in an external physical device, such as USB disk. This way an extra layer of security is provided by keeping the private key isolated from online malicious attacks.
- *Mnemonic Phrases:* This kind of wallet should provide during the set up a seed phrases for recovery situation when the private key has been lost. This seed is a sequence of words that can be used to restore a non-custodial wallet and should be stored offline (note on a paper) for security reasons.

4.2.3 Technologies

In general, digital wallets are built using programming languages such as Python, Go and Rust. The selection of the programming language depends on the platform the user demands, which can be an extension of the browser, a desktop or mobile application. Concerning the connection and interaction of wallet with the blockchain, WebSockets and RESTful APIs are commonly used for this purpose.

OpenDSU (Decentralized Software Unit)

OpenDSU [62] can be utilized for the implementation of PISTIS wallet and the key management. Is an implementation of the DSU technology which applies in decentralized applications (dApps), digital identity systems and secure data sharing. More specifically, OpenDSU is an open-source technology that provides a decentralized way for data management, storage, and access, based on the Distributed Ledger Technology (DLT).

Firefly software wallet

The Firefly[50] is a software wallet designed to provide to users a dashboard for managing their digital assets on the IOTA ecosystem. Two different options are offered named Firefly IOTA and Firefly Shimmer, where Shimmer is the official validation network of IOTA (testnet). This wallet has been built using Rust programming language and it is secured by Stronghold for address generation and transactions.

MetaMask software wallet

MetaMask[51] is a well-known non-custodial software wallet which acts as a browser extension. It enables users to interact with Ethereum blockchain and several Dapps, but not with IOTA mainnet or testnet. One of its key functionalities is the possibility to create multiple accounts where each one has a unique address and private key. MetaMask works as an extension of popular web browsers like Chrome and Firefox. The private keys are stored locally in users' browser, thus making essential privacy issue to ensure that the browser and operating system are secure.

[50] https://firefly.iota.org/
[51] https://metamask.io/

Ledger Nano hardware wallet

Ledger Nano[52] is a hardware device that connects to your computer via USB which stores the recovery phrase and the private keys, adding an extra level of security. It works cooperatively with the Firefly software wallet to securely access your funds.

4.3 Smart Contract Execution Component

4.3.1 Overview

Smart Contract Execution is a core component of PISTIS. It executes all the stored and valid data trading chaincode and makes the data artefacts available to the demand side. This component is responsible for the initiation of authorization and observability operations (e.g., triggering the Transaction Auditor) prior to any contract execution in order to check the status of trading/sharing actions in a secure and proper manner. Smart Contract Execution component is a robust, secure, and highly performant component for building and executing decentralized applications (dApps) and smart contracts.

4.3.2 Methods

Smart Contract Execution is tightly coupled with the underline DLT technology, while the stored chaincode will be checked and verified before the actual execution.

4.3.3 Technologies

This component is relevant to the underline DLT technology and the supported smart contracts. More details will be provided in sections 6.2 and 6.3.

4.4 Trading Preprocessor

4.4.1 Overview

The Trading Preprocessor component is a semi-automated service that provides the data owner/supplier with the option to decide on the trading value of a data asset prior to publishing it to the Data Market Exchange. To do so, the data owner/supplier needs to verify and evaluate the output of the Data Quality Assessment and the Monetisation XAI Engine module and assign a value to the data (the suggested or not). This component is necessary for the preparation activities for a contract.

4.4.2 Methods

A smart contract will assist the owner/supplier to assign a value considering PISTIS suggestions to the data. The information regarding the sharing profile and the trading value from data owner/supplier before publishing the data to the Data Market Exchange will be stored in the smart contract.

4.4.3 Technologies

This component is also relevant to the underline DLT technology and the supported smart contracts. More details will be provided in sections 6.2 and 6.3.

[52] https://wiki.iota.org/use/wallets/firefly/user-guide-ledger/

5 PISTIS DATA EXCHANGE MARKET

The PISTIS Data Exchange Market bundle serves as the primary financial infrastructure for trading data within the PISTIS ecosystem. In PISTIS, a conceptual Stablecoin is used as an additional payment method to connect assets with real monetary values. This approach helps to avoid the volatility of cryptocurrencies, which can make trading challenging and speculative. To convert real monetary values into this system, a PISTIS Stablecoin Exchange Desk service is utilized. Funds are inserted into the platform through an Open Banking Interface, which deposits them into a FIAT Wallet owned by the platform. The Stablecoins are then distributed to the PISTIS Stablecoin Custodial Wallets, and stakeholders have the option to either keep them there (entrusting them to PISTIS) or transfer them to their own wallets. The PISTIS Market Exchange Facility oversees market activity, providing insights and facilitating transfers of stablecoins between wallets.

5.1 PISTIS MARKET EXCHANGE FACILITY

5.1.1 Overview

This component is meant to focus on the monetary transactions happening between different wallets and expose those to other relevant PISTIS components like the “Market Insights”. This exposure to other components will facilitate the services like prices forecasting and market insights.

5.1.2 Methods

The solely method that should be met is the seamlessly respond to other components about which parties and when proceed into a data and consecutively into a value transaction More specifically, this component should accept wallets addresses, amount of value exchange, and perhaps the type of the dataset that was purchased.

5.1.3 Technologies

The aforementioned method could be implemented by designing a Web API (see section 5.4.3)

5.2 PISTIS PLATFORM STABLECOIN CUSTODIAL WALLET

5.2.1 Overview

This component is considered to be hosted on PISTIS Cloud platform and will serve the needs of the platform’s owner in terms of managing his digital assets. Given that every transaction will generate a commission fee to be kept by PISTIS, there is a need for a wallet and consecutively an address for the value to be transferred.

5.2.2 Methods

The methods comprising a custodial wallet were described in section 4.1.

5.2.3 Technologies

Technologies and software stacks used for the implementation of a custodial wallet was described in section 4.1.

5.3 PISTIS Platform StableCoin Exchange Desk

5.3.1 Overview

This component, in general, is responsible for accepting FIAT currency and transforming it to Stablecoin and conversely, establishing the link to real monetary values (euros). PISTIS stablecoin will be used as a payment method to link assets to real monetary values and at the same time avoiding the volatility of cryptocurrencies which makes trading difficult and speculative. The user will be charged a small fee for every transaction depending on the amount of money wanting to convert. It should be noted that the Exchange Desk will not offer trading options nor exchanges to a different crypto rather than PISTIS stablecoin.
At the technical level this component will be hosted on PISTIS cloud, interacting with the user with a UI.

5.3.2 Methods

Liquidity Provision: The exchange ensures sufficient liquidity in its stablecoin trading pairs, enabling smooth transactions and minimal price slippage.

Deposits and Withdrawals: Users can deposit and withdraw stablecoins to and from their exchange accounts. Users should be able to deposit euros through various methods (bank transfer, credit/debit card)

Transaction Fees: Define and implement a transparent fee structure for trading and withdrawals. Stablecoin exchanges may charge trading fees for executed trades, deposit and withdrawal fees, and other service-related fees. Transaction fees can be charged as a percentage of the trade value or a flat fee.

Stablecoin Price Oracle: Integration of price oracles for receiving real time exchange ratios and current prices.

Types of stablecoins – PISTIS stablecoin Euro-pegged 1:1

<u>Algorithmic stablecoins Seigniorage-Style</u>

The defining characteristic of a pure algo-seigniorage coin is the absence of collateral. The mechanism design comprises the case wherein if the demand increases, the issuer will create and sell additional stablecoins to maintain the peg to the Euro. If the demand decreases, the issuer issues a second asset, typically a bond, and sells it against the stablecoin to reduce the aggregate supply. The bond is a promise to future stablecoins. Algorithmic stablecoins attempt to stabilise a decreasing price by introducing bonds and/or stocks as a “institutional” guarantee to increase the stablecoin supply in the future.

<u>Asset-backed Collateralised stablecoins: On-chain</u>

The most well-known (successful) on-chain collateralised stablecoin is DAI stablecoin, which is pegged to the US dollar. DAI is based on a set of smart contracts on the Ethereum blockchain. To mint new ones, a user must send Ether “funds” to one of the smart contracts, which are sequentially used as collateral for a loan in DAI. The interest rate is known as the stability fee. The volatility of such stablecoin is low, which is considered a major advantage.

<u>Asset-backed Collateralised Stablecoins: Off-chain</u>

A collateralised stablecoin as e.g., Tether is associated with an exchange platform (Bitfinex) and the issuing company (Tether). The off-chain collateral could generate lack of transparency and censorship resistance. For such a stablecoin to preserve the ratio of 1:1 versus a fiat currency, there should be a stability mechanism on the backstage.
In the framework of PISTIS project no reserve asset will be provided to back the stablecoin, named as fiat-collateral mechanism. Instead, the liquidity mechanism will be based on a crypto-collagenized method, which is described in the following section 5.3.3.

Stablecoin risks

A major problem with many types of stablecoins, is that price stability requires collateral of at least 100%. For that reason, in PISTIS we will discard the idea that stability can be attained solely through an "algorithmic economy".

Moreover, on-chain collateral has many benefits over off-chain collateralised crypto-assets. On-chain promotes instant transparency - as demonstrated by the DAI stablecoin. In particular, each user (via its wallet) can verify that the collateral is effectively present in real-time. On the contrary, off-chain collateralised stablecoins potentially render a single point of attack and the threat of a sudden closure by malicious agents or even institutional regulatory bodies, due also to lack of transparency and very high volatility caused by speculators and arbitrageurs. Also, the creation and operability of those stablecoins, entails severe costs (e.g. audits and banking relation) and is often issued with a profit motive. This results in a permanent desire to engage in ***fractional reserve banking*** as in the traditional economic legacy systems. Via a decentralised on-chain collateral solution à la DAI, and when the collateral is on-chain, the aforementioned threats are non-existent.

5.3.3 Technologies

Viviswap API

This API[53] could potentially be used for this PISTIS component to provide an exchange mechanism between fiat money and cryptocurrencies. The Viviswap API can convert a certain amount of euros to IOTA tokens, simply by knowing your IBAN account number and your IOTA account address. Though the PSD2 API (see section 5.5.3) confirms that the IBAN you provided has enough balance in euros. If confirmation is successful, then it sent the amount of euros, indicates by the user, to Viviswap bank account. The Viviswap is then responsible to convert the euros to IOTA tokens using an exchange rate and commissioning a service fee. Lastly, the IOTA tokens are sent to users'crypto wallet.

At this point it was considered appropriate to describe the existed stability mechanisms of stablecoins. The two major categories of Stablecoins based on the underneath liquidity mech are named as Algorithmic Stablecoins, On-Chain and Off-Chain Collateralized Stablecoins [10].

- Algorithmic Stablecoins: The main characteristic of this type is the absence of collateral. In case of rising demand for the stablecoin, the issuer creates and sells additional stablecoins in order to maintain the peg to the fiat money that has been

[53] https://api-service.viviswap.com/docs/#section/Introduction/API-spaces

chosen. In instances decreasing demand the issuer introduces a secondary asset and sells it against the stablecoin in order to reduce the available quantity. This secondary asset is usually a form of bond which functions as a commitment to a forthcoming increase of stablecoin supply in the future [13].

- On-Chain Collateralized Stablecoins: This type can also be found as Crypto-Collateralized Stablecoins. It maintains its value by being backed by a reserve of other crypto assets. For reasons that have to do with the high correlation between cryptocurrencies, this type of stablecoin is backed by collateralization ratios typically above 150%.
- Off-Chain Collateralized Stablecoins: This type can also be found as Commodity-Collateralized Stablecoins. The stability is derived from a reserve of tangible, real-world assets, typically commodities like precious metals or fiat money. The issuer holds a reserve of the chosen commodities of an equivalent amount of stablecoins in circulation.

For PISTIS project it is believed that the most appropriate type of stablecoin is an on-chain collateralized one. Such a coin is the XSD [54] which is a fully on-chain decentralized Collateralized Debt Position (CDP), hard-pegged to the US Dollar, working on Ethereum and IOTA-Shimmer network.

5.4 PISTIS Platform FIAT Wallet

5.4.1 Overview

A FIAT wallet serves as a digital container for securely storing fiat currencies, such as coins, banknotes and credit cards. The primary functions of a FIAT wallet include money management, safekeeping of cash while facilitating everyday financial transactions. As a technology enables payments from mobile devices and further enhances convenience, security and efficiency. There are a number of digital FIAT wallets from various financial institutions and digital payment providers, including PayPal, Apple Pay, Google Pay and Visa wallet.

5.4.2 Methods

Overall, this component should either be a standalone software binary or a library that is responsible for the following operations.

1. Communicates with the Open Bank Interface (see section 5.5.1)
2. Provides a REST API for other PISTIS components.
3. Stores as less as possible information needed for each operation.

Some of its key functionalities are:

a) Add a new bank account.
    i. Reception of the PISTIS account ID.
    ii. Reception of user’s bank account credentials.

[54] https://www.lendexe.fi/learn/whitepaper/#toc_Networks

iii. Reception of user's IBAN number.
iv. Storage of PISTIS fiat account ID and/or IBAN, credentials to its software vault.

b) Deletion of an account.
   i. Reception of account ID.
   ii. Deletion of user's related info from vault.
c) Request of PISTIS stablecoin (cash in)
   i. Reception of account ID, IBAN, amount of Euros and the DLT address.
   ii. Checks if the requested amount of PISTIS coins is greater of equal the amount of Euros in fiat wallet.
   iii. In case the bank balance in Euros is enough then a bank transaction towards the PISTIS platform IBAN is performed, including also the user's DLT address as metadata.
d) Request of Euros (cash out)
   i. Reception of relevant notification to the main application, that a user requested a cash out transaction.
   ii. The main app waits for a notification from the open banking API that the transaction has been completed.
   iii. Send as notification to the main application the requested amount of money.

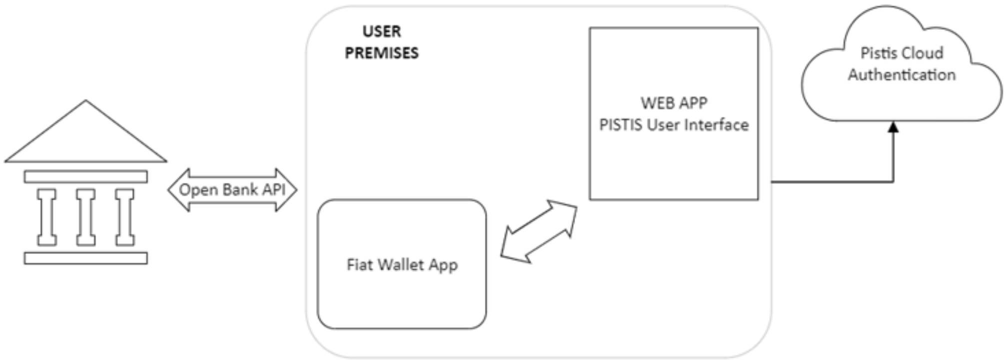

Figure 5-1: Diagram between fiat wallet and main app

5.4.3 Technologies

This component is expected to be implemented by using a modern compiled programming language, such as Rust[55] or Go[56]. The following paragraphs, of this section, summarize the protentional usage of these programming language within this PISTIS component.

Rust programming language

Rust has a growing ecosystem of cryptographic libraries that enable developers to implement secure cryptographic operations required for blockchain applications. These libraries support functions like key generation, digital signatures, and hashing, which are foundational to cryptocurrency protocols [41].

Go programming language

[55] https://www.rust-lang.org/
[56] https://go.dev/

Several blockchain projects, including some popular cryptocurrencies, have been built using Go. One of the most notable blockchain projects developed in Go is Ethereum[57]. Go's standard library and ecosystem of third-party packages make it well-suited for building tools and utilities related to blockchain explorers, transaction parsers and price tracking applications.

REST API

The Rest APIs[58] is a widely adopted architectural style for establishing interaction with a computer system in order to retrieve information of perform a function, over the internet, providing a set of principles and constrains. For an organization offers a way to share resources and information without compromising security issues. In general resources are represented as URLs and different HTTP methods are used to perform actions over the resources. Rest APIs are scalable and platform-independent making them ideal for IoT, mobile apps and serverless computing. For the PISTIS FIAT wallet this technology would be useful for initiating communication between the digital wallet, the Bank, the UI and the backend.

WebSockets

While the aforementioned API facilities mechanisms for data fetching and interaction between different applications, WebSockets [22] are mainly designed for real-time, continuous communication. WebSockets are a communication protocol that provides real-time bidirectional communication channels between a client and a server, over a single TCP connection. The key characteristic of this protocol is the long-lived communication channel establish between client and server without the need to initiate a new request for each interaction.

5.5 OpenBank & AMM/DEX Service Architecture

5.5.1 Overview

The platform to host the cryptocurrencies, NFTs and wallet new protocols, will involve also the development of an AMM/DEX (or simpler functional architecture of an OpenBank) which will provision e.g., fundraising, lending, DEXs, yield farming, stablecoin yield aggregators, derivative and synthetic product protocols for NFTs and PISTIS stablecoin.

5.5.2 Methods

The various methods applied for the operation of this module are the following:

- Account Information Services (AIS): Allowing third-party providers to access account information, including details about account balances, transaction history and other relevant data.
- Payment Initiation Services (PIS): Allowing third-party providers to initiate payments on behalf of the users without the need to log in to separate platforms.

[57] https://ethereum.org/en/developers/docs/programming-languages/golang/
[58] https://www.redhat.com/en/topics/api/what-is-a-rest-api

- Security and Privacy: Robust security measures to protect customers' sensitive financial data, including data encryption, strong customer authentication (SCA)[59] and compliance to regulations like GDPR.

High level architecture design

The Open Banking concept entails the access of third-party payment services to access customers' baking information's such as transactions, payment history and current balance, using APIs.

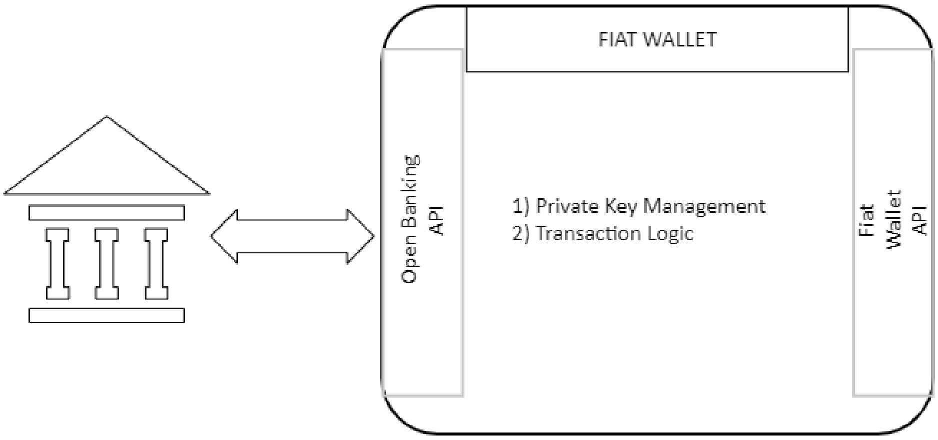

Figure 5-2: High level architecture design of OpenBank Interface service

Low level (advanced) architecture design

The "**Bank**" design could also incorporate ***a trading Automated Market Maker / Data Exchange operability (AMM/DEX)***, as commonly incorporated via the traditional account brokerage financial system operating in modern economies. Specifically, it could entail the provision of capabilities such as e.g., trading, investing, fundraising, lending, DEXs, yield farming, yield aggregators, derivative products, bonds, stocks and other crypto-asset protocols as well as tradable and (non)sharable (s)NFTs and PISTIS stablecoins.

Technically, a cross-chain oracle/aggregator could be developed via various platforms e.g., utilizing solidity and web3modal in the form of an xy=k AMM curve or using Web3 OpenDAX (Openware) generating Apps or a "yield farming" type platform which will facilitate conversion to SC on BNB Smart Chain/ETH under the current prototypes of Uniswap, Curve, PancakeSwap etc. Importantly, the topology will address both *technical and economic risks* i.e., multi-transaction attacks, ordering attacks, miner extractable value, oracle manipulation etc.

5.5.3 Technologies

PSD2 Open Banking API

The PSD2 Open Banking API, offered by the European Depositary Bank, is a financial solution that follows the principles of the Second Payment Services Directive (PSD2)[60] regulation. In

[59] https://ec.europa.eu/commission/presscorner/detail/en/MEMO_17_4961
[60] https://www.europeandepositarybank.com/psd2-open-banking-api/

Europe, due to PSD2, banks are required to develop their own Open Banking APIs and provide them for free. The intent of PSD2 regulation is to encourage Banks to expose their accounts and payments APIs to AISP (Account Information Service Provider) and PISP (Payment Initiation Service Provider)[61]. Open Banking API architecture provides secure communication between different computer systems, only with permission by the user who authorises a third-party app to access his banking data. The aforementioned API empowers users with real-time access to their account information, transaction history, and balances, while also facilitating seamless payment initiation and fund transfers. In the concept of PISTIS project the most important functionalities of PSD2 Open Banking API are the Account Information Services (AIS) and the Payment Initiation Services (PIS). The first one allows third-party providers to access information regarding account balances, transaction history and other relevant account information. The second one enables third-party providers to initiate payments on behalf of the users, without the need from the user to log in to separate platform.

OAuth 2.0 and OpenID Connect

These are authentication and authorization protocols commonly used in Open Banking implementations. OAuth 2.0 is an authorization framework that allows applications to obtain limited access to user's resources on another system without exposing the user's credentials[62]. OpenID Connect is built on top of OAuth 2.0 providing a standardize way to authenticate users and verify their identities[63].

Thegraph.com and Web3 OpenDAX (Openware) generating Apps

To ensure scalability and efficiency, the platform might also utilize a blockchain indexing service such as deployed on **thegraph.com**. Moreover, a selection will be decided between **Web3modal** in the form of an xy=k AMM curve, or **Web3 OpenDAX (Openware)** generating Apps including "yield farming" functionalities. All of the above, facilitate conversion to ETH, IOTA, SC, BNB Smart Chain etc., similarly to some already successful operating paradigms and prototypes based on Curve (Uniswap, PancakeSwap etc). A draft architecture is depicted in the following skeleton.

[61] https://ieeexplore.ieee.org/abstract/document/8668107
[62] https://oauth.net/2/
[63] https://openid.net/developers/how-connect-works/

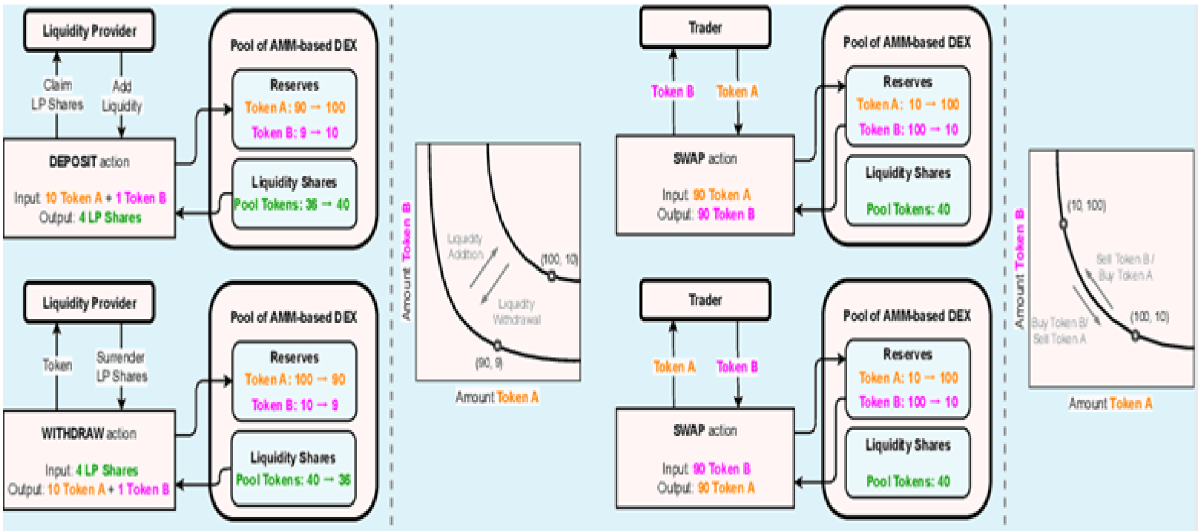

Figure 5-3: Tentative architecture of OpenBank with thegraph.com

DEX/OpenBank Technical and Economic Risk Mitigation

Reentrancy: The OpenBank/DEX could cope with a contract which is potentially vulnerable to a re-entrancy attack if it delegates control to an untrusted party, by calling it with a large enough gas limit, in particular during a withdraw-funds function from the OpenBank that checks for the internal wallet balance of a user, and sends money to update the user's wallet balance. If the receiver is a smart contract, it can then repeatedly re-enter the OpenBank contract of the owner's account, and drain the wallet funds (stablecoin and NFT values).

Single transaction sandwich attacks: If an attacker attempts to manipulate an instantaneous our DEX/AMM price both via the front- and back-running of another user's wallet transaction, the imbalance would exploit composable contracts which rely on the manipulated price, reverse the imbalance to cancel out the cost of the first step.

Multi-transaction sandwich attacks: Usually, the deterministic price on an AMM is altered prior to and after some other target transaction has been executed, in order to profit from temporary imbalances in the user's OpenBank liquidity reserve of its wallet account (stable coin and NFTs). This is the reason why the OpenBank module, obtains a DEX form: the instantaneous AMM price is simply a ratio of AMM reserves and imbalances can be created simply by changing this ratio. Yet, beyond a simple Bank functionality, the DEX will not allow for "swaps" creating imbalances, which, if left unbalanced, will incentivize arbitrageurs to perform the reverse actions to balance the entire OpenBank pool, and profit from slippages and smart contract bugs, to liquidate the entire Bank out of all users' wallet funds.

Anonymity and Privacy: The DEX/AMM operability promotes pseudo-anonymity. This means that if an agent's real-world identity can be linked to an on-chain address, all the actions undertaken by the agent through that address are observable.

Overcollateralization: As opposed to a trustless system without strong identities or legal recourse, with overcollateralization creating the economic incentive for a monetised deposit swap of assets, or pre-purchase or loan (e.g., NFT via an LLC structure, or a stablecoin transaction) in PISTIS we insure the user against losses. For example, using on-chain tokens, as prices evolve over time, volatility could allow automated deleveraging: if an agent's level of collateralization falls below a protocol-defined threshold, an arbitrager in the system can reduce the user's liquidity or value of the account in the OpenBank and create exposure.

Miner Extractable Value: There are consensus (Ledger blockchain protocol) risks which could render the MEV exceed the block reward. The OpenBank can be left uninsured when double spending of stablecoins or their net worth of an NFT, in base layer can create malicious incentives.

Market Manipulation: While assuming the oracle of data origin (off- or on-chain) follows a best practice implementation and is non-malicious, yet the manipulation of the market evaluation of an asset in a user's wallet over a certain time period can offer an opportunity to malicious attackers to profit. Such an attack could trigger liquidations. An example using PISTIS stablecoin could entail thin liquidity for the 1:1 Euro-pegged PISTIS coin, traded at a temporary price of €1:10 over some minutes on a major centralized crypto-asset exchange or Oracle, before returning to its intended €1 value. As a result, the OpenBank would report a value of the stablecoin or NFT in a user's account, diverging the actual 1:1 peg, which will create manipulation opportunities.

Oracle Manipulation: Centralized oracles serve as a single point of failure and are vulnerable to the provider behaving maliciously, whilst decentralized oracles may use on-chain data, most notably on DEXs (specifically AMMs) for crypto-to-crypto Bank transactions. We will utilise mostly decentralized oracles (such as Chainlink[64]), hence we mitigate this problem by aggregating data feeds from multiple sources (e.g., by calculating the median) and relying on curating reliable sources.

Composability: Stablecoins (and NFTs) can be repeatedly tokenized and interchanged between users, via rehypothecation. An example is the use of flash loans for manipulating instantaneous DEX/OpenBank valuations for complex synthetix products of specific data and metadata structures, not corresponding to the price evaluation of users' wallets.

[64] https://chain.link/

6 DATA EXCHANGE GOVERNANCE

Data Exchange Governance is a core component of PISTIS Market Exchange Platform for data exchanges and includes all the services that are necessary for the operation of the DLT network to operate within PISTIS.

6.1 SMART CONTRACT TEMPLATE COMPOSER

6.1.1 Overview

Nowadays, smart contracts, which initially referred to the automation of legal contracts in general, have gained extensive interest due to the emergence of the blockchain technology. Legal contracts are fundamentally a legally binding agreement between two or more parties, with each party committed to fulfilling its obligations, while the agreement must be enforceable by law, often via a centralized legal body. On the other hand, smart contracts replace trusted third parties or mediators between contracting parties. They make use of this with the assistance of code execution that is automatically disseminated and validated by network nodes in a decentralized blockchain. Moreover, they enable transactions between untrusted parties without the necessity for direct contact between the parties, reliance on third parties and intermediary commission costs [58]. The smart contracts, comprising low-level code scripts running on a blockchain platform, find various important applications (e.g., crowdfunding) in the real world. Despite their increasing use, the development of smart contracts still comprises a difficult procedure mainly due to their special design and applications [63].

In this context, a Smart Contract Template Composer will offer an intuitive interface to data owners in order to select from a list of ready-made contract templates and use them for the creation of smart contracts containing the agreement details for the trading of their datasets. Moreover, such a template composer will enable data owners to customize these example templates and formulate their smart contracts based on their specific needs. The Smart Contract Composer will provide to all participating parties the opportunity to view and negotiate the associated data trading terms that will finally lead to an agreement. This composer will enable the transformation of such contract templates in the corresponding code that will be written in the blockchain and activate the related transactions.

6.1.2 Methods

Byzantine fault-tolerant consensus algorithms are the algorithms that allow digital security through decentralization to form a blockchain that supports smart contracts[65]. Byzantine fault-tolerant consensus algorithms can address the presence of malicious nodes in the environment and ensure the security and activity of the blockchain. Such algorithms are the following [64]:

[65] Smart contracts on blockchain – explained, Intellipaat Blog. Available at: https://intellipaat.com/blog/tutorial/blockchain-tutorial/smart-contracts-blockchain/?US#no3 (Accessed: 27 July 2023).

PBFT Algorithm

The PBFT algorithm (Practical Byzantine Fault-Tolerance) was first proposed by Barbara Liskov et al. in 1999. Inspired by the solution of the Byzantine general problem based on the oral message and signature message proposed by Lamport in 1982 [35], the algorithm reduces the complexity of the Byzantine fault-tolerant algorithm from the exponential level to polynomial level, and it increases the fault-tolerant performance by one-third on the premise of ensuring the stability and security of the original algorithm, making the Byzantine fault-tolerant algorithm more suitable for practice. The PBFT algorithm focuses on solving the problem of transaction ordering in distributed nodes that may have Byzantine errors.

PoW Algorithm

The PoW algorithm (Proof of Work) was originally designed to resist spam attacks. However, it was further promoted with its use in the Bitcoin system. In the Bitcoin white paper, Satoshi Nakamoto officially used the PoW algorithm in distributed applications. This is also the first time that a Byzantine fault-tolerant consensus algorithm has been used in an open and public blockchain network. The PoW algorithm aims to facilitate the consistency of data and the security of consensus through computing power competition.

PoS Algorithm

Due to drawbacks of the PoW algorithm, such as waste of computing power and low efficiency in reaching consensus, the PoS algorithm (Proof of Stake) was proposed in 2011 and practically applied in the "Peercoin" digital currency in the following year. Opposing to the PoW algorithm where the nodes obtain the accounting right through the competition of computing power, the PoS algorithm selects the node with the highest stake in the system as the accounting node, promoting the concept of tokens.

DPoS Algorithm

In the DPoS algorithm (Delegated Proof of Stake) the number of currencies held by each node in the system is the basis for the number of votes allocated to it. Through voting, nodes that are more credible are elected as the decision makers in the consensus process. According to the number of votes obtained, multiple decision makers are allocated, and decision makers take turns to obtain the accounting rights of the block. Every node in the system may become a decision maker. Through the multi-decision makers mechanism, the accounting rights will not be concentrated on a single node, which hinders excessive centralization.

PoH Algorithm

The PoH algorithm (Proof of Humanity) is designed for gathering society-related taxes. In the PoH algorithm, the probability of a node becoming the leader with accounting rights depends on the amount of its donation to the blockchain network. The PoH algorithm replaces the computing power of the traditional PoW algorithm through the donation system to compete for the right of accounting. To participate in the accounting, nodes pay a relatively small fee and do not purchase expensive computing equipment to compete for computing power, resulting in reduction of energy consumption and costs. In addition, through donations, most nodes in the blockchain network have the opportunity to benefit and thus, become willing to maintain the stable operation of the blockchain network.

CW-PoW Algorithm

The CW-PoW algorithm (Compute and Wait Proof of Work) is based on the PoW algorithm and the nodes need to solve the target value to compete for the accounting right to reach a consensus. However, multiple rounds of calculation and evaluation are established and the difficulty of solving the target value in each round is much lower than that of the PoW algorithm. Using layer-by-layer competition, the node with the smallest standard deviation from the target value in each round is selected as the accounting node.

6.1.3 Technologies

Solidity[66]

Solidity is an object-oriented, high-level language for implementing smart contracts. Smart contracts are programs that govern the behavior of accounts within the Ethereum state. Solidity is a curly-bracket language designed to target the Ethereum Virtual Machine (EVM). It is influenced by C++, Python, and JavaScript. Solidity is statically typed, supports inheritance, libraries, and complex user-defined types, among other features.

Bamboo[67]

Bamboo is a programming language for Ethereum contracts. Bamboo makes state transition explicit and avoids re-entrance problems by default.

Vyper[68]

Vyper is a contract-oriented, pythonic programming language that targets the Ethereum Virtual Machine (EVM). Vyper makes possible to build secure smart-contracts with a simple compiler implementation. Vyper code is mainly human-readable and hinders the writing of misleading code. Simplicity for the code readers is one of the core advantages of Vyper.

DAML[69]

Digital Asset Modeling Language (DAML) is a multiparty application platform from Digital Asset with privacy and interoperability competences. This open-source smart contract language is used to create composable applications based on an abstract DAML ledger model. It also helps with agreement modeling and operates on several blockchain systems.

Golang[70]

The Go programming language is an open-source project that can be used for the implementation of smart contracts. Its concurrency mechanisms make it easy to write programs that get the most out of multicore and networked machines, while its novel type system enables flexible and modular program construction.

[66] https://soliditylang.org/
[67] https://github.com/pirapira/bamboo
[68] https://docs.vyperlang.org/en/stable/
[69] https://docs.daml.com/daml/intro/0_Intro.html
[70] https://go.dev/

6.2 PUBLIC LEDGER

6.2.1 Overview

Distributed Ledger Technologies (DLT) have received growing attention in recent years as an innovative method of storing data. It is a fully decentralized record-keeping system used to record information. Thus, such a technology is of paramount importance for storing the smart contracts, lineage information and data transactions in the context of PISTIS. Permissionless, or public, ledgers are seen by some as the purest form of Blockchains [11] and a typical example of public ledger is Bitcoin network.

6.2.2 Methods

Several kinds of DLTs exists in the literature, and every individual of them have different methods to work. The most known DLTs are the blockchain, the Hashgraph, the Holochain and the Directed Acyclic Graph (DAG) [45]. These projects share the common aspects of distributed, consensus, flexible, and peer-to-peer platforms. However, even though there are many similarities among the different DLTs, there are also some architectural differences [19]. Below we provide with more detail a summary for these DTL mechanisms. In the context of PISTIS we will focus on the blockchain technology as the most well established for identity management and data trading, while on top of that we will use directed acyclic graph technology for monetization purposes and energy efficiency. Thus, blockchain oracles will also be utilised to handle the synchronisation of these two different types of DLTs. However, more details and elaboration on the selection of two different technologies will be provided in the architecture deliverable D4.1. In a nutshell, blockchain oracles are entities that connect blockchains to external world and enable smart contracts to be executed based upon inputs and outputs from the real world.

Blockchain

The blockchain has been around for over a decade as a proven technology for recording transactions in a decentralized manner using a distributed database. Blockchain technology is one of the most well-known uses of DLT, in which the ledger comprises blocks of transactions. A blockchain is a transparent and immutable globally distributed ledger, distributed databases, who have global agreement by all its users [51]. Current literature showcases various distributed data-sharing methods with promising results [38], [23] utilising the distributed trust mechanism of blockchain and smart contract technology, to deliver secure access control mechanisms, as well as tracking, managing and enforcing of the respective data sharing agreements; removing the need for a trusted 3rd party, since being on a distributed public ledger, these transactions are visible to all, while on the same time offer a permanent audit trail, increasing truthfulness and trust in the data sharing operations. In parallel, Zero Knowledge Proof (ZKP) can be integrated to blockchain systems to ensure privacy.

Hashgraph[71]

Hashgraph is a form of distributed ledger technology that connects transactions with each other in the form of a directed graph. Unlike typically DLTs, hashgraph achieves transaction success exclusively by consensus. To do so, the nodes on the hashgraph can encounter fairness

[71] https://hedera.com/

based on virtual voting and gossip methods. Thus, through a system called 'gossip about gossip' the transaction information is distributed within the network. These gossip events are then documented in the form of the hashgraph, and every participant receives a copy of the hashgraph using a hash function.

Holochain

Holochain[72] is an agent centric hybrid DLT network of blockchain, BitTorrent, and GitHub. Every user has its own chain of transactions which is called the source chain. It distributes data across nodes to limit centralized control over data flow. Users can store data using specific keys, by utilising distributed hash tables (DHT). Distributed hash tables work by converting transaction information into hash functions and distributing these hash functions randomized within the network. This data remains in physical places distributed throughout the world.

Directed Acyclic Graph[73]

Directed Acyclic Graph (DAG) is a popular distributed ledger technology that uses consensus methods in such a fashion that transactions that succeed just require the network's majority support. DAG is based on a structure of connecting blocks and forming a chain, while only one transaction is contained inside the block. In order to execute a transaction, it is necessary to validate to other transactions before and by this process a DAG is formed.

6.2.3 Technologies

Several DLT technologies exist for the aforementioned DLT methods. The most common DLTs are described below. **Ethereum**

Ethereum[74] is a well-known, decentralized, and open source blockchain that enables the possibility of building and running smart contract the Decentralized Applications (DApps). Ethereum introduced in 2015 and is classified as a public blockchain technology, where developers can contribute and interact anonymously. After the birth of Ethereum, smart contract applications have gradually become prevalent, while many other platforms have been derived [3].

HyperLedger Fabric

HyperLedger Fabric[75] is an open source blockchain, created by IBM and further developed by the Hyperledger Foundation. Fabric supports pluggable consensus algorithms (e.g., PBFT, Raft, and Kafka), allowing one to choose the consensus protocol that best suits their needs. Additionally, it offers privacy for communication via the design of channels and a membership mechanism to restrict channel access.

Hyperledger Besu

Hyperledger Besu [76] is an advanced Ethereum client that facilitates confining a permissioned and privacy enabled Ethereum network using the privacy group feature. It is more suitable to

[72] Holochain, Website: https://github.com/holochain/holochain-proto
[73] https://en.wikipedia.org/wiki/Directed_acyclic_graph
[74] Ethereum, Website: https://ethereum.org/en/
[75] HyperLedger Fabric, Website: https://www.hyperledger.org/use/fabric
[76] Hyperledger Besu, Website: https://www.hyperledger.org/use/besu

develop applications that require security, high-performance in private transaction processing since it is scalable, reliable, and offers secured off-chain privacy [47]. In other words, Hyperledger Besu offers both the security and privacy of a permissioned blockchain, as well as the flexibility and compatibility of a public blockchain. It supports both Proof of Work (PoW) and Proof of Authority (PoA) consensus protocols.

Corda

Corda[77] is an open source blockchain, created by R3 and designed to operate better privacy level on transaction. Corda enables different business entities to transact without incurring expensive transactional costs. It is designed for managing private transactions by creating business network membership allowed to access their data.

Hedera

Hedera Hashgraph is an efficient consensus mechanism that makes transactions reach finality very fast, while keeping the time and order of transactions fair, with no security trade-offs [7]. The patented algorithm is the first Asynchronous Byzantine Fault Tolerant consensus mechanism, designed and it has been mathematically proven through Coq, a proof management system developed at INRIA. Hedera provides services such as cryptocurrency, smart contracts written in Solidity and a file storage, where files of any size can be appended to the ledger.

IOTA's Tangle

IOTA's Tangle[78] is one of the most known networks that use the DAG data structure. In this case, miners/nodes can undertake dual functions that nodes in the blockchain do independently. This means that a Tangle miner can originate and approve a transaction at the same time.

6.3 Smart Contract Execution Engine

6.3.1 Overview

As blockchain has developed, smart contracts have become more popular [39]. Smart contracts are a form of automated digital contract in which the terms of the transaction are embedded in computer code, to be automatically fulfilled by the software upon acknowledgement of a particular input. More specifically, smart contracts are 'a set of promises, specified in digital form, including protocols within which the parties perform on these promises' [59]. On top of that, the technological basis for smart contracts is DLT, which offer a decentralized platform to perform smart contracts. The management and sharing of data are performed through the use of smart contracts and DLTs and are of paramount importance. In PISTIS, as already mentioned, the underline DLT will be the Hyperledger Besu, and thus the supported smart contracts will be in solidity. Such a component could also utilise a trusted component (e.g., TPM or TEE) for assuring its trust. Thus, the smart contract

[77] Corda, Website: https://www.r3.com/corda-platform/
[78] https://www.iota.org/get-started/what-is-iota

execution engine will act as a secure blockchain oracle where the smart contract execution will be conducted inside a TEE.

6.3.2 Methods

In this section we provide the most known smart contract methods relevant to compilation/deployment and oracles.

Smart Contract Compilation & Deployment

In a nutshell, to deploy a smart contract, a transaction containing the compiled code (e.g., contract's bytecode) of the smart contract is needed. Several promising tools exist to assist the development, testing and deployment of the contract such as Hardhat, Truffle, and Remix. These tools provide distinct features making the process of testing and deploying smart contracts a lot easier.

Blockchain Oracles

PISTIS's will require a secure oracle to support the interoperability between the two adopted DLTs, the HyperLedger Besu for identity management and the IOTA's Tangle for monetization. More details on the selected approach of PISTIS will be provided in the reference architecture in deliverable D4.1. Blockchain oracles can be used as inbound and outbound plug-in, meaning that not only incoming data, from outside world to smart contracts and PISTIS platform, but also outcoming data from PISTIS platform to outside world, will be monitored and filtered by the oracles. In parallel, the oracles can be supported and embedded within a trusted execution environment (TEE) so that it can run secure operations (e.g., secure oracles).

6.3.3 Technologies

In this section we provide the most known smart contract development and testing tools and oracle types.

6.3.3.1 Smart Contract Development and Testing Tools

Remix IDE

Remix IDE[79] allows developing, deploying, and administering smart contracts. It is a browser-based blockchain tool or writing, testing, unravelling, and deploying smart contracts either in the browser or locally.

Hardhat

Hardhat[80] is a development environment to compile, deploy, test, and debug a smart contract.

Truffle

Truffle [81] is development environment, testing framework, build pipeline for Ethereum applications. This tool offers a complete ecosystem for dApps development.

[79] https://remix.ethereum.org/
[80] https://hardhat.org/hardhat-runner/docs/getting-started#overview
[81] https://trufflesuite.com/

IOTA Smart Contracts

Since the last protocol upgrade known as IOTA 2.0[82], it is possible to support smart contracts. This transition from version 1.5 (Chrysalis) to 2.0 (Coordicide) supports the creation of smart contract using a virtual machine called Wasp. This virtual machine provides the necessary infrastructure for creating and deploying smart contracts, including the important features of contract templates, execution environments and communication outside of the chain using oracles.

6.3.3.2 Blockchain Oracles

In general, each type of blockchain oracle involves some combination of fetching, validating, computing upon, and delivering data to a destination. Blockchain oracles can be categorised to the following main categories[83].

Input Oracles

Input Oracles, as its name suggest, transmit information from the real-world (off-chain) and delivers it onto a blockchain network for smart contract consumption. An example of this type of oracle is one that provides to a smart contract what the temperature is measured by a sensor.

Output Oracles

Output Oracles are the opposite of the Input Oracles, which allow smart contracts to send information from smart contracts to the external world (e.g., commands to off-chain systems that trigger them to execute certain actions). A typical example of this type of oracle could be a smart lock. For instance, if funds are deposited to an address, the smart contract provides this information through an outbound oracle to a mechanism that unlocks the smart lock.

Cross-chain Oracles

Cross-chain oracles can read/write information between different blockchains, enabling the interoperability functionality. Typical examples could be moving data and assets between blockchains, such as using data on one blockchain to trigger an action on another or bridging assets cross-chain so they can be used outside the native blockchain they were issued on.

Computation Oracles

Computation Oracles or compute-enabled Oracles use secure off-chain computation to provide decentralized services that are impractical to do on-chain due to technical, legal, or financial constraints. For instance, one might use a computation oracle to perform a computationally intensive regression calculation to estimate the yield of a bond contract.

[82] https://wiki.iota.org/learn/research/iota-2.0-coordicide/
[83] https://chain.link/education/blockchain-oracles#types-of-blockchain-oracles

6.4 TRANSACTION AUDITOR

6.4.1 Overview

This component as its name suggest is necessary for auditing transactions and will be triggered prior to any smart contract execution. Transaction Auditor service will be quired to monitor the performed transactions and to provide auditable evidence upon request.

6.4.2 Methods

Every time that a transaction is performed a new smart contract instance will be generated that will be used as a certificate (including the metadata of the transaction, the pointer of the smart contract trading transaction and PISTIS platform signature) that the transaction was valid.

6.4.3 Technologies

This component is relevant to the underline DLT technology and the supported smart contracts. More details will be provided in sections 6.2 and 6.3.

6.5 ON/OFF PLATFORM CONTRACT INSPECTOR

6.5.1 Overview

Upon the creation and signing of a smart contract, the On/Off Platform Contract Inspector will be responsible for the continuous monitoring of the status of the smart contract. It will check the usage of the underlying dataset that is related with the smart contract and will make a thorough comparison with the contract terms, by exploiting mechanisms that will operate both on-platform (checking of contract terms) and off-platform (monitoring of dataset after trading). In this sense, it will explore and discover any potential breaches in the ways that the dataset has been used based on the initially agreed and signed contract terms and will communicate such events to the data owner through an intuitive user interface that will offer him/her the opportunity to take any necessary actions. The On/Off Platform Contract Inspector will also analyse the usage of the traded dataset and provide the data owner with useful statistics that will help him/her understand how the dataset has been exploited with a series of detailed metrics.

6.5.2 Methods

Event listeners is the most common and prominent method for detecting specific actions or states on the blockchain that signal usage of the dataset or changes to the smart contract. Smart contract events can be used to log important actions and state changes in real time. Logging events and state transitioning allow for efficient and straightforward reading of historical records without the need to search through the entire blockchain.

Activity logging in smart contract operation has specific definition and implementation in each blockchain. Event logging is an essential feature in smart contracts, especially in Ethereum, which has a built-in mechanism for this purpose. In other blockchain technologies (e.g., IOTA) smart contracts activity can be recorded as state transitions. Hyperledger Fabric, unlike Ethereum, does not have a native concept of "events" built into the smart contract (chaincode) language itself. However, Hyperledger Fabric does provide a mechanism to log and emit events from chaincode, which can then be picked up by client applications.

Ensuring that data shared off-chain is used as intended and adheres to any agreed-upon terms is a challenging task. Unlike on-chain actions, which occur on transparent and immutable ledgers, off-chain activities aren't as easily verifiable and are more susceptible to misuse or breaches of trust. However, methods that help to reassure or monitor the use of off-chain data do exist.

One such popular method is Digital rights management (DRM) [16]; DRM aims to control and monitor how digital content is accessed and consumed. While these systems are not fool proof, they can deter unauthorized usage.

6.5.3 Technologies

To access the activity data of the smart contracts, each platform provides its own API and/or set of libraries. Blockchain access libraries and APIs enable developers to interact with blockchain networks. They simplify tasks like querying the blockchain, sending transactions, and monitoring events. An indicative list of such libraries contains web3js[84] and etherjs[85] for Ethereum, SDKs for IOTA[86] and Hyperledger Fabric[87], etc.

To facilitate monitoring the activity of smart contracts along with the rest of the blockchain, online tools or platform that allows users to view information about blocks, transactions, addresses, and other data on a blockchain. These explorers provide a visual and user-friendly way to navigate the often-complex datasets within a blockchain, making them invaluable tools for both casual users and developers alike. It is very common each blockchain technology to be supported by a respective explorer, but cross platform explorer does exist such as Blockchair[88].

Log collection and analysis can be further benefitted by log management systems. Log management systems are tools that handle vast amounts of computer-generated log messages (logs). These systems aggregate, analyse, and offer ways to interpret all types of logs for various purposes: from debugging and security monitoring to compliance and user behaviour tracking. The most significant features and benefits of log management systems include:

- Aggregation: Collects log data from various sources, such as applications, servers, network devices, and more, into a centralized location.
- Storage: Safely stores logs in a structured manner, often with considerations for redundancy and fault tolerance.
- Analysis: Offers capabilities for parsing and interpreting log data, which can then be used for alerting, visualization, and deeper analysis.
- Real-time Monitoring: Allows users to monitor logs in real-time, which can be crucial for identifying and addressing ongoing issues or threats.

[84] https://web3js.readthedocs.io/en/v1.10.0/
[85] https://docs.ethers.org/v5/
[86] https://wiki.iota.org/iota.rs/libraries/java/api_reference/
[87] https://hyperledger-fabric.readthedocs.io/en/latest/sdk_chaincode.html

[88] https://blockchair.com/

- Search & Filter: Provides efficient ways to search through vast amounts of log data using keywords, patterns, and filters.
- Alerting: Sends notifications based on specific triggers or anomalies in the log data.
- Retention: Ensures logs are retained for a set period, catering to both operational needs and compliance requirements.
- Access Control: Offers mechanisms to control who can access the logs and what they can do with them, ensuring data privacy and security.
- Visualization: Graphs and dashboards offer visual representation of log data, helping in faster understanding and decision-making.

The most popular tools for log management are:

- Splunk[89]: A solution that provides real-time log analysis, visualization, and more. It's popular for its powerful search capabilities.
- ELK Stack (Elasticsearch, Logstash, Kibana)[90]:
    - Elasticsearch: A search and analytics engine.
    - Logstash: A log aggregation tool.
    - Kibana: Offers visualization capabilities for Elasticsearch.
- Graylog[91]: Built on top of Elasticsearch, it offers powerful log aggregation and analysis capabilities.
- Datadog[92]: While primarily a monitoring and analytics platform, it offers extensive log management features.

[89] https://www.splunk.com/
[90] https://www.elastic.co/elastic-stack
[91] https://graylog.org/
[92] https://www.datadoghq.com/

7 CONCLUSIONS

The deliverable D3.1 has documented the methods and technologies that have been considered by the WP3 partners for the design and implementation of the components that will comprise the PISTIS Market Exchange Platform. A thorough review has been conducted to explore and specify the candidate solutions for the operations of the platform.

In the following next project period, following the definition of the first version of the architecture and elaborating on the user stories and requirements coming from WP1, the preferred options for each component presented above will be evaluated and selected by the responsible partners, catering for the highest degree of interoperability between the components, expandability and adherence with standards and the directions provided in global initiatives such as GAIA-X and IDSA.

This activity will mark the commencement of the development activities under WP3, that will be described in the corresponding deliverables D3.2, D3.3 and D3.4.

8 REFERENCES

[1] Agarwal, A., Dahleh, M., & Sarkar, T. (2019). A Marketplace for Data: An Algorithmic Solution. Proceedings of the 2019 ACM Conference on Economics and Computation, 701–726. https://doi.org/10.1145/3328526.3329589

[2] Ahituv, N. (1980). A Systematic Approach Towards Assessing the Value of an Information System. MIS Quarterly, 4(4), 61–75. https://doi.org/10.2307/248961

[3] Amirmohammad Pasdar, Young Choon Lee, and Zhongli Dong. 2023. Connect API with Blockchain: A Survey on Blockchain Oracle Implementation. ACM Comput. Surv. 55, 10, 2023, 39 pages.

[4] Askham, N., Cook, D., Doyle, M., Fereday, H., Gibson, M., Landbeck, U., Lee, R., Maynard, C., Palmer, G., & Schwarzenbach, J. (2013). The Six Primary Dimensions for Data Quality Assessment—Defining data quality dimensions. DAMA UK. [Crossref]

[5] Attard, J., Debattista, J., & Brennan, R. (2019). Saffron: A Data Value Assessment Tool for Quantifying the Value of Data Assets. [Crossref]

[6] Attard, J., Orlandi, F., & Auer, S. (2017). Exploiting the Value of Data through Data Value Networks. Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance, 475–484. https://doi.org/10.1145/3047273.3047299

[7] Baird, L., Harmon, M., & Madsen, P. (2019). Hedera: A public hashgraph network & Governing Council.

[8] Batini, C., Castelli, M., Viscusi, G., Cappiello, C., & Francalanci, C. (2018). Digital information asset evaluation: A case study in manufacturing. ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 49(3), 19–33. https://doi.org/10.1145/3242734.3242737

[9] Belleflamme, P., (2018). Economic Models of Data, Working Paper, August 2018.

[10] Berentsen, Aleksander, and Fabian Schär. "Stablecoins: The quest for a low-volatility cryptocurrency." The economics of Fintech and digital currencies (2019): 65-75.

[11] Brennan, Charles, & William Lunn. 2016. Blockchain: The Trust Disrupter. London: Credit Suisse. As of 14 March 2017

[12] Brennan, R., Attard, J., Petkov, P., Nagle, T., & Helfert, M. (2019). Exploring Data Value Assessment: A Survey Method and Investigation of the Perceived Relative Importance of Data Value Dimensions: Proceedings of the 21st International Conference on Enterprise Information Systems, 200–207. https://doi.org/10.5220/0007723402000207

[13] Bullmann, Dirk, Jonas Klemm, and Andrea Pinna. "In search for stability in crypto-assets: are stablecoins the solution?." Available at SSRN 3444847 (2019).

[14] Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal, 14(2), 1–10. https://doi.org/10.5334/dsj-2015-002

[15] Chang, W.L. NIST Big Data Interoperability Framework: Volume 1, Definitions; NIST Big Data Public Working

[16] Cohen, J. E. (2003). DRM and Privacy. Communications of the ACM, 46(4), 46-49.

[17] Diepeveen, S., & Wdowin, J. (2020). The Value of Data—Literature Review. Bennett Institute for Public Policy, Cambridge in partnership with the Open Data Institute. [Crossref]

[18] Dodds, L., (2019), “How can we describe different types of dataset? Ten dataset archetypes”, [URL] accessed on August 30, 2023.

[19] El Ioini, N., & Pahl, C. (2018). A review of distributed ledger technologies. Springer, Cham, In OTM Confederated International Conferences, On the Move to Meaningful Internet Systems, (pp. 277-288).

[20] Even, A., & Shankaranarayanan, G. (2006, November 10). Value-Driven Data Quality Assessment. Proceedings of the 2005 International Conference on Information Quality. MIT IQ Conference, MIT, Cambridge, MA, USA. [Crossref]

[21] Faroukhi, A. Z., El Alaoui, I., Gahi, Y., & Amine, A. (2020). Big data monetization throughout Big Data Value Chain: A comprehensive review. Journal of Big Data, 7(1), 3. https://doi.org/10.1186/s40537-019-0281-5

[22] Fette, Ian, and Alexey Melnikov. The websocket protocol. No. rfc6455. 2011.

[23] Firdaus, M; Rahmadika,S.; Rhee, K.-H. Decentralized Trusted Data Sharing Management on Internet of Vehicle Edge Computing (IoVEC) Networks Using Consortium Blockchain. Sensors 2021, 21, 2410.

[24] Fleckenstein, M., Obaidi, A., & Tryfona, N. (2023). A Review of Data Valuation Approaches and Building and Scoring a Data Valuation Model. Harvard Data Science Review, 5(1).

[25] for data-based value creation in information-intensive services. Int. J. Inf. Manag. 2018, 39, 121–135.

[26] Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2020). Datasheets for Datasets. ArXiv:1803.09010 [Cs]. http://arxiv.org/abs/1803.09010

[27] Ghorbani, A., & Zou, J. (2019). Data Shapley: Equitable Valuation of Data for Machine Learning (arXiv:1904.02868). arXiv. https://doi.org/10.48550/arXiv.1904.02868

[28] Group: Gaithersburg, MD, USA, 2015.

[29] Holland, S., Hosny, A., Newman, S., Joseph, J., & Chmielinski, K. (2020). The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards. In D. Hallinan, R. Leenes, S. Gutwirth, & P. De Hert (Eds.), Data Protection and Privacy: Data Protection and Democracy (pp. 1–26). Oxford: Hart Publishing. https://doi.org/10.5040/9781509932771.ch-001

[30] Kannan, K., Ananthanarayanan, R., & Mehta, S. (2018). What is my data worth? From data properties to data value. ArXiv:1811.04665 [Cs]. http://arxiv.org/abs/1811.04665

[31] Kastius, A. and Schlosser, R. (2022) ‘Dynamic pricing under competition using reinforcement learning’, Journal of Revenue and Pricing Management, 21(1), pp. 50–63. Available at: https://doi.org/10.1057/s41272-021-00285-3

[32] Kaufmann, M. (2019). Big Data Management Canvas: A Reference Model for Value Creation from Data. Big Data and Cognitive Computing, 3(1), Article 1. https://doi.org/10.3390/bdcc3010019

[33] Koniakou, V., Baloutsos, S., & Farmakis, T. (2022). Data pricing recommendation tool. i3-Market. [Crossref]

[34] Kostick-Quenet, K. et al. (2022) ‘How NFTs could transform health information exchange’, Science, 375(6580), pp. 500–502. Available at: https://doi.org/10.1126/science.abm2004.

[35] Kuhn, E. (2019, August 13). Let’s Talk About Data Pricing—Part I. Medium. [URL]. Accessed: September 14, 2023

[36] Lamport, L. The Byzantine Generals Problem. Acm Trans. Program. Lang. Syst. 1982, 4, 382–401.

[37] Lim, C., Kim, K.-H., Kim, M.-J., Heo, J.-Y., Kim, K.-J., & Maglio, P. P. (2018). From data to value: A nine-factor framework for data-based value creation in information-intensive services. International Journal of Information Management, 39, 121–135. https://doi.org/10.1016/j.ijinfomgt.2017.12.007

[38] Lo, O.; Buchanan, W.J.; Sayeed, S.; Papadopoulos, P.; Pitropakis, N.; Chrysoulas, C. GLASS: A Citizen-Centric Distributed Data-Sharing Model within an e-Governance Architecture. Sensors 2022,22,2291.

[39] Macrinici, D.; Cartofeanu, C.; Gao, S. Smart contract applications within blockchain technology: A systematic mapping study. Telemat. Inform. 2018, 35, 2337–2354.

[40] Mawer, C. (2015, November 10). Valuing Data is Hard. Silicon Valley Data Science. URL: https://www.svds.com/valuing-data-is-hard/

[41] Mindermann, K., Keck, P. and Wagner, S. (2018) "How Usable Are Rust Cryptography APIs?," 2018 IEEE International Conference on Software Quality, Reliability and Security (QRS), Lisbon, Portugal, 2018, pp. 143-154, doi: 10.1109/QRS.2018.00028.

[42] OECD. (2013). Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value (OECD Digital Economy Papers, No. 220). OECD. http://dx.doi.org/10.1787/5k486qtxldmq-en

[43] OECD. Data-Driven Innovation: Big Data for Growth and Well-Being; OECD Publishing: Paris, France, 2015.

[44] Ohbyung Kwon, Namyeon Lee, Bongsik Shin, Data quality management, data usage experience and acquisition intention of big data analytics, International Journal of Information Management, Volume 34, Issue 3, 2014, Pages 387-394, ISSN 0268-4012

[45] Panwar A. and V. Bhatnagar V. (2020), “Distributed Ledger Technology (DLT): The Beginning of a Technological Revolution for Blockchain,” Feb. 2020. doi: 10.1109/IDEA49133.2020.9170699.

[46] Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. Communications of the ACM, 45(4), 211–218. https://doi.org/10.1145/505248.506010

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

[48] Project: 101093016 — PISTIS — HORIZON-CL4-2022-DATA-01 DESCRIPTION OF THE ACTION (PART A) (Associated with document Ref. Ares(2022)7911957 - 16/11/2022)

[49] Rajasekaran, A. S., Azees, M., & Al-Turjman, F. (2022). A comprehensive survey on blockchain technology. Sustainable Energy Technologies and Assessments, 52, 102039. https://doi.org/10.1016/j.seta.2022.102039

[50] Rea, N., & Sutton, A. (2019). Putting a value on data. PricewaterhouseCoopers. [Crossref]

[51] Risius, M.; Spohrer, K. (2017) “A Blockchain Research Framework,” Bus. Inf. Syst. Eng., vol. 59, no. 6, pp. 385–409

[52] Sajko, M., Kornelije, R., Miroslav, B. (2006) How to calculate information value for effective security risk assessment, Journal of Information and Organizational Sciences, 30(2), 263-278.

[53] Sammon, D., Nagle, T. (2017). The Data Value Map: A framework for developing shared understanding on data initiatives. AIS Electronic Library (AISeL). https://hdl.handle.net/10468/5167

[54] Shen, J. and Shafiq, M.O. (2020) 'Short-term stock market price trend prediction using a comprehensive deep learning system', Journal of Big Data, 7(1), p. 66. Available at: https://doi.org/10.1186/s40537-020-00333-6

[55] Sim, R. H. L., Xu, X., & Low, B. K. H. (2022). Data Valuation in Machine Learning: 'Ingredients', Strategies, and Open Challenges. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, 5607–5614. https://doi.org/10.24963/ijcai.2022/782

[56] Slotin, J. (2018). What Do We Know About the Value of Data? Global Partnership for Sustainable Development Data.

[57] Suratkar, S., Shirole, M. and Bhirud, S. (2020). "Cryptocurrency Wallet: A Review," 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP), Chennai, India, 2020, pp. 1-7, doi: 10.1109/ICCCSP49186.2020.9315193.

[58] Szabo, N. 1(996). 'Smart Contracts: Building Blocks for Digital Markets'. As of 8 March 2017

[59] Taherdoost, H. (2023). Smart Contracts in Blockchain Technology: A Critical Review. Information, 14(2), 117.

[60] Tufiş, M. (2021). Deliverable 4.5—Release of the Big Data Valuation component v2 [Deliverable]. EU Horizon 2020 Safe-DEED project. [Crossref]

[61] Tufiş, M., & Boratto, L. (2021). Toward a Complete Data Valuation Process. Challenges of Personal Data. Journal of Data and Information Quality, 13(4), 1–7. https://doi.org/10.1145/3447269

[62] Ursache, Cosmin, Michael Sammeth, and Sînică Alboaie. (2022) "OpenDSU: Digital Sovereignty in PharmaLedger." arXiv preprint arXiv:2209.14879

[63] Viscusi, G. and Batini, C. (2014) 'Digital Information Asset Evaluation: Characteristics and Dimensions', in L. Caporarello, B. Di Martino, and M. Martinez (eds) Smart Organizations and Smart Artifacts. Cham: Springer International Publishing (Lecture Notes in Information Systems and Organisation), pp. 77–86. Available at: https://doi.org/10.1007/978-3-319-07040-7_9.

[64] Zou, W. et al., "Smart Contract Development: Challenges and Opportunities," in IEEE Transactions on Software Engineering, vol. 47, no. 10, pp. 2084-2106, 1 Oct. 2021, doi: 10.1109/TSE.2019.2942301.

[65] Xiong H, Chen M, Wu C, Zhao Y, Yi W. Research on Progress of Blockchain Consensus Algorithm: A Review on Recent Progress of Blockchain Consensus Algorithms. Future Internet. 2022; 14(2):47. https://doi.org/10.3390/fi14020047