
Persistent Personal Data Vaults Empowering a Secure and Privacy Preserving Data Storage, Analysis, Sharing and Monetisation Platform

# D2.3
# "Updated DataVaults Security Methods and Market Design"

| | |
|---|---|
| **Editor(s)** | Marina Cugurra (ETA), Weizhi Meng (DTU), Alexander Köberl (IFAT), Thanassis Giannetsos (UBITECH) |
| **Lead Beneficiary** | ETA |
| **Status** | Final |
| **Version** | 1.0 |
| **Due Date** | 30/06/2021 |
| **Delivery Date** | 21/07/2021 |
| **Dissemination Level** | PU |

DataVaults is a project co-funded by the European Commission under the Horizon 2020 Programme (H2020-ICT-2019-2) under Grant Agreement No. 871755 and is contributing to the BDV-PPP of the European Commission.

 

| | |
|---|---|
| **Project** | DataVaults – 871755 |
| **Work Package** | WP2 - Security Aspects, Privacy Considerations, Value Generation and Commercialisation Outlines in Personal Data Management |
| **Deliverable** | D2.3 – Updated DataVaults Security Methods and Market Design |
| **Editor(s)** | Marina Cugurra (ETA), Weizhi Meng (DTU), Alexander Köberl (IFAT), Thanassis Giannetsos (UBITECH) |
| **Contributor(s)** | Sotiris Koussouris (SUITE5), Miguel Angel Mateo Montero (ATOS), Marina Cugurra (ETA), Weizhi Meng (DTU), Alexander Köberl (IFAT), Thanassis Giannetsos (UBITECH), Maria Jose Lopez Osa (TECNALIA), Christina Tsilihkiri (OLYMPIACOS), Michail Bourmpos (PIRAEUS), Sébastien Hannay (ANDAMAN7), Ramon Ruiz (MIWENERGIA), Elena Palmisano, Paolo Boscolo (PRATO), Shaun Topham (AS) |
| **Reviewer(s)** | Nikos Achilleopoulos (MAGGIOLI), John Kaldis (UNISYSTEMS), Yury Glikman (FRAUNHOFER) |

| | |
|---|---|
| **Abstract** | This deliverable reports the updated and final results of WP2. |
| **Disclaimer** | The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.<br><br>© Copyright in this document remains vested with the DataVaults Partners |

## Executive Summary

The document reports the updated and final results of WP2 towards the development of secure, trusted, auditable and privacy-preserving platform for data sharing economies, capable of enhancing data privacy and ownership safeguarding (privacy by design) , data provenance and sovereignty checking mechanisms, whilst respecting prevailing GDPR legislation.

The privacy, legal and ethically relevant properties of DataVaults technology, including the overall platform and Personal App, as well as their services and tools, as resulting from the project's progress have been analysed, in conjunction with the same properties and data protection remarks at demonstrator level.

Additional, potentially applicable regulatory sources have been investigated, such as the Law on trust services and electronic identification and the Data Governance Act under development.

Relying on the mentioned analysis and on the updated survey on the regulatory framework relevant to DataVaults, the set of legal and ethical requirements previously identified have been enriched, still applying a systematic and holistic approach, driven by Fairness & Privacy-by-Design-and-by-Default enriched with the Protection Goals method. This list, in conjunction with the privacy, security and trust requirements, which have been categorized as *mandatory* and *desirable*, are driving the development of DataVaults solutions towards an enhanced (holistic) data sharing solution capable of safeguarding fundamental rights and individuals' control over their personal data, whilst meeting the businesses and Data Seekers' expectations at the same time, moving forward the vision of win-win personal data platform with core security, privacy and trust services towards the support of enhanced data sharing economies.

As described in Deliverables D2.1 "Security, Privacy and GDPR Compliance for Personal Data Management" and D2.2 "Personal Data Market Design, Contracts and Rules", one of the core services that will be leveraged by the DataVaults platform towards enhancing the security posture of both the user devices but also the platform itself is remote attestation; both for verifying the correct state of a data user's device as well as for the privacy-preserving platform authentication when accessing and interacting with the DataVaults platform (data owners sharing/uploading their data). In terms of design, DataVaults will leverage advanced crypto primitives in the context of both static and dynamic attestation; namely, Configuration Integrity Verification (CIV) and Direct Anonymous Attestation (DAA). The focus is on the provision of secure, robust, and efficient attestation, verification and privacy-preserving methods to check the internal state of a Data Owner – when accessing the DataVaults trading ecosystem – whose level of trust has not been verified, thus, enabling secure enrolment and platform authentication services.

Moving to Smart Contracts, they are a core technology to enable trusted and secure data sharing with Distributed Ledger Technology (DLT). This deliverable presents the technical background of private smart contracts and their application in DataVaults, in particular for the Access Policy Engine, Attribute Based Encryption (ABE) and Searchable Symmetric Encryption (SSE).

As part of the compensation schemes, we introduce details of the Personal Wallet for privacy-preserving value transfer and summarize the results from the consultation with Data Seekers to validate the usefulness of the design. Finally, the flow of data and value through the DataVaults' components is illustrated in detail to determine the security and privacy relevant interfaces.

This document, representing the update and final version of WP2 outcomes, is aimed at driving the further design and development of the core security, privacy and trust services of DataVaults Platform and Personal App towards reaching DataVaults' vision in a trustworthy, legally-sound and value-focused manner fostering a human-centric Data Economy, fairly benefitting all the actors involved.

# Table of Contents

## List of Figures

## List of Tables



## Terms and Abbreviations

| | |
|---|---|
| **ABE** | Attribute-based Encryption |
| **BDVA** | Big Data Value Association |
| **DAA** | Direct Anonymous Attestation |
| **DFD** | Data Flows Diagram |
| **DGA** | Data Governance Act (proposal) |
| **DLT** | Distributed Ledger Technology |
| **DMA** | Digital Market Act (proposal) |
| **DoA** | Description of the Action |
| **DPIA** | Data Protection Impact Assessment |
| **DSA** | Digital Service Act (proposal) |
| **ECD** | e-Commerce Directive |
| **EDPS** | European Data Protection Supervisor |
| **eIDAS** | eIDAS Regulation |
| **ePD** | Directive 2002/58/EC on privacy and electronic communications (ePrivacy Directive) |
| **ePR** | ePrivacy Regulation (proposal) |
| **ESL** | Ethics and Soft Law |
| **GDPR** | “General Regulation on data protection” 2016/679 |
| **HRs** | Human Rights Law |
| **IoT** | Internet of Things |
| **ITSL** | Telecommunication Law and/or Information Technology-Security Law |
| **WP** | Work- Package |
| **ODRL** | Open Digital Rights Language |
| **PDBR** | Platform-To-Business Regulation |
| **PCRs** | Platform Configuration Registers |

| | |
|---|---|
| **PDPL** | Privacy and Data Protection Law |
| **PID** | Personal Information Diagram |
| **PET** | Privacy-enhancing Technologies |
| **RFSJ** | Regulatory Framework in the selected jurisdictions |
| **SC** | Smart Contract |
| **SoTA** | State of the Art |
| **SSE** | Searchable Symmetric Encryption |
| **TPM** | Trusted Platform Module |

# 1 INTRODUCTION

This deliverable is aimed at presenting the updated and final outcomes of WP2 in relation to the main security and privacy methods, the process of enabling end-to-end security, privacy and intelligent handling of personal data, data anonymization and pseudonymization techniques, data protection algorithms, encryption and authentication methods, as well as in relation to the approaches and framework for smart contracts and DLT for fair and secure personal data sharing and management of transactions, monetization and compensation mechanisms supported by DataVaults. Furthermore, the document describes the final findings of the legal survey relevant to the data to be used by DataVaults, to the whole DataVaults technology and to the demonstration activities. The document outlines the related legal, ethical, security, privacy and trust requirements to be considered during the design of the platform and app.

## 1.1 DOCUMENT STRUCTURE

The document is structured as follows:

- **Section 2** contains an update of survey on the regulatory and ethical instruments relevant to DataVaults personal data management, taking into account the main technical choices taken and the ongoing legislative reforms and new pieces of legislation;
- **Section 3** describes the facts and aspects of the project relevant in order to provide the legal analysis and to elicit the legal and ethical requirements, including the privacy-relevant properties and personal data collection, processing and sharing in each service and tool. It reflects the project's progress since the version contained in D2.1;
- **Section 4** contains the updated list the legal and ethical requirements to DataVaults design, development and operation, as well the final version of the security, privacy and trust requirements elicited from a technical point of view;
- **Section 5** is focused on the security, privacy and trust aspects and investigates the platform authentication and attestation aspects for the data owners, the secure communication channel between them and the platform in the sharing/uploading of personal data, as well as the use of TPM building blocks and services for the secure key management;
- **Section 6** is composed of three main bundles: i) the features of the smart contracts in DataVaults, including their functionalities, the access policy, the transaction privacy, as well as ABE/SSE contracts; ii) the compensation schemes, including findings on the micropayment and on the Personal Wallet, besides outlining the outcomes of the consultation with the Data Seekers;
- **Section 7** draws conclusions.

# 2 UPDATE OF THE REGULATORY AND LEGAL FRAMEWORK RELEVANT TO DATAVAULTS

In D2.1 the survey on the regulatory and ethical reference framework relevant to DataVaults, on the basis of which the legal and ethical requirements were elicited, was mainly focused on:

- the Privacy and Data Protection Legislation: special attention was given to i) GDPR ("General Regulation on data protection" 2016/679); ii) to the "ePrivacy Directive" (Directive 2002/58/EC on privacy and electronic communications), which replaced the Directive 97/66/EC and was partially amended by Directive 2009/136/EC; and iii) to the national legislations in the countries of the DataVaults demonstrators (Greece, Spain, Belgium and Italy);
- the Human Rights Law, in particular analysing the European Convention of Human Rights[1] and the Charter of Fundamental Rights of the European Union[2].
- Ethics and Soft Law, consisting of quasi-legal instruments, such as the European Courts' case law. These instruments, though not necessarily legally binding, usually are very helpful, especially in filling the gaps of the legislation, in identifying safeguards, boundaries and obligations to ensure the legitimacy and fairness of the new technologies and in identifying the balance between competing interests on a case-by-case basis. For instance, the elicitation of the ethics and legal requirements referred to the EC's Communications "AI for Europe" (25 April 2018) and "Building Trust in Human-Centric AI" (8 April 2019) and the European Data Protection Supervisor's Opinion 7/2015 "Meeting the challenge of Big Data. A call for transparency, user control, data protection by design and accountability".

Taking into consideration both the project progresses and the new pieces of legislation, the legal survey was extended to cover other areas of law and new sources (already applicable and/or under development), which might be relevant for DataVaults development and/or future uptake, and therefore have been investigated for eventually introducing additional requirements.

## 2.1 LAW ON TRUST SERVICES, IDENTIFICATION AND AUTHENTICATION

In this area, we investigated the possible relevance for DataVaults of the Regulation 910/2014 on electronic identification and trust services for electronic transactions in the internal market (eIDAS), which repeals Directive 1999/93/EC[3]. This source is aimed at ensuring the proper functioning of the internal market, facilitating seamless digital transactions among individuals and businesses across the same, and at creating a climate of trust in online and digital transactions. According to Art. 2, it applies to electronic identification schemes notified by a Member State, and to trust service providers established in the Union.

This Regulation consists of two main parts: one concerns the electronic identification, whilst the other regards the trust services (electronic signatures and other trust services).

It sets the conditions for the recognition of electronic identification means of natural and legal persons, the rules for trust services (especially for electronic transactions), besides introducing a legal framework for electronic signatures, electronic seals, electronic time stamps, electronic

[1] The European Convention on Human Rights, adopted in 1950 and entered into force in 1953.

[2] Charter of Fundamental Rights of the European Union, 2016/C 202/02.

[3] Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures

documents, electronic registered delivery services and certificate services for website authentication. eIDAS identifies three types of electronic signature, namely simple, advanced and qualified.

According to the Article 3, c. 16 of eIDAS, a trust service is "an electronic service normally provided for remuneration which consists of:

(a) the creation, verification, and validation of electronic signatures, electronic seals or electronic time stamps, electronic registered delivery services and certificates related to those services, or

(b) the creation, verification and validation of certificates for website authentication; or

(c) the preservation of electronic signatures, seals or certificates related to those services".

It might be relevant to DataVaults the provisions regarding the electronic registered delivery services, since it can fall into such concept. In fact, the electronic registered delivery service is defined by eIDAS a "service that makes it possible to transmit data between third parties by electronic means and provides evidence relating to the handling of the transmitted data, including proof of sending and receiving the data, and which protects transmitted data against the risk of loss, theft, damage or any unauthorised alterations" (Art. 3, (36) eIDAS).

The obligations of non-qualified electronic registered delivery service providers includes to verify that requirements of the Regulation applicable to (all) TSPs[4] are met:

- Data processing and protection (art.5)
- Liability and burden of proof, including limitation of use of the services (art.13)
- Access to person with disabilities (art.15)
- Risk management and security breach notification (art.19)

Interoperability is key for eIDAS, which also distinguishes between normal trust services and qualified trust services, imposing certain obligations to the provider of the latter to prevent and minimize the impact of security incidents or loss of integrity of its services.

On the other hand, art. 2 (2) eIDAS states that this regulatory source does not apply to "the provision of trust services that are used exclusively within closed systems resulting from national law or from agreements between a defined set of participants". This means that in case of private blockchain the regulation might not be fully applicable in some cases.

The eIDAS Regulation states that the processing of personal data must be carried out in accordance with the GDPR and respecting its principle of confidentiality and security of processing: as clarified in its Recital 11, the authentication for an online service should concern processing of only those identification data that are adequate, relevant and not excessive to grant access to that service online.

Anyway, in case DataVaults foresees to use electronic identification for its users, either natural or legal persons, this Regulation can become applicable for the whole project and/or for some of its demonstrators and should be investigated especially in the context of the wallets and the smart contracts. Its electronic identification (eID) tools can be used for the identification of users, as they broadly offer enhanced security and accuracy, swifter and less costly processes, while they may mitigate risk of fraud, identification theft and legal challenges.

[4] Trust Service Providers.

On the other hand, the concept of self-sovereign identity (SSI)[5][6] could also present advantages for the purpose of DataVaults deployment and use and should therefore be investigated, including its compliance with eIDAS.

Sovrin[7] argued that the "self-sovereign identity (SSI) is a term used to describe the digital movement that recognizes an individual should own and control their identity without the intervention of administrative authorities. SSI allows people to interact in the digital world with the same freedom and capacity for trust as they do in the offline world." Furthermore, "Blockchain and SSI are natural complements, making the perfect symbiosis" [8]: the user is able to individually create and manage his/her identify thanks to the use of distributed ledger technologies (e.g. blockchain), without the involvement of a third party, but often making use of the "decentralized identifier" (DID) associated with an entity. Such entity using SSI to authenticate itself can be an individual (natural person) and therefore, in this case, the DID usually relates to an identified or identifiable person (thus being personal data).

The SSI enables sovereignty for individuals over their digital assets and credentials, often by using digital wallets. In case the individual presents such assets and credentials to a third party to prove ownership, the public, decentralized, and immutable registry (such as a blockchain network) can be employed: the cryptographic proofs of the asset or credential were registered and are kept in a standardized and trustable way.

Nonetheless, the question whether eIDAS is already suitable for SSI and blockchain technology is still open, as well as whether, on the one hand, the smart contracts could be considered electronic documents and, on the other hand, the means used to sign blockchain transactions could be considered electronic signatures, with all the legal consequences it implies. Some scholars[9] argue that the eIDAS Regulation will need some adjustments to become the legal and trust framework for SSI in the European Union: it was created as a legal framework supporting a digital identity metasystem mainly based in delegated authentication, which is more limited than the self-sovereign approach which enables, among other things, pseudonymity and selective disclosure mechanisms.

In the US system the situation is not exactly the same and some authors underlined that blockchain transactions can constitute, or evidence, electronic signatures and that, virtually, all transactions stored on a blockchain, and retrievable in perceivable form, constitute an electronic record under the US law[10][11].

In conclusion, for the purposes of DataVaults it should be investigated how to ensure the electronic identification and to get the verifiable credential (on the basis of a national digital identity), where necessary for accessing to online public services.

On the other hand, from the viewpoint of the smart contract itself, the debate is still ongoing whether and to what extent and conditions, these can give rise to legally binding and

[5] Marcos Allende Lopez, "Self-sovreign identity. The future of Identity: self-sovreignity, Digital Wallet, Blockchain", 2020.

[6] Domingo, Ignacio Alamillo. 'SSI EIDAS Legal Report - How EIDAS Can Legally Support Digital Identity and Trustworthy DLT-Based Transactions in the Digital Single Market', April 2020.

[7] Sovrin, Sovrin Trust Assurance Framework, 2019. Retrieved from https:// sovrin.org/wp-content/uploads/Sovrin-Trust-Assurance- Framework-V1.pdf

[8] Marcos Allende Lopez, "Self-sovreign identity. The future of Identity: self-sovreignity, Digital Wallet, Blockchain", 2020.

[9] Marcos Allende Lopez, "Self-sovreign identity. The future of Identity: self-sovreignity, Digital Wallet, Blockchain", 2020

[10] Notably the Electronic Signatures in Global and National Commerce Act, "ESIGN": public law 106-229, June 30, 2000

[11] Therefore under certain legislation, blockchain platforms may constitute or store electronic records and electronic signatures and thus may be used to evidence, or give effect to, electronic or smart legal contracts

enforceable contracts and whether this necessarily requires the identification of the individual pursuant to eIDAS.
The smart contract satisfies the elements of a contract under several national laws, such as Spanish Civil Code and, therefore, smart contract code represents a valid mechanism to define the parties' contractual rights and obligations as a matter of contract law in many jurisdictions. Therefore, "under certain circumstances, and if so decided by the parties, smart contracts can fulfill the elements of a legally binding contract under common law and civil law systems[12]". Though the parties may act pseudonymously, it is necessary a link (including off-chain) to their real identity to provide for valid consent, which is a crucial element of a contract under several national systems. However, even if its deployment does not give rise to a legally binding contract, the smart contract may still affect legal relations (either between the parties or with third parties) and therefore may have legal effects.
At the same time, both smart contracts and conventional natural language contracts can coexist in relation to the same (or related) subject matter and create together the entire legal framework within which a smart contract operates. This is the case of the so-called "external smart contract", where "the code does not form the entirety of the parties' legal agreement, but merely automates the performance of some of its terms[13]". The code merely automates the performance of some of the conventional contract's terms. In this case the legal relationship is intended to be governed by the natural language version of the contract, rather than by the code. In the internal model, on the contrary, the code could either encompass the entire agreement between the parties, or, alternatively, could form only an integral part of the legally binding contract (rather than the entirety of the contract), and would supersede any other clauses written in natural language: the code would be given legal effect and is an integral part of the agreement.
Principally, it is necessary to refer to the governing law applicable to the smart contracts in order to determine whether these give rise to legally binding contracts, whether personal identification is necessary or not according to eIDAS, as well as to evaluate the effects of the DTL/blockchain, and, ultimately, to ensure that the model chosen meet local law requirements. However, considering that the DataVaults offering can constitute an electronic registered delivery service according to eIDAS (Art. 3, (36) eIDAS), such Regulations and the obligations established for the providers of such services have to be taken into account in the design, development and future use of DataVaults.

## 2.2 REGULATION ON THE FREE FLOW OF NON-PERSONAL DATA

The Regulation 2018/1807 on a framework for the free flow of non-personal data in the European Union, adopted by the EC applies to any form of data other than personal data, as defined in Art.4.1 of the GDPR. It is functional to create a comprehensive and coherent approach to the free movement and portability of data in the EU. Notably, its main objectives are to further promote the free movement of data and data processing services (Recital 4), whilst facilitating cross border availability of data, enhancing legal certainty and creating a level playing field through a single set of rules for all market participants. It supplements and complements the GDPR in issues related to non-personal data within the Digital Single Market, primarily concerning business and public sector users of data storage and processing services.

[12] Smart Contract Alliance, "Smart Contracts: is the Law Ready?", 2018.
[13] Smart Contract Alliance, "Smart Contracts: is the Law Ready?", 2018.

This instrument should be taken into account in relation to the non-personal data (such as insights, other derivatives, data related to the Persona and data completely anonymized) in the Project.

## 2.3 E-COMMERCE DIRECTIVE

Another important legislative source to be considered is the Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (e-Commerce Directive). Considering that the DataVaults services will be normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of the service, it falls under the scope of this directive and its provision are potentially relevant for DataVaults, to the extent that it will offer an information society service. Being a Directive, the national provisions implementing the Directive would need to be considered in each country where the DataVaults will be adopted. Section 4 on Intermediary Liability may be particularly relevant in case of illicit third-party content. Pursuant to such Directive:

- Easily reachable information to be provided to the user (Art. 5 e-Commerce Directive): name of the service provided, the geographic address at which the service provider is established, details of the service provider, including e-mail address, the register and registration number (in case of registration in a trade or similar public register) and tax registration number (in case of VAT);
- Information to be provided for the conclusion of a contract with a consumer in a clearly, comprehensively and unambiguously manner and prior to the conclusion of the contract (Art. 10 (1) e-Commerce Directive: i) the different technical steps to follow to conclude the contract; ii) Whether or not the concluded contract will be filed by the service provider and whether it will be accessible; iii) The technical means for identifying and correcting input errors prior to the placing of the order; iv) The languages offered for the conclusion of the contract;
- Contract terms and general conditions provided to the recipient must be made available in a way that allows him to store and reproduce them (Art. 10 (2) e-Commerce Directive);
- The Consortium should neither i) initiate the transmission, select the receiver of the transmission and select or modify the information contained in the transmission, as regards the transfer of content data (Art. 12 e-Commerce Directive), nor ii) monitor the data;
- In case of awareness of illegal activity or information, the Consortium must act expeditiously to remover or to disable access to the information (Art. 14 e-Commerce Directive)

## 2.4 PLATFORM-TO-BUSINESS REGULATION - P2BR

The Regulation 2019/1150 on promoting fairness and transparency for business users of online intermediation services (Platform-to-Business Regulation - P2BR) is a set of rules in the area of business platforms for creating a fair, transparent and predictable business environment for smaller businesses and traders on online platforms, in order to enable

consumers to receive the highest quality goods and services. The P2BR, which is part of the legislative measures promoted by the EC for the Digital Single Market strategy, foresees a list of measures ensuring transparency and fairness with the intent to temper the natural asymmetries that characterize the relationship between the platforms and their suppliers, establishing a fair and trustworthy innovation-driven ecosystem. Its Article 2 describes the requirements of the intermediation services (platforms) that fall into the scope of its application: "(a) they constitute information society services within the meaning of the European Electronic Communication Code; (b) they allow business users to offer goods or services to consumers, to facilitate the initiating of direct transactions between those business users and consumers, irrespective of where those transactions are ultimately concluded; (c) they are provided to business users based on contractual relationships between, on the one hand, the provider of those services and, on the other hand, both those business users and the consumers to which those business users offer goods or services". This definition of intermediaries describes only the services that have a direct relationship with business users and their clients without a clear threshold, applying indistinctively to all types of platforms falling in such criteria. The two main principles set by the P2BR are transparency and fairness. In particular:

- transparency obligations are foreseen for providers of intermediation services to inform, through clear, unambiguous and readily available contractual terms and conditions, about the treatment, the criteria used to rank their products and the requirements to suspend or terminate their services;
- Fairness should be achieved through the settlement of effective out-of-court redress mechanisms such as internal handling systems for business users and mediation procedures. To settle disputes, a list of independent mediators should be provided as part of the contractual terms and conditions prepared by the intermediaries.

In relation to DataVaults, considering whom the platform intends to offer its services to, it could fall within the P2BR scope. Nevertheless, it is still not entirely clear whether the it is applicable. It mainly depends whether DataVaults offering can be considered an online intermediation service, especially because, whilst it is likely that the data providers are businesses, it is unlikely that the data receivers are consumers, as requested by the definition of the online intermediation service, which is in principle applicable only for business users[14]. However, the positive answer seems the most reasonable.

## 2.5 DIRECTIVE ON CERTAIN ASPECTS CONCERNING CONTRACTS FOR THE SUPPLY OF DIGITAL CONTENT AND DIGITAL SERVICES

Another instrument analyzed is the Directive 2019/770 on certain aspects concerning contracts for the supply of digital content and digital services, which must be transposed into national law by 1 July 2021 (the date of entry into force of the transposition rules shall be 1 January 2022). Given that contracts are crucial for DataVaults, it is paramount to consider the

[14] Such services must have the following characteristics: being information society services, i) allowing business users to offer goods or services to consumers for facilitating the initiating of direct transactions between such business users and consumers ii) and provided to business users on the basis of contractual relationships between the provider of those services and business users (which, in turn, offer goods or services to consumers).

EU framework related to contractual agreements, that may be applicable in the context of the project. From a consumer policy perspective, considering the steps taken by the EC to implement a "digital update" of consumer contract law, it is widely recognized that consumers should enjoy the same level of protection under consumer contract law, whatever the object of consumption is. This Directive aims at the maximum harmonization and at introducing mandatory contractual liability for the non-conformity of digital content with the contract. It also extends the information duties as well as the right to withdraw from a contract in case of "free digital services" contracts, where consumers provide personal data instead of paying a fee. The Directive is directed to protect the consumer, understood as "any natural person who, in relation to contracts covered by this Directive, is acting for purposes which are outside that person's trade, business, craft, or profession" (Art.2.6)[15]. The Directive applies to "contracts of an indefinite or fixed duration which were concluded before the application date and provide for the supply of digital content or digital services over a period of time, either continuously or through a series of individual acts of supply, but only as regards digital content or a digital service that is supplied from the date of application of the national transposition measures", with the exception of the provisions on the modification of the digital content or digital service and the right to redress.

In relation to contractual agreements and consumer protection, also the following pieces of legislation can be considered: Directive 93/13/EEC on unfair terms in consumer contracts and Directive 2019/2161 (amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU) as regards the better enforcement and modernization of Union consumer protection rules.

## 2.6 SECURITY LAW

Though from a legal point of view the requirements related to security are mainly coming from the GDPR and the ePD, it is useful to understand the latest legislative developments in this area. Cybersecurity has been identified as one of the highest priorities for the EU: the achievement of a secure and safe environment is a precondition to enhance trust and to boost business opportunities. In this area of law, it is important to mention the Directive 2016/1148 on Security of Network and Information Systems (NIS) and the recently approved Regulation (EU) 2019/881 (Cybersecurity Act[16]).

The NIS Directive was part of the 2013 EU Cybersecurity strategy, comprising binding and non-binding legal instruments aimed at establishing a high standard of security across the European Union. It applies to:

- operators of essential services: any private or public entity that falls under one of the categories referred to in Annex II of the NIS. They are considered essential for the maintenance of critical societal and economic activities (Art 4 NIS); and
- Digital Service Providers: legal persons providing a digital service. There are Different Types of Digital Service Providers with a cross-border nature, listed in Annex III of the NIS, and they include online marketplace, online search engine or cloud computing service. They have to comply with a set of security and notification obligations to ensure the integrity and security of their services are subject to ex-post supervisory control by competent national authorities.

[15] Member States can extend the protection afforded to other persons who are not qualified as consumers.

[16] Regulation (EU) 2019/881 of the European Parliament and of the Council on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act).

The Cybersecurity Act was included in the Cybersecurity Package. It provides rules on the creation of an EU cybersecurity certification scheme for ICT products, ICT services, and ICT processes and aim to improve the cross-border coordination, besides promoting EU standards. The cybersecurity certification schemes for ICT products, ICT services, and ICT process might be of interest for DataVaults Consortium, since it can enhance security and trust in the DataVaults platform.
As regards the EU Encryption framework, the following documents are particularly interesting in view of DataVaults development: the ENISA[17] Opinion Paper on encryption (2016) and the European Electronic Communications Code (EECC), established with the Directive 2018/1972. This code, in its security provisions, makes reference to encryption protocols and explicitly, to the end-to-end encryption.

## 2.7 REGULATORY REFORMS UNDER DEVELOPMENT

Vast reforms are underway and an update of the European regulatory landscape was announced in terms of the Commission's Mission Statement for 2019-2025. The following are expected to be the most significant changes, relevant to DataVaults deployment and use.

### 2.7.1 The overarching framework

#### 2.7.1.1 A European Strategy for Data

The new European Data Strategy was presented along with the Commission's Communication on "Shaping Europe's digital future": data are embraced as the "lifeblood of economic development"; therefore, the EC aims at renewing its overarching framework to achieve the proper balance between, on the one hand, the wide availability and use of data and, on the other hand, the high preservation of privacy, security, safety and ethical standards. Aspects related to data ownership and data governance are going to be addressed and/or reframed. The Strategy is motivated by the need to put people first in developing technology and to defend and promote European values and rights in how the technology is designed and deployed in the real economy. The Strategy sets out a programme of policy reforms, already started with the Data Governance Act, the Digital Services Act, the Digital Markets Act and the Cybersecurity Strategy.

#### 2.7.1.2 New Deal for Consumer

Considering the envisaged role of individuals within DataVaults in their role as a data owner, it is an opportunity to follow the developments in terms of the European consumer protection framework, as already mentioned. More specifically, we ought to closely follow the developments related to so-called "New Deal for Consumers" initiative, adopted in 2018.[18] This initiative is functional to achieve a stronger and better enforced consumer protection rules in light of a growing risk of EU-wide infringements and at modernising EU consumer protection rules in view of market developments.

[17] European Union Agency for Network and Information Security
[18] Communication of the Commission of 11 April 2018—A New Deal for Consumers, (COM)2018, 183 final.

#### 2.7.1.3 Digital Compass Communication

The Communication "2030 Digital Compass: the European way for the Digital Decade"[19] has been produced by the EC in response to the President von der Leyen in the State of the Union Address (September 2020) announcement that Europe should secure digital sovereignty with a common vision of the EU in 2030, based on clear goals and principles. In particular, the European Council invited the EC to present a comprehensive Digital Compass to accelerate Europe's digital transformation, setting out digital ambitions for 2030 and outlining key milestones and the means of achieving these ambitions and intensifying actions defined in the strategy for Shaping Europe's digital future[20]. The Digital Compass Vision for 2030 relies on empowered citizens and businesses: "the European way to a digitalised economy and society is about solidarity, prosperity, and sustainability, anchored in empowerment of its citizens and businesses, ensuring the security and resilience of its digital ecosystem and supply chains" with four cardinal points for mapping the EU' trajectory:

- digitally skilled population and highly skilled digital professionals
- Secure and performant sustainable digital infrastructures
- Digital transformation of businesses
- Digitalisation of public services

The document also underlines the need to full respect of EU fundamental rights, including the freedom of expression (including access to diverse, trustworthy and transparent information), the freedom to set up and conduct a business online, the protection of personal data and privacy and right to be forgotten and the protection of the intellectual creation of individuals in the online space. It is envisaged the definition of a comprehensive set of digital principles allowing to inform users (besides guiding policy makers and digital operators), including, for instance, a secure and trusted online environment, the access to digital systems and devices that respect the environment, accessible and human-centric digital public services and administration, ethical principles for human centric algorithms and access to digital health services. The EC proposed to include these set of digital principles and rights within an interinstitutional solemn declaration between the European Commission, the European Parliament and the Council, as well as to carry out an annual Eurobarometer exercise specifically dedicated to monitoring the perception of citizens regarding the respect of their rights and values, and to what extent they feel the usefulness of the digitization of the society.

### 2.7.2 The Data Governance Act

On November 25, 2020, the EC published its draft Data Governance Act[21], which, as already remarked, is part of its 2020 European Strategy for Data, together with the Digital Services Act and the Digital Markets Act.

The DGA has been conceived to play a vital role in ensuring the EU's leadership in the global data economy, whilst empowering users to stay in control of their data. The DGA sets out

[19] COM(2021) 118 final. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, "2030 Digital Compass: the European way for the Digital Decade".

[20] COM (2020) 67 final "Shaping Europe's digital future".

[21] COM (2020) 767 final. Proposal for a Regulatorion of the European Parliament and the Council on European data governance (Data Governance Act)

policy measures and investments designed to capitalize on European vast quantity of data and, hence, to give the EU businesses a competitive advantage. The envisioned framework is expected to boost data sharing, encouraging a greater reuse of data by increasing trust in data intermediaries and strengthening various data-sharing mechanisms across the EU. In addition, the DGA will support the creation of EU-wide common, interoperable data spaces in strategic sectors (part of which are common to DataVaults demonstrators), such as health, energy and mobility, which, in turn, are meant to bring benefits to citizens. The proposal provides a broad definition of data: "any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audio-visual recording." This definition includes personal data as defined in the GDPR, which apply simultaneously to the DGA, as remarked also by some of its recitals and provisions. The explanatory memorandum which accompanies the DGA also underlines that its provisions and measures are fully compliant with the data protection legislation and increase in practice the control that individuals have over the data that they generate. This is an important element for DataVaults.

Many of its rules are potentially relevant for DataVaults. They include, among others:

- conditions for reuse of public sector data, which are subject to existing protections (such as intellectual property, commercial confidentiality and data protection);
- obligations on providers of various types of intermediation services within data-sharing services. New European rules on neutrality are defined to allow novel data intermediaries to function as trustworthy organisers of data sharing;
- establishment of a European Data Innovation Board, composed of experts and chaired by the European Commission;
- a set of measures to increase trust in data-sharing, due to the fact that the lack of trust is currently a major obstacle and results in high costs;
- data altruism, providing its concept and the possibility for organizations to register as "Data Altruism Organization recognized in the Union";
- measures to give the individuals the control on the use of the data they generate, in particular by making it easier and safer for companies and natural persons to voluntarily make their data available for the wider common good under clear conditions.

### 2.7.1 The Digital Service Act

The European Digital Service Act (DSA)[22] is expected to update and reform the framework established by the e-Commerce Directive, addressing the topics of intermediary liability and safety rules for digital platforms, including transparency, information obligations and accountability for digital services providers. At the same time, there is a strong call for maintaining the core principles of the e-Commerce Directive, its measures having the consumer protection at their core and the protection of fundamental rights in the online environment, as well as online anonymity wherever technically possible. In fact, the DSA builds on the key principles set out in the e-Commerce Directive, which is still applicable, seeking to ensure the best conditions for the provision of innovative digital services in the internal market, to contribute to online safety and the protection of fundamental rights, whilst setting a robust and durable governance structure for the monitoring and supervision of providers of intermediary services. The liability rules for providers of intermediary services, set out in the

[22] COM(2020) 825 final. Proposal for a Regulation of the European Parliament and the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC.

e-Commerce Directive, are maintained and are now included in this Act to ensure an effective harmonisation across the Union, and to avoid legal fragmentation. Therefore, the DSA deletes Articles 12-15 in the e-Commerce Directive and reproduces them, maintaining the liability exemptions of such providers, as interpreted by the Court of Justice of the European Union.
Since the adoption of the e-Commerce Directive, novel information society (digital) services have emerged, which, on the one hand, have deeply contributed to societal and economic transformations in the European Union and worldwide but, on the other hand, have brought new risks and challenges, both for society as a whole, and for individuals using such services.
The DSA, which is envisaged to be a standard-setter at global level, addresses the online marketplaces and consumer trust in the digital economy, while respecting users' fundamental rights and advocating for rules to underpin a competitive digital environment in Europe.
Clear responsibilities and accountability are defined for providers of intermediary services, and in particular online platforms, including marketplaces. Due-diligence obligations are set for certain intermediary services in order to improve users' safety online across the entire Union and improve the protection of their fundamental rights. Certain online platforms have the obligation to receive, store, partially verify and publish information on traders using their services in order to ensure a safer and more transparent online environment for consumers.
A higher standard of transparency and accountability is set for certain platform, as well as obligations to assess the risks their systems pose and to develop appropriate risk management tools to protect the integrity of their services against the use of manipulative techniques. However, the operational threshold for service providers in scope of these obligations includes only online platforms with a significant reach in the European market (currently set to more than 45 million recipients of the service).
The DSA is without prejudice to:

- the e-Commerce Directive, which defines the current EU legal framework regulating digital services
- the GDPR Regulation (EU) 2016/679 (the General Data Protection Regulation) and other Union rules on protection of personal data and privacy of communications.

The DSA will be complemented by further actions under the European Democracy Action Plan[23], aimed at empowering citizens and building more resilient democracies across the Union.

### 2.7.2 The Digital Market Act

The Digital Market Act[24] might be relevant to DataVaults in the future. Its objective is "to allow platforms to unlock their full potential by addressing at EU level the most salient incidences of unfair practices and weak contestability" in view of allowing end users and business users alike to reap the full benefits of the platform economy and the digital economy at large, in a contestable and fair environment. Nevertheless, its scope of application concerns "markets characterised by large platforms, with significant network effects acting as gatekeepers".

[23] COM(2020) 790 final.
[24] COM(2020) 842 final. Proposal for a Regulation of the European Parliament and the Council on contestable and fair markets in the digital sector (Digital Markets Act).

### 2.7.3 The Proposal of Regulation on Privacy and Electronic Communications (E-privacy Regulation)

Another legal instrument under development to monitor is the ePrivacy Regulation [25], intended to update European privacy framework, repealing the ePrivacy Directive, for a better alignment of the provisions of such Directive with those of the GDPR, while addressing the new challenges to privacy, brought about by the significance advancement of technology the last two decades. In fact, albeit objectives and principles of the existing framework remain sound and relevant, the essential technological, economic and business progresses, together with the ever-increasing penetration of the Internet in various aspects of the life and its vital role in the Digital Single Market, call for the modernization of the Directive. The choice of a Regulation is meant to improve the harmonization. As clarified in the proposal itself, it will be "lex specialis" to the GDPR: it will fine-tune and complement the GDPR as regards electronic communications data that qualify as personal data, whilst all matters concerning the processing of personal data not covered by the proposal remain regulated by the GDPR. In other words, in order to better understand their different scope and nature, it should be remarked that, on the one hand, the GDPR focuses on the protection of personal data and ensures the free flow of personal data across the European Union, while on the other hand, the ePrivacy Regulation refers to the protection of privacy when that data are being communicated electronically, thereby representing a medium and technology specific legal instrument compared to the GDPR (and its protections extend to legal persons as well).

The dual objective of the ePR is " the protection of fundamental rights and freedoms of natural and legal persons in the provision and use of electronic communications services, and in particular, the rights to respect for private life and communications and the protection of natural persons with regard to the processing of personal data" and the aim to ensure the "free movement of electronic communications data and electronic communications services within the Union, which shall be neither restricted nor prohibited for reasons related to the respect for the private life and communications of natural and legal persons and the protection of natural persons with regard to the processing of personal data." (Art.1.1 and 1.2)

In accordance to the existing draft, the scope of the ePR, similarly to the GDPR, is essentially global and, as provided in Article 2, it "applies to the processing of electronic communications data carried out in connection with the provision and the use of electronic communications services and to information related to the terminal equipment of end-users", clarifying that issues related to intermediary liability will be addressed in accordance to Directive 2000/31/EC.

The "electronic communication network" and "electronic communications service" are broadly conceived to bring also within the scope of the proposed ePR the "over-the top" services, and machine-to-machine communications in IoT and smart-environments context. Thereby, this source includes also new players providing electronic communications services and guaranteeing that new, yet unprecedent services will be covered.

Similarly, to the ePrivacy Directive, due attention is given to cookies.

[25] COM/2017/010 final. Proposal for a Regulation of the European Parliament and the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC

The general rule has been upheld in Art. 5: electronic communications data are to be kept confidential and "listening, tapping, storing, monitoring, scanning or other kinds of interception, surveillance or processing" is prohibited. However, the proposed Regulation sets a number of exceptions to provide some flexibility (Art.6).
Communication content and metadata are regulated by Article 7 and the proposed Regulation upholds also the general consent rule for market communications as well. When electronic contact details (e-mail) are obtained from existing customers in connection with a sale or purchase of a product or service, these details may be used for direct marketing communications regarding the same person's similar goods or services, provided that the customers are given an easy way to object, free of charge, both when the details originally are collected and then at each time a message is sent. When electronic communications services are used to send the direct marketing messages, the marketing nature of the communication must be indicated and the person on whose behalf the message is sent must be identified. The end-users must be informed about how to exercise their right to withdraw their consent for receiving such messages (Art.16)
When in force, the ePrivacy Regulation will apply to DataVaults, and it is advisable to remain vigilant and follow the developments, in order to develop the characteristics of its services and technological components in accordance with the draft Regulation. However, it is still possible that the draft will be modified or withdrawn, while is not clear when it will enter into force.

### 2.7.4 Proposal for a Directive on measures for a high common level of cybersecurity across the Union

It is also noteworthy to follow the regulatory developments in the area of security. In fact, on 16 December 2020, the EC adopted a proposal for a Directive on measures for a high common level of cybersecurity across the Union, repealing Directive (EU) 2016/1148. This proposal is directed to introduce systemic and structural changes to the current NIS Directive for covering a wider set of entities across the Union, with stronger security measures, such as mandatory risk management, minimum standards and relevant supervision and enforcement provisions. As highlighted by the European Data Protection Supervisor[26], it is essential to integrate "the privacy and data protection perspective in the cybersecurity measures stemming from the Proposal or from other cybersecurity initiatives of the Strategy in order to ensure a holistic approach and enable synergies when managing cybersecurity and protecting the personal information they process", and that "all cybersecurity systems and services involved in the prevention, detection, and response to cyber threats should be compliant with the current privacy and data protection framework".
In parallel, the EC and the High Representative of the Union for Foreign Affairs and Security Policy issued a Joint Communication titled "The EU's Cybersecurity Strategy for the Digital Decade", whose overall objective is to ensure a global and open internet with strong safeguards for the risks to security and the fundamental rights, in a multi-stakeholder model.

[26] European Data Protection Supervisor, "Opinion 5/2021 on the Cybersecurity Strategy and the NIS 2.0 Directive", 2021.

# 3 UPDATE OF THE FACTUAL BASIS FOR THE LEGAL AND ETHICAL ANALYSIS AND FOR THE REQUIREMENTS ELICITATION

## 3.1 DATAVAULTS DATA MANAGEMENT AND SERVICES

This section provides an update to the facts and aspects of the project which have been presented in deliverable D2.1 which were used to provide the legal analysis and to elicit the legal and ethical requirements, dwelling upon the privacy-relevant properties and personal data collection/processing/sharing in the main services and tools, as well as details upon the data categories, data sources and purposes of processing.

The overall data management and services to be offered by the overall infrastructure which are relevant to safeguarding security and privacy, are part of the first version of the DataVaults architecture has been developed under T5.2 and is shown in the next figure.

This architecture was designed having in mind the analysis conducted in WP2 regarding the security, privacy and ethical requirements and the overall technological developments and design work that has been carried out in the other project's WPs.

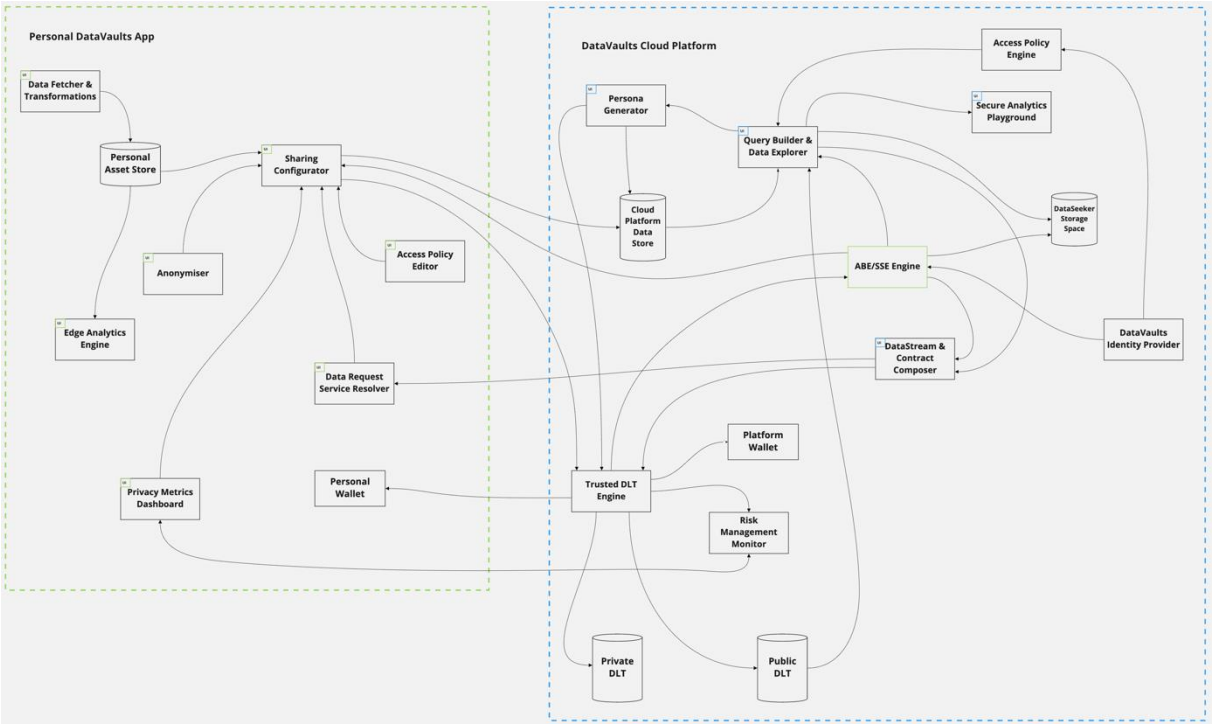


**Figure 1 –** DataVaults Architecture v1

As such, comparing this architecture with the conceptual architecture which was used in D2.1 reveals little difference (from a technical perspective), as the intention of the consortium was to refine the envisaged in the DoA architecture as necessary, to cover all the needs that have been expressed by the different stakeholders, and in the same time not to the main security an privacy principles which have been set by WP2.

### 3.1.1 Technical Components

#### *3.1.1.1 The DataVaults Cloud Platform*

The security by design architectural blueprint of the DataVaults Cloud Platform (the eastbound part of the overall architecture) is presented in the following figure, where colours are used to annotate the various components that facilitate privacy and security guarantees when it comes to data.

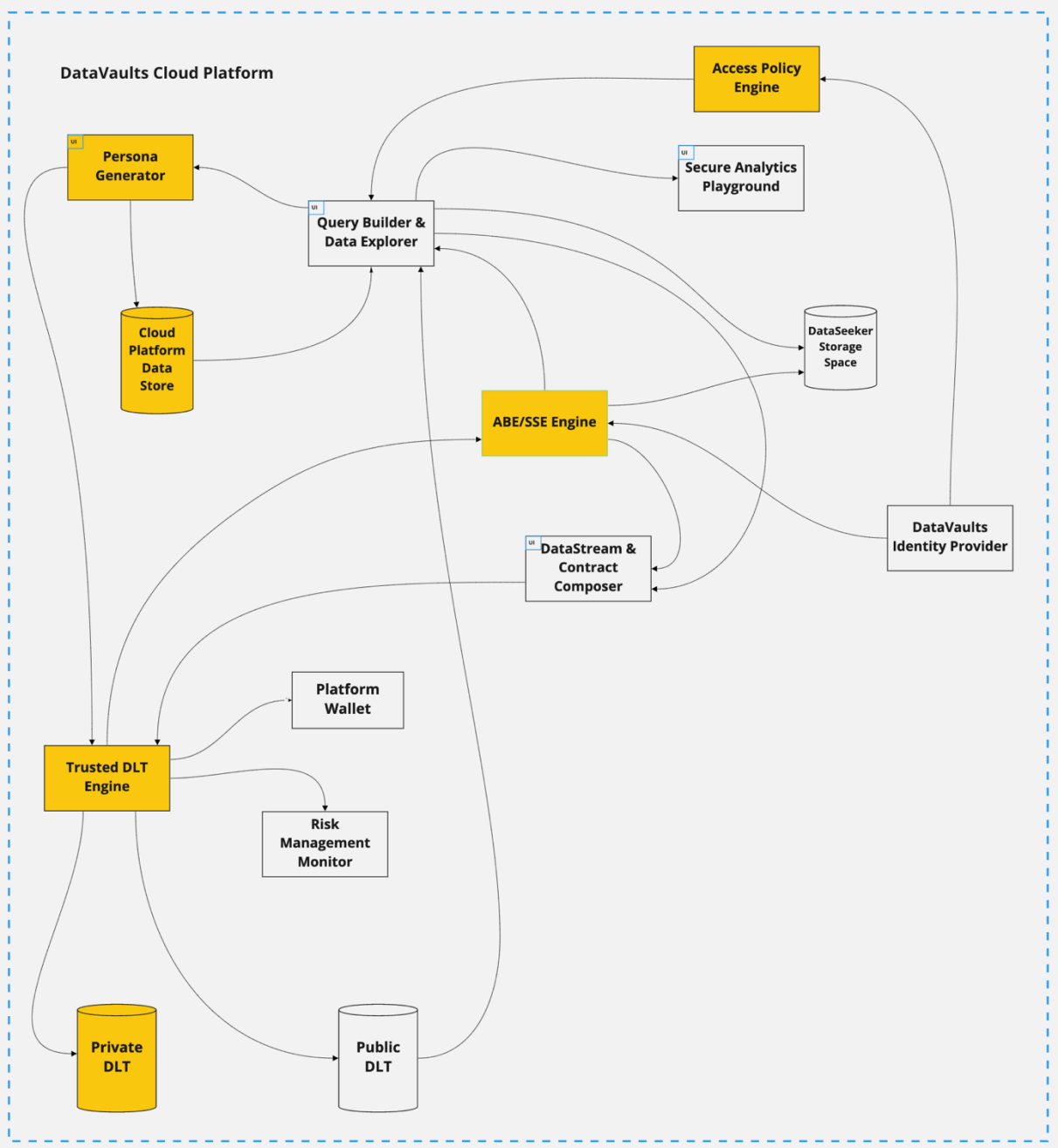


**Figure 2 -** Security and Privacy relevant components highlighted in the architectural blueprint of the DataVaults Cloud Platform part

As shown in the annotated version of the first version of the architectural figure above, the components that are relevant to the handling of personal data at the DataVaults Cloud Platform are the following:

- The **Access Policy Engine**. This is an infrastructure that handles access to the data based on the attributes that are described in the data contracts signed between the Data Owners and the Data Seekers. These attributes are selected by a Data Owner and are evaluated against the attributes of a Data Seeker which are coming from the platform's identify provider.
- The **Persona Generator Engine** (new addition from D2.1). Used merging together data from similar data owners, to generate personas which in order to guarantee the privacy of the data owners and non-traceability
- The **ABE/SSE Engine** (new addition from D2.1). This is used for disclosing the personal data to data seekers which have acquired the data based on smart contract, by unencrypting the data (ABE Engine) and for performing queries over the encrypted data, without disclosing the whole of the dataset to data seekers who have not a valid contract (SSE Engine)
- The **Trusted DLT Engine** and its **Private Ledger** (new addition from D2.1). This component is used to store the sharing configuration files of the Data Owner and execute the different data sharing transactions, without disclosing the real identity of each Data Owner and safeguarding that no changes to the way data are shared can happen, due to the immutability of the ledger.
- The **Cloud Platform Data Store** (update from D2.1), which includes the "**Secure Storage Containers**" and the "**Encrypted Searchable Data Lake**" (not shown explicitly in the architecture). This is an encrypted data storage facility that is used to store, in an encrypted manner the data which are shared by the Data Owners/Individuals. For technical implementation reasons, the concept of having one secure storage container per data owner is at the moment not considered, however all the data which will be necessary to be secured on the platform will be encrypted, providing in that manner the security and privacy guarantees required.

It is noted that the "Anonymiser" component, part of which was placed in the cloud platform side in D2.1, is now moved to the Personal DataVaults app side (see next sub-section)

#### *3.1.1.2 Personal DataVaults App*

As discussed previously, the Personal Data App of DataVaults is a core component of the overall architecture which is tasked with the collection of the personal data of individuals and is operated at the premise/side of each individual.

The following figure presents the main components (annotated with colours) which are relevant for safeguarding the privacy and the security of the Data Owner.

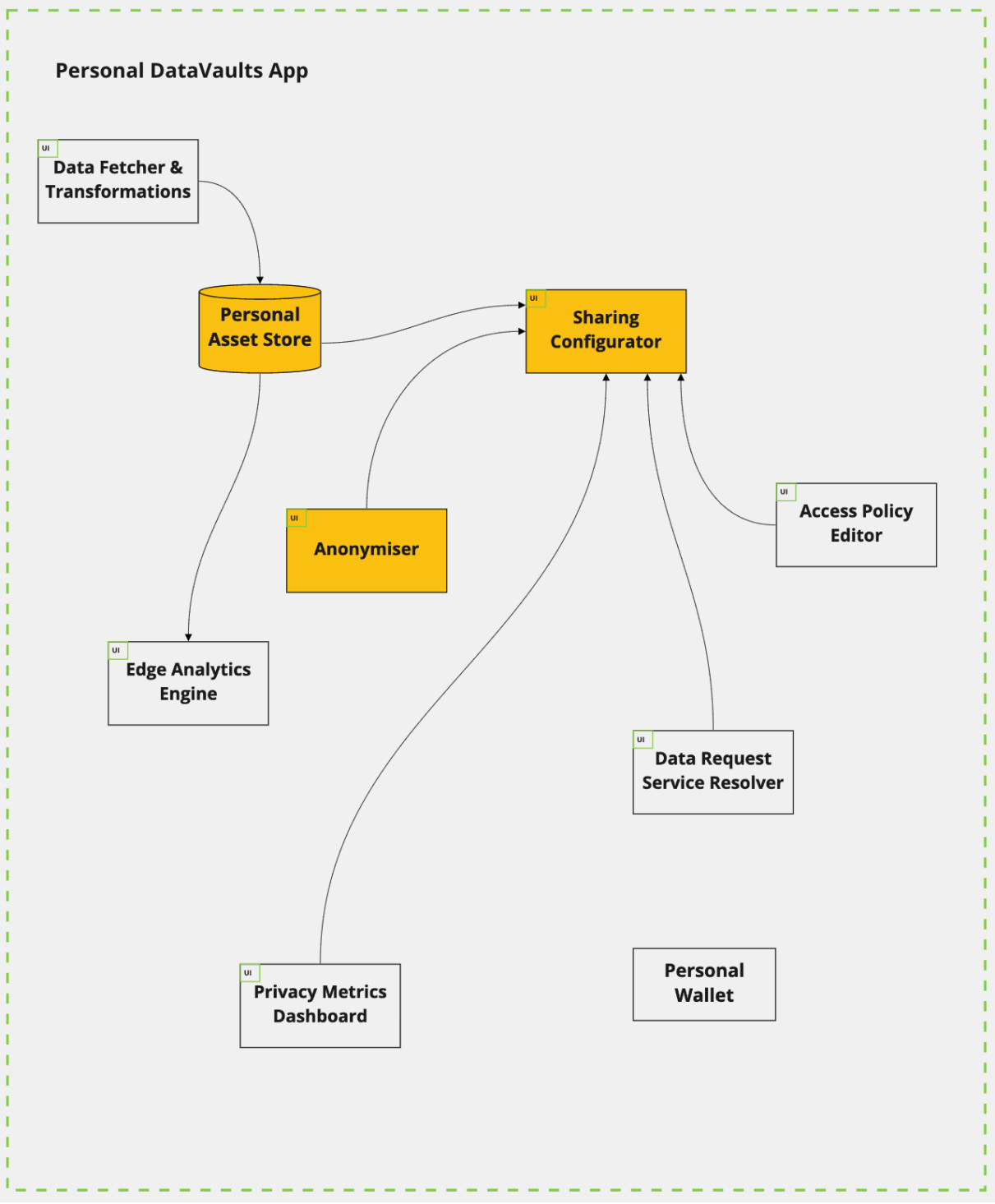


**Figure 3** - Security and Privacy relevant components highlighted in the architectural blueprint of the Personal DataVaults App part

The Personal Data App, includes the following services and components that are relevant to the handling of personal data:

- The **Personal Asset Storage**. This is the local storage container for the Personal DataVaults App where data resides, in an encrypted or not state.
- The **Anonymiser Engine**. A component used for manipulating data at the user's side, for uploading anonymous data to the core data platform

- The **Sharing Configurator** (update from D2.1),, which includes the **TPM DAA module** (as described in D2.1) and the BlockChain StarterGo Kit component (see deliverable D5.2), with the former used to infrastructure, based on TPM technology that allows the Personal DataVaults App to be attested the Personal DataVaults App towards the DataVaults Cloud Platform, and the second one for digitally signing the transactions to be performed, thus both providing trust and security guarantees relevant to the personal data.

### 3.1.2 High-Level Data in DataVaults

Regarding the data sources, the types of personal data and the format of such data, no changes have been performed in the overall concept or architecture which would impose changes to ones provided in deliverable D2.1

### 3.1.3 Data Subjects and other actors

Regarding data subjects and the other actor involved, both the evolution of the pilot scenarios, as well as the technical infrastructure elements developed in the previous period do not alter the already identified deliverable D2.1 subjects

For completeness reasons, the identified roles are presented in the following table.

**Table 1. Demonstrators and actors**

| Demonstrator | Data subjects | Controllers | Processors | Recipients |
|---|---|---|---|---|
| **OLYMPIACOS Demonstrator – Sports and activity personal data** | Club Members, fans and athletes | Olympiacos | New market segmentations and marketing campaigns companies, if any | Sponsors, NGOs, Federations, and any entity that asks for data to the controller |
| **PIRAEUS Demonstrator – Strengthening entrepreneurship and mobility** | Citizens, visitors | Local municipal authorities | Local authorities (transport departments). | Entities dedicated to cultural activities as museums. Olympiacos could act in this case, as the receptor of the data, local entrepreneurship associations |

| Demonstrator | Data subjects | Controllers | Processors | Recipients |
|---|---|---|---|---|
| **ANDAMAN7 Demonstrator – Healthcare data retention and sharing** | Patients/App Users | Andaman7 | doctors, hospital | third parties in the health sector (e.g.: clinical trial, research) |
| **MIWENERGIA Demonstrator – Smart home personal energy data** | users/customers | MIWENERGIA | other companies | other companies to offer services |
| **PRATO Demonstrator – Personal data for municipal services and the tourism industry** | citizens of Prato | Prato Municipality jointly with the Textile Museum and CGIL-CAAF fiscal support centre | Prato Mobility Office, Palazzo Pretorio Museum | City services and institutions, tourism companies and guide, third party aggregators and others |

### 3.1.4 The DataVaults Data Life Cycle: collection, processing, storage, sharing personal data and derivatives

The DataVaults MVP that has been developed and finalised in WP1 (see deliverable D1.4), has verified the overall data lifecycle (shown in the next figure) that was envisaged in the DoA and no changes are applied to this workflow, remaining the same as described in D2.1

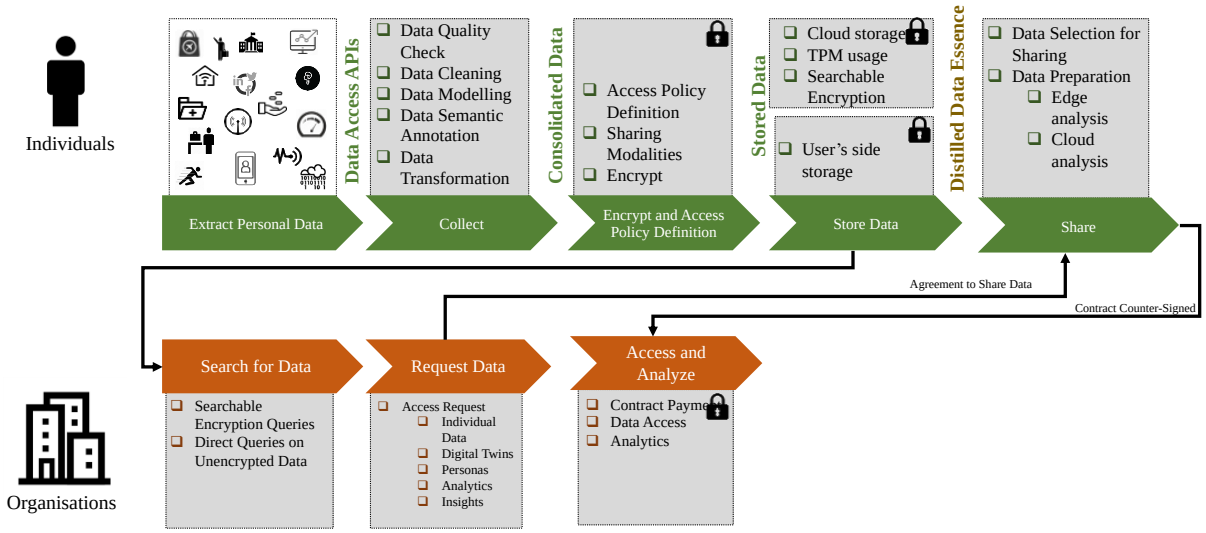


**Figure 4** - High-level overview of the data life cycle within DataVaults

## 3.2 DataVaults Demonstrators

### *3.2.1* Demonstrator #1 – Sports and Activity Personal Data

There are no changes for the scenarios, or relevant changes in the legal and ethics situation affecting this demonstrator.

**Scenario A.** Current users of Olympiacos will be able to connect to DataVaults to store all or part of their personal data (after explicit consent). This storage can be used as a backup to retrieve data when lost. This can also be used anonymized and unencrypted by different organizations such as sponsors/NGOs/Federations/Local authorities who want to run a campaign/host an event for the club members/fans. Users will be able select what services they want to subscribe to and what kind of data will be shared.

**Scenario B.** Athletes will be able to connect to DataVaults to collect their data (coming from various sources such as training reports and medical exams) and store them in DataVaults on their smartphone. This will make the data available to the doctors, coaches and trainers so as to adapt their strategies and plans based on them covering the athlete's expectations and offering the appropriate medical and sport equipment. As regards privacy and data protection, an initial remark is that, before obtaining any information, first the participants must sign a consent form assuring they know about the main objectives of the project, how the data are going to be processed and their rights. Regarding privacy and data protection the users will have always the ownership and right to decide about them.

Modifications on these scenarios maybe necessary upon the finalization of the DataVaults platform and the approval of the available through the platform data.

### 3.2.2 Demonstrator #2 – Strengthening Entrepreneurship and Mobility

There are no significant changes for the scenarios, or relevant changes in the legal and ethics situation affecting this demonstrator, at this stage. The scenarios are as follows:

**Scenario A.** Smart mobility services for individuals. This scenario will engage both OLYMPIACOS and PIRAEUS and will use GPS route data, shared by citizens as well as by the members taking part in the OLYMPIACOS demonstrator, to better schedule the mobility strategy and the relevant services within the city. The data will be collected through the DATAVAULTS platform/app. The area of interest will be the surroundings of the OLYMPIACOS sport venues.

**Scenario B.** Empowering local entrepreneurship. In this scenario, the data to be provided by the DataVaults users will be used to better understand consumer behaviours and preferences, with the aim to strengthen the local economy through activities that can be brought forward by the Municipality. Moreover, Piraeus will invite local entrepreneurship associations (i.e., the Piraeus Traders Association) and other interested stakeholders to either join the platform or act as 2nd tier data seekers, to test the aspects of the project that have to do with value generation and sharing with entities not directly using personal data but that access the derivatives of the latter. This scenario meets the on-going activities of PIRAEUS about the city's Open Trade Centre associated, *inter alia*, to the improvement of the local economy through restructuring of the market infrastructures and the deployment of smart applications.

**Scenario C.** Services for personalized cultural and touristic experiences. This scenario will build on data analyzed from the profiles and preferences of the DataVaults platform/app users and from data provide by an application of the Municipality of Piraeus called "Pireapp", in order to create services that target tourists and citizens visiting the city of Piraeus. During this scenario, the data to be analyzed will generate reports that will assist the departments of the Municipality to better design their strategies regarding the services offered to meet the touristic and cultural event demand. This scenario is both aligned and complementary to the Digital Strategy20 of PIRAEUS in terms of implementing an integrated Destination Management System, engaging citizens and visitors in the interactive definition of the cultural content of interest through the analysis of public (i.e., museums & touristic organisations) and private (i.e., travel agencies, cruise operators, booking organizations, etc.) data sources.

Modifications to the scenarios maybe necessary upon the finalization of the DATAVAULTS platform/app, the initial discussions between the Municipality of Piraeus and UBITECH (the partners responsible for these pilots) and the finalization of the available through the platform data.

### 3.2.3 Demonstrator #3 – Healthcare Data Retention and Sharing

Some precisions were added to the scenario (a) that was described in D2.1: The Andaman7 platform can also act as a data seeker for third party companies. This will give us the ability to collect some client specific data entered in Andaman7.

Scenario (b) has not changed.

There are no significant changes from a legal and ethical perspective. Treatment of health data still falls into GDPR sensitive data which means that we still need robust data protection safeguards and explicit consent of the source to exchange, store or process such data.

The Belgian Law of 30 July 2018 still applies which means that we should also:

- indicate which categories of persons have access to the data and explain their relation to the processing of the personal data
- maintain a list of these categories of persons for the Belgian data protection authority
- ensure that the designated persons are subject to a legal, statutory or equal contractual obligation to maintain the confidential character of the personal data.

However, as the Andaman7 platform can act as an intermediary service provider between the data owner and the final data seeker, in scenario (a), some precisions must be added:

- As stated in GDPR, the consent should mention the service provider and the final data seeker.
- Both of them should be compliant with GDPR and agree to terms specified in the consent.
- They also should be located in a country that is recognized by the European Commission as providing adequate protection.

### 3.2.4 Demonstrator #4 – Smart home Personal Energy Data

There are no changes for the scenarios, or relevant changes in the legal and ethics situation affecting this demonstrator.

There is a change in the electric bills structure in Spain, which is planned for June 2021.This is made by the government, changing the price for each energy period and the calendar governing those periods. This change will not introduce modification in the data that are going to be uploaded to DataVaults through our API.

Furthermore, the amended tariff structure will not introduce a change in data protection and ethical considerations, so we consider that it is not relevant for the project, either for the deliverable.

This legal modification introduces more importance to have the data of energy consumption for developing properly our scenario a, enhancing the utility of it. But it is not a variation of the scenario.

If this “frame” modification is relevant to be described in the deliverable we can do it quickly.

### 3.2.5 Demonstrator #5 Personal data for municipal services and the tourism industry

The original scenarios reported in D2.1 have been slightly changed to accommodate the pilot actions inside the more consolidated technological approach proposed in the project. Three different scenarios have been drafted, as reported below.

**Scenario 1. Access to personal data for the analysis of mobility solutions**

In this scenario, the Mobility Office acts as a Data Seeker and can access the DataVaults platform to look for citizens’ personal data in order to accomplish different types of activities:

1. to plan mobility solutions in the city through the access to personal data provided by citizens (position, means of transportation, itineraries, etc.),

2. to identify adequate samples of citizens for the sending of surveys on traffic and mobility.

In case of activity 1, the Office might require accessing personal data, like profiles including mean of transportation and GPS position, shared by citizens in the "anonymous" or "persona" form, and available for Data Seekers classified as "public administration". Once the Mobility Office has downloaded the requested "persona"/"digital twin" datasets (profile, localisation, means of transportation, etc.), they can use the visualisation/analysis tools provided by the DataVaults platform to support the creation/revision of the mobility plans in the city. Data collected with the DataVaults platform can be also integrated with other data already owned by the Office.

In case of activity 2, the Office will require "eponymous" personal data shared by citizens, since this option will allow to build more specific citizens' samples to whom address questionnaires/surveys focused on mobility issues. Alternatively, the Office could require an anonymous selection of citizens and have the platform directly managing the sending of the survey.

In all cases, a compensation for citizen is issued by the DataVaults platform on the basis of the specific contract stipulated between the platform and the Data Seeker, by taking the data access policy set by the data owner into account.

**Scenario 2. Access to personal data for the improvement of cultural offer in the city**

In this scenario, a Cultural Institution in the city acts as a Data Seeker and can access the DataVaults platform to look for citizens'/visitors' personal data in order to accomplish different types of activities:

1. to carry out data analysis for the improvement of its cultural offer,
2. to define adequate samples of citizens/visitors for the sending of surveys/market campaigns.

In case of activity 1, the Cultural Institution might require data like personal profile including cultural preferences and GPS position, that citizens/visitors have shared in the "anonymous"/"persona" form, allowing access to data seekers like cultural institutions. Once the Cultural Institution has downloaded the requested "persona"/"digital twin" datasets (profile, localisation, cultural interests, etc.), they can use the visualisation/analysis tools provided by the DataVaults platform to support the creation/revision of plans for the improvement of its cultural offering in the city. Data collected with the DataVaults platform can be integrated also with other data already owned by the Institution.

In case of activity 2, the Cultural Institution will require "eponymous" data, to identify the most suitable samples of citizens/visitors to whom address more tailored surveys on its cultural offer. Alternatively, the Office could require an anonymous selection of citizens and have the platform directly managing the sending of the survey.

In all cases a compensation for citizen is issued by the DataVaults platform on the basis of the specific contract stipulated between the platform and the Data Seeker, by taking the data access policy set by the data owner into account.

**Scenario 3. Access to personal data for the delivery of personal certificates**

In this scenario, a third-party requiring citizen's certificates (e.g., estate agency, fiscal support centre, utilities, banks, etc.) accesses the DataVaults platform as a Data Seeker, with the objective of acquiring personal documents shared by the citizen as "eponymous" data, to conclude a given process started by the data owner or to get update for its own databases to provide services. Although the certificates are managed by the administration, the document exchange is carried out in a totally transparent way without the administration being aware of it and this makes the whole process more compliant with the GDPR legislation. Moreover, the automatic, authorised regular downloading of personal certificates by third parties reduces burdens for citizens and interested data seekers.

**Data collection and management**

The pilot scenarios foresee the collection of different types of personal data, such as preferences on social networks, cultural interests, participation in cultural events, geolocation, preferences on mobility and civil certificates. Some of the data will be collected by data owners through a connection to the population registry and CRM of the Municipality of Prato, while other will be extracted from other sources like social networks and user's smartphone (e.g., geolocalisation systems). Eventually, other information, like for example cultural interests, might be provided directly by data owners on request from data seekers.

Adequate security and privacy-preserving measures for storage and handling of such data will be adopted, using state-of-the-art technologies for secure storage, delivery and access of personal information, as well as managing the rights of the users. In particular, security will be guaranteed by the implementation of a specific blockchain structure that includes access control, the composition of the management of the fiduciary consent, the authentication of participation and the preservation of privacy. Users will be able to define the configuration of the conditions for sharing their personal data, which will be represented by specific smart contracts managed by the DataVaults platform.

Data collection and management procedures will be in accordance with the national and European legislation framework that was described in D2.1.

# 4 LEGAL, ETHICAL, SECURITY, PRIVACY AND TRUST REQUIREMENTS

## 4.1 LEGAL AND ETHICAL REQUIREMENTS

The following table sets the legal and ethical requirements for the design, development and validation of DataVaults cloud-based platform and Personal App, as well as, to some extent, for the future operation of them, clearly laying out a first guideline for legal compliance and ethically-sound activities and results, without forgetting checkpoints. This requirement list reflects the project's progress, including the better shaping of its services, solutions and demonstrators including their privacy-relevant properties and personal data collection, processing and sharing features, as well as considerations on the data categories, data sources and purposes of processing. On the other hand, the list takes into account the enriched legal review, where additional areas of law were analyzed, as well as the regulatory reforms under development and their accompanying documents have been studied.

As in the initial release if the requirements, the elicitation was guided by a systematic and holistic approach, driven by Fairness & Privacy-by-Design-and-by-Default enriched with the Protection Goals method[27].

The requirements, though in some case binding (when directly deriving from the legislation, such as GDPR), in some cases are quite challenging and need to be interpreted taking into account the SoTA, the research nature of the project and the risk-based approach fostered by GDPR itself. This demands for a certain degree of flexibility in the assessment of the adequateness of measures and technological solutions, to be specifically established on a case-by-case basis, considering a set of circumstances rotating around the severity of the risks and the reasonable efforts to face with them. In addition, in other cases, where not directly imposed by the legislation, the requirements have to be interpreted more than recommendations or preferable requirements. This is clearly stated in the description of each of them.

**Table 2. Legal and Ethical Requirements.**

| N. | Short name | Description | Supporting DataVaults Tool | Phase | Nature |
|---|---|---|---|---|---|
| **R1** | **Fairness and Lawfulness** | Fairness can be explained through the concepts of loyalty and good faith to be respected in all the steps of any personal data processing. The lawfulness implies that the data processing should be performed according not only to applicable data protection legislation, but also to any other applicable law and regulation, including provisions that other than legislative acts from a strict legal interpretation. GDPR itself (art. 6) lays down legal bases on which the lawfulness of processing relies. Fairness obligations are also required by the P2BR for the intermediation services (platforms), though in the different meaning of settlement of effective out-of-court redress mechanisms such (as internal handling systems for business users) and mediation procedures. | The whole system and app | All | PDPL, HRs, ESL, P2BR |

[27] More details on this can be found in the requirement list itself (under R15), and in the Section 4.2.1, under R8.

| N. | Short name | Description | Supporting DataVaults Tool | Phase | Nature |
|---|---|---|---|---|---|
| **R2** | **Purpose limitation and legitimate aim** | This principle requires that i) DataVaults technologies serve a specific, explicit and legitimate purpose; ii) the data have to be collected for such a purpose and not further processed in a way incompatible with it; iii) adequate safeguards against misuse have to be taken. | The whole system and app | All | PDPL, ESL |
| **R3** | **Data minimisation** | DataVaults must embed in its developments tools and measures to comply with the data minimization principles. According to art. 5 GDPR, personal data shall be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed". The benefit potentially arising from the use of that kind of data has to be clear. This principle also requires to adopt anonymization and pseudonymization that can be invoked by the data owner, including adopting safeguards for mitigating the risks of re-identifying the individuals and for minimising possible linkability and actual linkages. | Core DataVaults platform: Access Policy Engine, Risk Management Monitor, Anonymizer Engine Personal Data App: Risk Privacy Metrics Dashboard Access Policy Engine, Anonymizer Engine, TPM DAA module | All | PDPL, ESL |
| **R4** | **Data Accuracy** | "Personal data shall be…accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay" (Article 5, letter d GDPR). This principle is connected with the data quality and trust, as well with the data security and integrity and with the technical and organization measures that need to be taken. | Core DataVaults platform: Secure Storage Containers Personal Data App: Secure Storage facility, Data Feeder and Transformation, TPM DAA module | All | PDPL, ESL, ITSL |
| **R5** | **Integrity and Confidentiality** | Personal data must be protected with appropriate controls to ensure the integrity, confidentiality and availability of the data. Personal data shall be "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures" (Article 5, letter f GDPR). | Core DataVaults platform: Encryption/Decryption Engine, Secure Storage Containers, Access Policy Engine. Personal Data App: Encryption/Decryption Engine, Secure Storage facility, Access Policy Engine. | All | PDPL, ITSL, RFSJ |
| **R6** | **Storage Limitation** | The storage limitation requirement is set forth in Art. 5 (1) (e) GDPR, requiring that personal data must either be erased or anonymised as soon as it is no longer necessary for the purpose to identify the natural person. As regards the data processing in the demonstrators, this requirement will have limited application due to the privilege for scientific research, for which personal data may be retained | Core DataVaults platform: Secure Storage Containers. Personal Data App: Secure Storage facility. | R, Ex | PDPL, RFSJ |
| **R7** | **Transparency** | The personal data processing in DataVaults must be inspired to full transparency, functional to grant an adequate level of clarity of it, including all privacy-relevant properties and actions. The information to | Core DataVaults platform: Access Policy Engine. Personal Data | All | PDPL, ESL, RFSJ, |

| N. | Short name | Description | Supporting DataVaults Tool | Phase | Nature |
|---|---|---|---|---|---|
| | | the data subject is fairly considered as one of the fundamental rules of a lawful personal data processing and enables the data subject to correctly enforce his/her rights under the GDPR: in other words, the adequate level of transparency is a prerequisite for all kinds of control and intervention. The minimum list of mandatory information to be provided with the data subject are listed in GDPR (Art. 13).<br>On the other hand, under P2BR transparency obligations are foreseen for providers of intermediation services to inform, through clear, unambiguous and readily available contractual terms and conditions, about the treatment, the criteria used to rank their products and the requirements to suspend or terminate their services.<br>In addition, also the e-Commerce Directive imposes information obligations for the conclusion of a contract with a consumer and liabilities in relation to them (Sect. 4). | App: Access Policy Engine, TPM DAA module | | P2BR, ECD |
| **R8** | **Privacy and Data Protection by Design and Privacy by Default** | Privacy-by-design and by default need to be in the focus of attention within DataVaults. Art. 25 GDPR expressly sets forth that, considering the set of circumstances, the controller shall implement, appropriate technical and organisational measures: "such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects"; "for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility". | The whole system and in particular: Encryption/Decryption Engine, Secure Storage Containers, Anonymizer, Access Policy Engine | All | PDPL, ESL, ITSL |
| **R9** | **Avoidance of discrimination (including social sorting) and of harm** | In line with the European Charter of Fundamental Rights, which prohibits any kind of discrimination (Article 21), in DataVaults efforts should be directed to avoid that the overall system architecture and/or the demonstrators facilitate any kind of discrimination (race, gender, age, religion, disabled) or social sorting, as well as to cause undue or unjustified harm to anyone, including wrongfully stigmatisation. This is also aligned with the recommendations set forth in the position papers and other soft law instruments promoted by the EC (such as those of the Big Data Value Association). | The whole system | All | PDPL, HRs, ESL |
| **R10** | **Informed Consent** | The GDPR (Article 4) defines the "consent of the data subject" as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her". DataVaults must have a very strong focus on the consent requirement set forth by the GDPR, aiming at implementing consent processes capable of enabling a much better control of individuals over their personal data, taking into account | The Personal data app | All | PDPL, ESL, HRs |

| N. | Short name | Description | Supporting DataVaults Tool | Phase | Nature |
|---|---|---|---|---|---|
| | | privacy-by-design and by default in relation to this, as well as the data subjects' rights and corresponding obligations of data controllers and processors. The data subject's informed, explicit and free given consent is one of the criteria for rendering the data processing legitimate. | | | |
| **R11** | **Set of requirements referring to the voluntary participation to DataVaults demonstrators** | The following requirements apply to DataVaults demonstrators: i) Recruitment Procedures for the selection of the voluntary participants for the piloting operations have to be set and followed, in order to avoid any sort of discrimination/social sorting. These procedures need to be assessed by the Ethics Advisory Board of the project; ii) informed consent has to be obtained: the pilot partners must inform voluntaries and distribute the consent form, to be signed by each voluntary before the piloting operations start; iii) volunteers' dignity has to be safeguard and direct/indirect incentives for participation must not affect it. | The Personal data app | D | PDPL, HRs, ESL, RFSJ |
| **R12** | **User Control** | DataVaults must concretely ensure to individuals to retain and exercise real control over their personal information. User control is required not only by GDPR, but also by the upcoming ePrivacy Regulation (ePR). | Personal data app: Privacy Metrics Dashboard, Access Policy Editor, Identities Wallet | All | PDPL, HRs, ESL |
| **R13** | **Data subject's rights** | In DataVaults the data subjects must be effectively entitled to exercise a range of rights, specifically laid down in the Articles 12 –22 GDPR, including:<br>- Transparent communication (Art. 12 GDPR);<br>- Information on the controller's identity and the processing itself, including the means and purposes of the processing. There are two cases: personal data collected from the data subject (Art. 13 GDPR) and personal data not obtained from the data subject (Art. 14 GDPR);<br>- Right of access (Art. 15 GDPR);<br>- Right to rectification of inaccurate data (Art. 16 GDPR);<br>- Right to erasure, 'right to be forgotten' (Art. 17 GDPR);<br>- Right to restriction of processing (Art. 18 GDPR);<br>- Right to receive a notification from the controller regarding rectification or erasure of personal data or restriction of processing (Art. 19 GDPR);<br>- Right to data portability (Art. 20 GDPR);<br>- Right to object (Art. 21 GDPR);<br>- Protection against automated decision-making, including profiling (Art. 22 GDPR). | Personal data app: Privacy Metrics Dashboard, Access Policy Editor, Identities Wallet, Data Request Resolver, Data Picker<br><br>Core DataVaults platform: Access Policy Engine, DataVaults Private Brokerage Engine | All | PDPL, ESL |
| **R14** | **Enforcement** | DataVaults smart contract should be developed as flexible and pragmatic solutions, capable of providing certainty, predictability, auditability, and ease of enforcement not only to contractual provisions, but also to data protection legislation via enabling technological tools. DataVaults system should not only create tools giving people | Core DataVaults platform: Open Ledge, DataVaults Open Brokerage Engine, Contract Composer, | All | PDPL, ITSL, RFSJ |

| N. | Short name | Description | Supporting DataVaults Tool | Phase | Nature |
|---|---|---|---|---|---|
| | | ownership of the data which they and the devices they own generate, but it is also recommended to start considering steps forward towards for the enforcement of data subjects' rights and, in general, of the GDPR rules, besides the usual data policies (use limitation, flow control, data transfer restrictions, etc.). | Private Ledger, DataVaults Private Brokerage Engine, Access Policy Engine | | |
| **R15** | **Fairness by Design** | DataVaults technology needs to be conceived and developed following the fairness by design approach, in order to ensure that individuals' privacy and real control over their data is afforded to it. Both the substantive and the procedural dimension of fairness are deemed necessary. | The whole system | All | ESL |
| **R16** | **Effective "sharing the wealth" paradigm** | DataVaults should deliver a personal data framework and platform capable of offering benefits to all the stakeholders involved (citizens, businesses, governments, research world, civil society organisations, etc.) and of adhering to the European values, e.g., democracy, privacy, safeguards and equal opportunities. Thereby it should be consistent with the win-win paradigm, promoted by the soft law and, in primis, by the EC and its PPPs such as that with BDVA. | The whole system | All | ESL |
| **R17** | **Privacy Notice** | According to GDPR, a set of information have to be provided to the data subjects, both in case the personal data are collected from the data subject (Art. 13), and in case personal data have not been obtained from the data subject (Art. 14).<br>In relation to DataVaults, it is important to refer to Art. 13 which mention, among others, the following information to be provided i) the identity and the contact details of the controller, ii) the contact details of the data protection officer, where applicable; iii) the purposes of the processing and the legal basis; iv) the recipients or categories of recipients of the personal data, if any; v) if applicable, the intention to realize transfer personal data to a third country; vi) data storage; vii) data subjects' rights viii) the existence of automated decision-making, including profiling, and ix) the secondary use.<br>It is also recommendable to define and use adequate Privacy Metrics, easy to understand from any individual, including non-expert. They should consider, as a parameter, the availability and kind of input data for each context and use case. The individual should be informed of the privacy risk. | The Personal data app; Privacy Metrics Dashboard | All | PDPL, HRs, ESL, RFSJ |
| **R18** | **Data breaches** | Mechanisms should be established in DataVaults to ensure that, in case of personal data breach and if it is likely to result in a risk to the rights and freedoms of natural persons, the notification requirement set forth by Art. 33 and 34 GDPR can be fulfilled. However, the legislator sets a number of exceptions that need to be considered as well. The notification has to be done to the individuals concerns and to the supervisory authorities (with undue delay, and, if feasible, within 72 hours). As for the data owners, | The Personal data app | All | PDPL, ESL |

| N. | Short name | Description | Supporting DataVaults Tool | Phase | Nature |
|---|---|---|---|---|---|
| | | it could be explored a notification mechanism through DataVaults personal data app itself. | | | |
| **R19** | **Accountability** | The principle of accountability requires organisations to be compliant with GDPR and to be able to demonstrate compliance: "the controller shall be responsible for and be able to demonstrate compliance with". DataVaults is recommended, therefore, to provide the tools for respecting the accountability principle and the documentation requirement, including documenting the legal basis, the purposes and the means of a specific processing operation types (e.g., in an index of procedures describing the processing operations in conjunction with the technical and organisational circumstances) along the entire value chain.<br>DataVaults technology is recommended to support the documentation and demonstration of compliance with all privacy-related policies, procedures and practices in various ways. | The DataVaults Operations manual | All | PDPL, ESL, RFSJ |
| **R20** | **Record of processing activities** | DataVaults solution is recommended to provide the tools for complying with the obligations set forth by GDPR, Art. 30: "Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility" specifying also the information that has to be contained in the recording. | The DataVaults Operations manual | All | PDPL, RFSJ |
| **R21** | **Data Protection Impact Assessment** | In case it is likely that the data processing in DataVaults results in "a high risk to the rights and freedoms of natural persons" a Data Protection Impact assessment, pursuant to Art. 35 GDPR (and D10.2: POPD - Requirement No. 2) will be carried out, to evaluate the impact of the envisaged operations on the protection of personal data.<br>As for DataVaults demonstrators, it has to be remarked that, according to Art. 35, c. 4, 5 and 6, the competent National Data Protection Authority for each of the countries involved could have established a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment. They must consult their respective DPO on this aspect, also taking into account the differences from common situations, due to the research purposes of the data processing in DataVaults.<br>DataVaults is strongly committed to operationalize the risk-based approach encouraged by the GDPR and specific tools and services will be devoted to this. | Core DataVaults platform and personal data app: Risk Management Service and Risk Exposure Dashboard | D (and potentially R) | PDPL, ESL, RFSJ |
| **R22** | **Application scrutiny to local/national boards if required by national legislation concerned** | GDPR doesn't require a general notification requirement to the supervisory authorities. Such an obligation is required only for those types of processing operations "which are likely to result in a high risk to the rights and freedoms of natural persons by virtue of their nature, scope, context and purposes. Such types of processing operations may be those which in, particular, involve using new technologies, or are of a new kind and where no data protection impact assessment has been carried out before by the controller…" (Recital 89). DataVaults demonstrators have to take this clarification into account and consult their | N/A | D | PDPL, RFSJ |

| N. | Short name | Description | Supporting DataVaults Tool | Phase | Nature |
|---|---|---|---|---|---|
| | | respective DPO, to assess if the notification is necessary or not, bearing also in mind the differences that could arise from national legislation. | | | |
| **R23** | **International Data Transfer** | In the post-project phase, the DataVaults solutions could be used in a wide data sharing ecosystem, potentially including flows of personal data to and from countries outside the Union and international organisations. Therefore, though it is not expected to have an impact on the demonstrator activities in DataVaults, it is recommended that the design and development of the solution envisage also the case of transfers of personal data to Third Countries (or international organisations) and consider the provisions of the Chapter 5 of the GDPR.<br>Tools should be provided for addressing the related data protection challenges and concerns, and thus complying with Chapter 5 of the GDPR, ensuring that its level of protection of natural persons is not undermined in particular when personal data are transferred from the EU to controllers or other recipients in Third Countries (or international organisations).<br>Special attention should be given to Art. 44 and Art. 46, respectively setting forth the general principle for transfers and the transfers subject to appropriate safeguards. Also, Recital 101 should be addressed. | N/A, though the Access Policy Engine could be used for example to exclude data being server to entities outside the EU | E | PDPL, RFSJ |
| **R24** | **Technical and organizational measures** | GDPR requires that all controllers shall implement appropriate technical and organisational measures designed to implement data-protection principles in an effective manner (Art. 25) and to ensure a level of security appropriate to the risk (Art. 32). As for the design and development of the DataVaults solution, technical measures are particularly relevant. | The whole system and in particular: Secure Storage Containers, Encryption/Decryption Engine, Access Policy Engine | All | PDPL, ITSL, RFSJ |
| **R25** | **Use of private environment/ cloud as much as possible** | In order to retain bigger control of the data being processed, it is recommended to use private environment as much as possible for the storage or processing of personal data. This especially applies to the Personal Data App (in particular the Secure Storage facility), operated at the premise/side of each individual through the personal devices that will be host environments for this App. The recommendation is relevant also for the corresponding components of the Core DataVaults cloud-based platform, the Secure Storage Containers. | Core DataVaults platform: Secure Storage Containers Personal Data App: Secure Storage facility | All | PDPL, ESL, ITSL |
| **R26** | **User and data protection friendly User Interface** | DataVaults consortium must develop user and data protection friendly User Interface (UI), that should facilitate as much as possible the user control features. It should be capable of collecting consent and constraints/restrictions, providing appropriate options for user information and control, thereby enabling the data subject to easily consent and exercise his/her rights set forth under data protection legislation, at national and European level. | Personal Data App: Privacy Metrics Dashboard | All | PDPL, ESL, ITSL |

| N. | Short name | Description | Supporting DataVaults Tool | Phase | Nature |
|---|---|---|---|---|---|
| **R27.** | **Measures in case of profiling** | DataVaults foresees the use of personas, in the sense of fictional individuals sharing the same, but obfuscated characteristics of specific groups of individuals. To build the personas, anonymous data from similar individuals has to be grouped. Therefore, it has to be investigated whether this implies or not "profiling" in the meaning provided by GDPR and therefore whether Art. 22 is applicable. In such a case, if an automated-decision making occurs and it produces in some way relevant effects on the data subjects, this aspect should be covered by informed consent. Furthermore, the suitable measures (including from a technical point of view) to safeguard the data subject's rights and freedoms and legitimate interests, have to be taken, ensuring at least "the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision". | Personal Data App: Privacy Metrics Dashboard, Anonymiser, Identities Wallet<br><br>Core DataVaults platform: Anonymiser | All | PDPL, ESL, RFSJ |
| **R28** | **Appointment of Data Protection Officer** | The Consortium must appoint a DataVaults Data Protection Officer (DPO) among its Consortium members, for the handling and management of personal data in accordance with the existing provisions of GDPR and other relevant EU and national legislations. His/her responsibilities will be in line with Article 39 of the GDPR. | N/A | R | PDPL |
| **R29** | **Assignment of responsibilities** | In each of the demonstrators the data controller has to be identified, as well as the data processors and, in case, the data sub-processors). In relation to the role covered, each entity involved in the processing (data controller and data processor or sub-processor) is bound by obligations to be met and principles to be followed. These obligations are functional ensure that: i) the data processing conforms to privacy laws; and ii) the data subjects maintain the right to control what information is collected about them, how it is used, who has used it, who maintains it, and what purpose it is used for. Most duties and obligations are assigned to the data controller, who has the main responsibility for data, whilst the data processor has fewer and limited legal responsibility. It has to be pointed out that these roles are relevant also in relation to the design and development of DataVaults solutions, as well as for the post-project uptake. | N/A | All, but especially D | PDPL |
| **R30** | **Ethics Board set-up and involvement** | This requirement refers to the need to set-up and involve this committee to i) monitor ethical and legal issues in the project and report to the Commission; ii) work closely with the consortium in order to address the ethical and legal issues and data privacy concerns, that may arise from accessing user related information. It will periodically report to the Commission on the implementation of the ethical, legal and data protection issues in project and compliance with applicable national and EU regulations | N/A | R | ESL |
| **R31** | **eIDAS Obligations** | Electronic identification (eID) for the users to offer enhanced security and accuracy, swifter and less costly processes, while they may mitigate risk of fraud, identification theft and legal challenges. | Smart Contract, Overall System | All | eIDAS |

| N. | Short name | Description | Supporting DataVaults Tool | Phase | Nature |
|---|---|---|---|---|---|
| | | Obligations of to the providers of the non-qualified electronic registered delivery service to prevent and minimize the impact of security incidents or loss of integrity of its services (art. 5, art. 13, art. 15, art. 19). electronic identification for its users. | | | |

**Phases**: R: Research phase, D: Demonstration phase, E: Exploitation phase, All: all the phases, both during the project and after its end;

**Nature**: PDPL: Privacy and Data Protection Law; HRs: Human Rights Law; ITSL: Telecommunication Law and/or Information Technology-Security Law; ESL: Ethics and Soft Law; RFSJ: Regulatory Framework in the selected jurisdictions; P2BR: Platform-To-Business Regulation; ECD: e-Commerce Directive; eIDAS: eIDAS Regulation.

## 4.2 SECURITY, PRIVACY AND TRUST REQUIREMENTS

In this part, we describe the technical requirements (in terms of data security, user privacy and operational assurance), which have been clustered in mandatory and desirable ones. The split differentiates the requirements that are needed for the demonstrators within the DataVaults project, and the possible requirements of a secure and privacy-preserving data sharing environment in general. Thus, they are the mandatory requirements that will drive the design and development of the core security, privacy and trust services of DataVaults platform.

**Table 3. Mandatory requirements.**

| Number | Short Name | Description | ID | Supporting DataVaults Tool |
|---|---|---|---|---|
| **SR1** | Integrity and Confidentiality | Personal data must be protected with appropriate controls to ensure the integrity, confidentiality and availability of the data including on-chain and off-chain data. | M | Trusted Platform Module, DataVaults Distributed Ledgers, Secure Storage Facility |
| **SR2** | Authorization and Access Control | The participating users should act according to the security and privacy policies, related to data sharing preferences, specified and deployed via smart contracts. Only authorized users should have access to the platform, its components and the shared data. In case such policies need to be updated, during runtime (e.g., specification of different user roles for accessing specific data models), this should be reflected | M | DataVaults Identity Access Management |

| Number | Short Name | Description | ID | Supporting DataVaults Tool |
|---|---|---|---|---|
| | | through the deployment of new smart contracts. | | |
| **SR3** | Non-repudiation and Accountability of Actions | Actions should be non-repudiable, and all system entities should be held accountable of their actions. | M | DataVaults Crypto Suite (i.e., DAA, Signatures, Attestation, Verification) |
| **SR4** | User-controlled Anonymity | When anonymization is desired by the users (thus, empowering user controlled-anonymity), users including their devices and actions should not be identifiable without breaching the non-repudiation requirement of their actions (SR3). Observers should not be able to infer private information and whether a user performed or will perform a specific action. Moreover, no observer should be able to link an action to the user or infer if two (or more) actions were performed by the same user (device). Non-repudiation should be checked and verified by the Trusted Component (TC) hosted by each user device. | M | TPM DAA Module, Data Anonymizer, Searchable Encryption |
| **SR5** | Conditional Anonymity | Users should be anonymous within a set of potential participants. In case a user deviates from system policies, the corresponding credentials should be retrieved and revoked. | D | TPM DAA Module |
| **SR6** | User-controlled Unlinkability | According the users' preferences, in order to achieve unlinkability, no action or transaction should be able to be directly linked back to the original initiator without breaching the non-repudiation requirement of their transactions (SR3). Non-repudiation should be checked and verified by the Trusted Component (TC) hosted by each user device. | M | TPM DAA Module, Attribute-based Encryption |
| **SR7** | Data Privacy | One key aspect of DataVaults is the privacy guarantees on the | M | DataVaults Crypto Suite (i.e., DAA, |

| Number | Short Name | Description | ID | Supporting DataVaults Tool |
|---|---|---|---|---|
| | | data stored. This: (i) should guarantee the protection of sensitive information, (ii) it should be hard for an adversary to learn the secret information required for any action (e.g., authentication, encryption, etc.), and (iii) credentials should be stored on user device and must be protect from eavesdropping / leakage. | | Signatures, Attestation, Verification) |
| **SR8** | Forward and Backward Privacy | The revocation of a credential should not affect the unlinkability of previously signed data messages. Also, recovering the identity of the user of a particular credential should not affect the privacy of other messages signed by the same user. | D | TPM DAA Module |
| **SR9** | Fairness | Misbehaving entities should not be able to exploit the incentive and trading mechanisms to increase their utility without making the requested contributions or sharing the appropriate (anonymized) data. | D | DataVaults Blockchain, Infrastructure, Smart Contracts |
| **SR10** | Trustworthiness and Operational Correctness | All system entities need to be able to provide verifiable evidence on the correctness (i.e., correct configuration) of their current state. The operational correctness aims to provide a more holistic view of the system by combining dynamic and static attestation data in order to produce guarantees on the operational trust state of the system. | M | DataVaults Crypto Suite (i.e., DAA, Signatures, Attestation, Verification) |
| **SR11** | Cryptography | Having strong cryptographic primitives is a fundamental requirement of any security-oriented system. What is needed towards this direction is a good source of entropy that will be utilized in a secure pseudo-random number generator | M | Trusted Platform Module as hardware-based Root of Trust |

| Number | Short Name | Description | ID | Supporting DataVaults Tool |
|---|---|---|---|---|
| | | (PRNG) so that the keys generated by the system are secure. To make good use of this source of entropy, we also must ensure that the cryptographic primitives deployed in a root of trust and related systems are fit for purpose. | | |
| **SR12** | Ledger Security | (i) Integrity of block data - no one can tamper with the data stored in ledger; (ii) Verification of block data - the information stored in the block is valid and verified; (iii) Mining validation - a block mined by a user is valid; (iv) Agreement on validation - a majority or all network users to reach an agreement on validation; (v) Membership authentication - provide access control over ledger (read & write rights) for authenticated users; (vi) Guarantee of actions - deliver a mechanism that a "promised" action will be performed successfully; (vii) Customized block data security - enable authenticated user to put various encrypted levels of data on ledger. | M | DataVaults TPM-enabled Blockchain computation and verification functionalities |
| **SR13** | Physical Security | Systems entities (user devices and infrastructures) should be adequately (physically) secured against side-channel attacks. | D | Trusted Platform Module as hardware-based Root of Trust |

# 5 SECURITY, PRIVACY AND TRUST CONSIDERATIONS FOR PERSONAL DATA SHARING

## 5.1 PLATFORM AUTHENTICATION AND ATTESTATION ASPECTS FOR THE DATA OWNERS

As described in Deliverables D2.1 [2] and D2.2 [3], one of the core services that will be leveraged by the DataVaults platform towards enhancing the security posture of both the user devices but also the platform itself is **remote attestation**; both for **verifying the correct state of a data user's device** as well as for the **privacy-preserving platform authentication** when accessing and interacting with the DataVaults platform (data owners sharing/uploading their data).

In terms of design, DataVaults will leverage advanced crypto primitives in the context of both static and dynamic attestation; namely, **Configuration Integrity Verification (CIV)** [4] and **Direct Anonymous Attestation (DAA)** [5, 6]. The focus is on the provision of secure, robust, and efficient attestation, verification and privacy-preserving methods to check the internal state of a Data Owner – when accessing the DataVaults trading ecosystem – whose level of trust has not been verified, thus, **enabling secure enrolment and platform authentication services**. For the former, in the context of secure and authenticated configuration integrity checks / results, each step of the process is measured, e.g., by computing a cryptographic hash over the software image and platform configuration information; the resulting measurement is stored in a way that allows it to be securely retrieved later. This relies on an underlying root-of-trust for guaranteeing unforgeability of measurements. In this context, we consider a **remote adversary that attempts to compromise the binary immutable files of the data owners' devices.**

In what follows, we go into details on these two attestation variants and their interactions with the other DataVaults components towards offering the aforementioned services.

### 5.1.1 DataVaults Attestation Services and Protocols

Remote attestation (RA) [7] is a security service to validate the integrity of a remote entity (Data Owner platform) when accessing the DataVaults platform. In case of collective attestation schemes, the goal is to ensure the integrity of a multiple set of user platforms interacting with the DataVaults network. Different RA schemes collect different information which will be integrated into an attestation report. Depending on the information in the attestation report (i.e., configuration of execution behavioral properties), the *verifier* – DataVaults Platform – can detect different types of attacks on the *prover*-device's integrity. There are two types of attestation variants that are leveraged in DataVaults: **Static attestation** which is a mechanism for enabling the **detection of manipulation of a device's static memory content**, e.g., program code or loaded binaries configurations. **Run-time attestation** provides information about a **device's run-time execution (executional behavior)** [8, 9], allowing the verifier to also detect more sophisticated types of attacks and malware targeting specific functions or processes of a data owner's platform, e.g., compromising the key management functions used for interacting with the DataVaults DLT Engine (Section 5.3).

**Static Attestation:** Static attestation allows the verifier (DataVaults Platform) to check the software code and configuration of a prover system (Data Owner Platform) before accessing the system, i. e., the software code and configuration that got loaded on a device. This information is provided to the verifier in a status report that is generated in a secure and

authentic way (typically by a trusted component – TPM in our case – Section 5.3) on the prover system. Existing approaches fall into three different categories, differing in the mechanisms and components they use to achieve the required authenticity for the attestation reports. **Software-based attestation mechanisms** cannot use cryptographic secrets to authenticate attestation report as they work without any trusted component that could protect such a secret. Approaches that are based on **secure hardware** for which the authentication secret is protected and managed by dedicated hardware modules, such as a TPM. **Hybrid approaches use trusted software that itself is hardware-protected** to manage the authentication secret, i.e., a trusted software-component, typically isolated in a TEE, is responsible for authenticating attestation reports.

**Dynamic Attestation:** In dynamic or run-time attestation the execution behaviour of the prover is reported to the verifier coupled with the static properties of the prover's software code and configuration. This is usually done by recording the path through the program to be attested that was actually executed by the prover device. However, to minimize the amount of information to be recorded and reported, in run-time attestation it is usually assumed that the adversary can only manipulate data dependent code branches, e.g., function call, indirect jumps or function returns, which are exploited by run-time attacks like ROP.

The main building blocks of the DataVaults Attestation Toolkit are: (1) the state and event monitoring, (2) state storage (and compression), (3) reporting, (4) validation and verification, and (5) assessment and reaction. These building blocks have been implemented so that they can run on a Data Owner platform equipped with a trusted component such as a TPM. There are dependencies between the building blocks, for instance, the verification building block can make a decision only when the data provided by the state monitoring and state compression building blocks is precise and contains sufficient information.

The state and event monitoring must provide (a) precise information about all the relevant state information on a device and (b) must ensure the correctness of the data. The monitor building block leverages the tracing mechanisms implemented on each edge device to collect relevant state information. The spectrum of monitoring reaches from static properties, e.g., code and configuration integrity, to dynamic properties like control-flow and data-flow information. The correctness of the monitoring can be ensured by different on-device security mechanisms provided by DataVaults, e.g., by relying on dedicated, tamper-proof hardware components that prevent from the modification of monitoring component and the interference during monitoring.

#### *5.1.1.1 Configuration Integrity Verification*

In DataVaults, the **secure enrolment and platform authentication services**, comprised of the Zero-Touch Configuration Integrity Verification (S-ZTP CIV) and Remote Attestation variants (including the enhanced Direct Anonymous Attestation (DAA) for privacy-preserving data sharing based on the use of group-based signatures and pseudonyms), focus on the provision of operational assurance and secure platform authentication prior to a data owner interacting with the DataVaults platform. As aforementioned, the architecture followed by DataVaults follows a **decentralized approach where the DataVaults Platform (Identity Provider (IdP))**

**acts as the verifier for attesting the secure state of a device requesting access to the overall system.**

In the current implementation of CIV, DataVaults leverages one of the most prominent trusted computing architectures, namely the Trusted Platform Module (TPM) (Section 5.3). TPMs have been devised as a component of trust that enable to check the security posture of a system and provide mitigation controls against attacks such as not allowing the system to correctly boot up in case of a compromise, thus, also allowing the detection of anomalies during the reboot process. Furthermore, TPMs act as one of the main components handling the operations related to key management, such as key creation, storage, destruction and duplication (Section 5.3). The CIV architecture implemented in DataVaults [2, 3], currently relies on a SW-based TPM. **However, in the next release, the goal is to migrate to a pure virtualized trusted component by leveraging the QEMU implementation approach already defined in the literature [10]**. QUEMU pass-through makes use of a hardware-based TPM (HW-TPM) of a third-party server and can strongly link multiple remote virtual guests to this host server. Therefore, our approach provides **a good basis for a cloud-based implementation where multiple guests will be offered their own TPM functionality**.

CIV is based on the attestation services inherent to a TPM. This is the process by which a platform can report in a trusted way the current status of its configuration. It can be used for enabling the subsequent dynamic detection of possible software vulnerabilities when the platform state deviates from what has been defined as “trusted”. The report can include as much information as required based on the already defined control policies (including the configuration and behavioural properties to be traced and verified). **The basis of the attestation are the measurements recorded in Platform Configuration Registers (PCRs).**

PCRs are one of the essential features of a TPM that allows it to act as a Root-of-Trust for Reporting (RTR) and Measurement (RTM). A PCR is a memory register that can store the entire output of a hash algorithm (e.g., 256 bits for SHA-256), and provides a method to cryptographically record a log of measurements corresponding to the software states that affect the security condition of a platform. In the context of Trusted Computing, such measurements are initiated by the RTM, and are expected to take place, at least, during the boot phase of the collection of system-resources responsible of maintaining the security policy of the system.

The PCRs can then be read to know the current status of the platform and be also signed to provide a secure report. The signed message can then be sent to the DataVaults IdP that will act as the verifier. It is worth noticing that the TPM does not check the measurements, that is, it does not know whether a measurement is trustworthy or not. The trustworthiness of the measured value comes when an application uses some PCR value in an authorization policy (*Attestation by Proof*), or the DataVaults IdP asks for an attestation of some value (Attestation by Quote) when a data owner’s platform is trying to access the system. **Attestation enables such clients to confirm whether a platform has been compromised. Additionally, the TPM offers means of certifying and auditing the properties of keys and data that cross the TPM boundary.**

In various platforms and devices, processes and files are loaded in parallel, driving an explosion of loading-order-paths that are almost impossible to match to a reference. Also, many processes often create their own files on the system (e.g., state, configuration files, logs etc.), files for which there can be no initial reference as for the executables. In addition, various

processes can directly or indirectly interact with each other, such as through inter process communication (IPC) or through successive file writes-reads, making it hard to evaluate the impact of an unknown process on the other ones.

To this end, in the context of DataVaults, we have introduced the **Configuration Integrity Verification (CIV)**, please see Figure 4**Fehler! Verweisquelle konnte nicht gefunden werden.**: **an architecture that allows to assess and/or preserve the integrity of a data owner's devices, at load time and during system execution, while ensuring predictability of the PCR values regardless of the order of loading of applications and reducing performance impact by dramatically reducing the number of TPM PCR extend.**

CIV is building on the Integrity Measurement Architecture (IMA) and Extended Verification Module (EVM) features of the Linux kernel and introduces new attestation features (**Attestation by Proof and Attestation by Quote**) based on the Clark-Wilson integrity model [11, 12]. It monitors the information flows between TCB processes and those outside the TCB and can prevent violations or record them in the TPM-protected IMA measurement list. CIV introduces a concept of *digest lists* to limit the reporting of measured software only to the case of unknown software (not added to the digest list). This approach ensures predictable PCR values and reduced usage of the TPM and, consequently, reduces the performance impact. It also introduces *Simple Remote Attestation (Simple RA)*, to minimize the effort of integrating Remote Attestation in existing distributed architectures, by using implicit attestation over existing secure protocols (e.g., TLS), while addressing the lack of dedicated standard attestation protocols and thus mitigating interoperability concerns. The CIV overview and the underpinning of the internal operations of the mechanisms, i.e., Attestation by Proof and Attestation by Quote, is depicted in Figure 5.

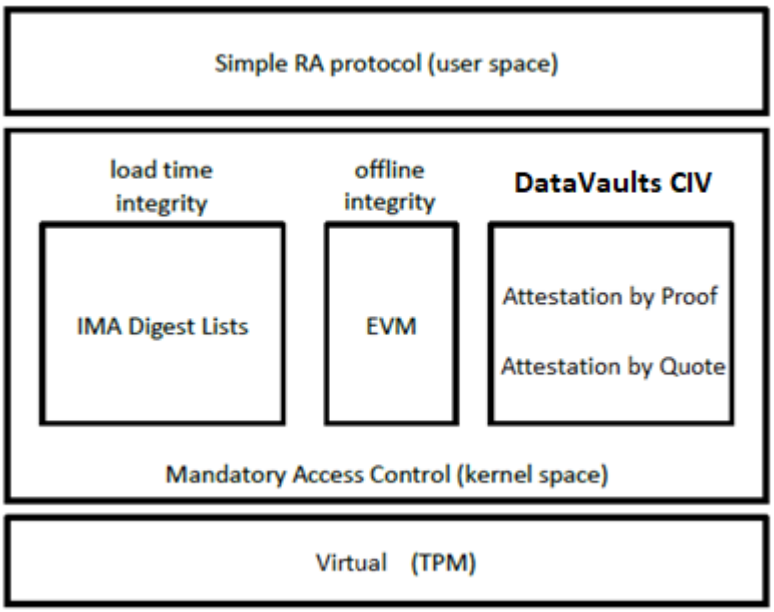


Figure 5 - DataVaults CIV

The workflow is the following: CIV verifies immutable files by searching for a file digest in the digest lists provided by the DataVaults IdP. Alternatively, CIV detects/prevents offline attacks on mutable files by verifying the HMAC and detects/prevents online attacks by restricting access (through the Attestation by Proof) to the processes that are able to modify those binary files. Measurements and attestation reports produced by CIV (only if the verification failed) are used by the DataVaults IdP Controller for authenticating a platform depending on its correct state, based again on the defined control policies. In this context, DataVaults provides the following two functionalities:

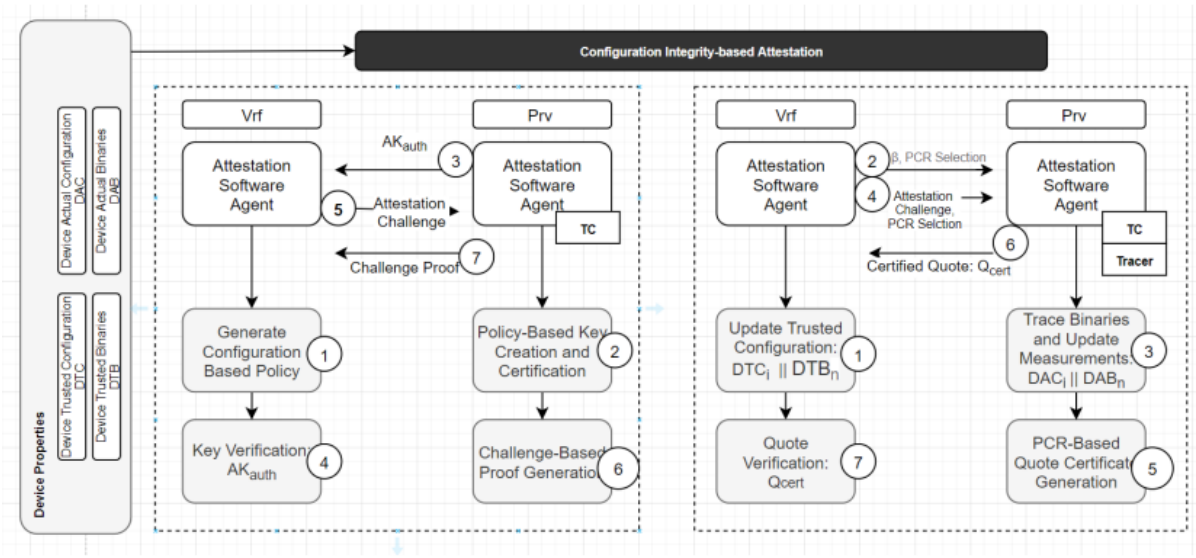


**Figure 6 - DataVaults Workflow of CIV: Attestation by Proof (Left) and Attetation by Quote (Right)**

**Attestation by Quote:** This is a functionality where the DataVaults IdP Controller requests from Data Owner to perform a TPM Quote operation for securely extracting its current integrity measurement list – prior to entering the system. A quote is a TPM command that generates a hash over a selected set of PCRs and then signs it with a signing key. In our case, we generate an Attestation Key (based on the TPM's Endorsement Key) to sign the quote, before sending it to the DataVaults platform. When received, the attestation security service verifies the signed quote, and follows up by comparing the quote itself against a whitelist. This has the advantage of being able to be resistant to unwished updates and configuration changes since it simply measures, signs, and transmits the integrity measurement list. However, there is one important challenge: *how to protect the derived Attestation Key (AK)*. For instance, if the key is protected by the PCRs (i.e., the PCRs has to be in a certain state in order to allow the loading of the key), then this limits the dynamicity of the process (to be able to be performed during runtime) since the AK is not resistant to changes on the VF. On the other hand, if we protect the key with a secret, then this secret must be securely stored on the host or be transmitted during attestation requests. In both cases, if an adversary is present on the target VF, she will be able to take control of the AK and, thus, the entire attestation process. To act against such a quote manipulation, a nonce is included in the attestation structure and the AK is declared as restricted, which means that it can only sign digests generated by a valid TPM. This is also enhanced with the use of appropriate "tickets" (i.e., short-term certificates that are created and signed by the TPM for verifying the parameters used during the AK creation) proving that a quote is generated within a TPM.

**Attestation by Proof:** Instead of requesting a TPM Quote, we have also implemented the functionality allowing the integrity verification to happen locally within the Data Owner's device. To build such a proof, the DataVaults IdP instructs the device to create an Attestation Key (AK), and seal it with a digest that reflects the correct PCRs. The VF can then send back the public part of the AK and signs it with its Endorsement Key. The security controller can now send a nonce to the device, which then loads the AK and signs the nonce. This can be done if and only if the current PCRs are in a correct state. If the system configurations are updated, the orchestrator could simply instruct the VF to create a new AK with an updated digest, to reflect the new configurations.

Overall, the offered CIV allows to assess the integrity of a Data Owner's platform, when accessing and interacting with the DataVaults ecosystem, so as to make sure of its correct state before granting access privileges to the offered data sharing and trading services. This enables the secure platform authentication both during access control but also during run-time – when initiated by the DataVaults platform itself.

### 5.1.2 Direct Anonymous Attestation (DAA)

As aforementioned, for privacy, DataVaults will also offer another variant (on top of the option for creating User PERSONAS) by leveraging advanced crypto primitives, namely **Direct Anonymous Attestation (DAA)** [5, 6] based on group signatures. Privacy requirements that are captured by DAA are the ones already documented in the ETSI TS 102 941 standard: **anonymity** *(ability of a user to use a DataVaults resource without disclosing its identity)***, pseudonymity** *(ability of a user to use a DataVaults resource without disclosing its identity while being accountable for that action)***, unlinkability** *(ability of a user to make multiple uses of DataVaults resources without others being able to link them together)***, and unobservability** *(ability of a user to use a DataVaults resource without others being able to observe that the resource is being used)***.**

Direct Anonymous Attestation (DAA) is a cryptographic protocol that allows a Trusted Platform Module (TPM) to serve as a trust anchor for a host platform it is embedded in. To do so, the TPM chip creates attestations about the state of the host system, e.g., certifying the boot sequence the host is running on. These attestations convince a remote verifier that the platform it is communicating with is running on top of trusted hardware and using the correct software. **A main design goal of DAA is that attestations are made in a privacy-preserving manner.** That is, the verifier can check that attestations originate from a certified hardware token, but it does not learn anything about the identity of the TPM. Another important feature of DAA is that it supports user-controlled linkability which is steered by a base name bsn. If a platform uses a fresh or empty base name, the resulting attestations cannot be linked whereas repeated use of the same base name makes the transactions linkable. A DAA can be seen as a special variant of group signatures with a central issuer controlling membership to the group of certified TPMs, and TPMs being able to sign anonymously on behalf of the group. Instead of the opening capabilities provided in group signatures, DAA controls privacy through the use of base names and user-controlled linkability. DAA will be used in DataVaults for **enabling Data Owners to both authenticate their platforms in a privacy-preserving manner but also share their data in an anonymous way by leveraging group-based pseudonyms** [6].

As described in D2.2, there is a two-step approach followed in DataVaults when a Data Owner has selected the use of group-based pseudonyms as the means for offering anonymity assurance:

1. **Platform Registration and Authentication:** A secure channel is established between the Data Owner's device and the DataVaults platform itself, with the help of DAA (i.e., DAA-SETUP and DAA-JOIN phases). As aforementioned, this enabled the Data Owner to authenticate him/herself (in a privacy-preserving manner) to the DataVaults IdP so that they can then exchange session keys. This provides a proof to the platform that the user controls a valid TPM, issued with access permissions.

2. **Pseudonym Creation:** Once the authentication was successful, the Device owner – through its TPM - can create anonymized DAA-identities (i.e., pseudonyms through the DAA-CREATE phase) that can be then used for accessing Blockchain related services, e.g., sharing/uploading data. Towards this direction, the DataVaults Quorum DLT Engine needs to also create a fresh Blockchain key-pair linked to this newly introduced DAA-identity. The Data Owner can create as many DAA-identities as he/she wants for being able to use a different pseudonym per shared data bundle, thus, achieving unconditional anonymity.
3. **Privacy-preserving Data Exchange**: Through the use of DAA-SIGN/VERIFY phases, the Data owner can share his/her data to the DataVaults platform by using a different (or the same depending on the preferred level of privacy assurance) pseudonym.

## 5.2 THE SECURE COMMUNICATION CHANNEL BETWEEN THE DATA OWNERS AND THE DATAVAULTS CLOUD-BASE PLATFORM WHEN UPLOADING/SHARING THEIR DATA

Currently, Public Key Infrastructure System (shortly PKI), the hierarchical trust relationship system, is the most widely used cornerstone technology to help secure the communication channels. However, PKI recently concerns the users due to various security breaches, i.e., the compromised PKI allows attackers to issue any valid keys to the victim and decrypt any secure connections within the system. To mitigate such issue, we introduce a new decentralized PKI system, called SecurePKI, by leveraging the blockchain technology, as shown in Figure 5. The basic system is consisted of two major parties: the end-device and the server.

- End-Device: IoT gateways or SBCs that have TPM platform built-in with capability of running an Ethereum light node. If any of them are not capable of such, a device that works as communication middleware is required. These devices are expected to have limited computational resources and storage space. Some even run-on battery support. Ethereum light node provide options for these devices to request nearby full-/partial-node servers to finish their transactions (E.g., send and retrieve data from Blockchain). This means the light nodes are only require meeting the capability of sending request, the task that is simplify enough for most of these devices.

- Server: Servers are assumed well-protected and managed by administrators. These machines located in the environment that computational resources are sufficient.

This blockchain-based secure connection can ensure the followings:
1. IoT devices can validate and recognize the validity of the network membership of a node.
2. Members can validate any transaction occurrences in the network.
3. Symmetric session key is applied during data transmission.

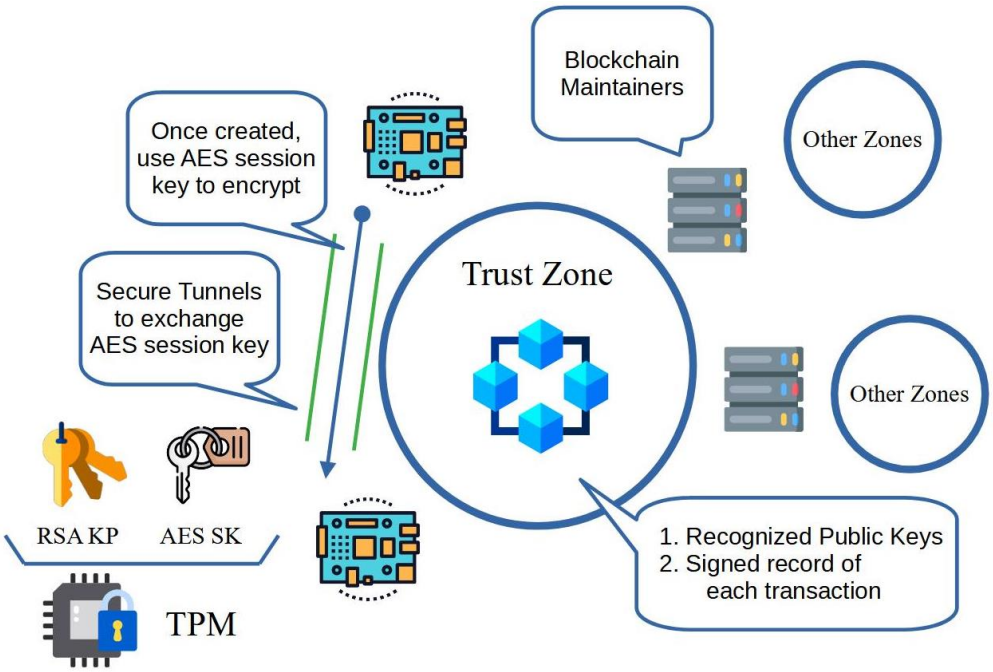


Figure 7 - High-level architecture of SecurePKI for DataVaults

### 5.2.1 To initialize a trust zone

Administrators must manually register the end-devices' public keys on to the newly created trust zone. All devices recognized the advertised key-entity bond on the Blockchain as valid. New nodes that do not registered and approved by the administrators will be denied by other devices.

### 5.2.2 To make a secure connection

To connect to other device securely, the originator will first check its session key table to see whether the timeout has been reached. If so, the device will establish a secure channel by asymmetric encryption, which the public-key bond must be advertised on the Blockchain. Once the temporarily secure channel has been established, agreement on creating a session key is established. Once both devices finished the session key establishment, such will communicate under the encryption of the session key. However, we cannot ensure that such packet may be altered. To prevent this issue, all occurred transactions is hashed and bonded with a query ID, which then uploaded to the Blockchain. All nodes can check the integrity of the transaction.

## 5.3 USE OF TPM BUILDING BLOCKS AND SERVICES FOR THE SECURE KEY MANAGEMENT THAT WILL ALSO BE USED IN THE ATTESTATION SERVICES

Secure key management is one of the main features provided by the TPM, enabling many advanced use cases. For attestation, it ensures that the proofs are correct and signatures originate from the authorized platform and cannot be forged. The following section gives an overview of the main functionality and presents a high-level description about their application for the attestation services.

The most important key of a TPM is the Endorsement Key (EK), which is uniquely generated for the chip during production. It is an asymmetric key pair, where the private part never leaves the secure storage. The manufacturer usually provides a certificate, to act as prove that the stored key is protected by a genuine TPM.

In addition to the EK, TPMs can hold many keys with a large variety of configurations. It can generate a new key based on internal (secret) seeds, with the help of the internal Random Number Generator (RNG), or imported externally. For the internal keys, it is ensured that a strong key with good entropy is created and the secret key cannot be extracted. The keys are organized in hierarchies, where the parent can be used to encrypt or sign the child. This scheme creates a chain of trust, where a key can be certified by the TPM.

Keys are generated based on attributes to limit the usage. This concept is important to understand the security properties of the advanced applications. The following lists the main configuration attributes when generating a TPM key:

Application type: Each TPM key can only be used for signing or encryption. Using the same key for both operations breaks the security of the key, because signing of data implicitly decrypts it (and vice versa). An adversary could thus trick the application to sign or decrypt arbitrary data. To prevent this potential vulnerability, the TPM design allows only one predefined operation type for each key.

Cryptographic algorithm: Different algorithms and key types (e.g., symmetric and asymmetric) are provided to support a wide range of applications.

Duplication restriction: As previously mentioned, there is the possibility to import keys. This could be useful for some applications, where multiple devices are externally treated as one (e.g. load balanced servers), or when encryption keys require a backup or migration to another machine. However, for other applications (e.g., attestation) a duplication would void the required identity and non-repudiation properties.

Usage restrictions: Unrestricted keys can be used flexible as replacement for general-purpose software implementations. The user supplies the input to the cryptographic operation and receives the result (e.g., data for signing or encryption). Restricted signing keys are required when only internal data should be signed. This is a strong requirement for the attestation, where the user should not be able to request a signature for arbitrary data to spoof an integrity measurement. Restricted decryption keys are used for key import, where the result of the operation should not be returned to the user and only be used internally by the TPM.

The previous list shows the large variety of key configurations to enable advanced cryptographic schemes. This is the main requirement to enable unforgeable signatures required by the attestation services. The second part of attestation is concerned with checking

the system state with the Platform Configuration Registers (PCRs). Those are special registers inside of the TPM, which can only be modified with an *extend* operation. It takes the current value, concatenates it with the user provided input and updates the value with the cryptographic hash of this data. A sequence of measurements can be made to check the system state and compare it to a precomputed control value.

In combination with the advanced key properties described above, it enables two attestation schemes. The theoretic foundation was developed in the CloudVaults [1] project, DataVaults will evaluate both approaches in a practical setting:

**Attestation by Quote:** The values of a selection of PCRs are signed with a restricted TPM key. The key configuration is certified by the TPM to guarantee that it is located on a genuine TPM and only internal data can be signed. Otherwise, an adversary could simply sign the expected data externally without using the PCR mechanism at all.
When requesting the quote, a nonce value is included to ensure freshness of the data and certify the current system state. An authority can then verify the signature, check the quote data and compare it to a precomputed reference value. One downside of this approach is the loss in privacy. Because the original PCR values need to be sent every time, the authority can use them to draw conclusions about the configuration of the system.

**Attestation by Proof:** TPM keys can also be linked to predefined PRC values, allowing them to be used only when the system is in a known-good state. This restricts the ability of producing a signature to a certified system configuration. In this scheme, the authority sends a nonce value as challenge and expects it to be signed by the TPM. If the client is able to produce this signature, it is a sufficient proof that the PCRs contain the expected values.

# 6 SMART CONTRACT, MICROPAYMENTS, COMPENSATION SCHEMES, DATA VALUE FLOWS

## 6.1 SMART CONTRACTS

The DataVaults platform uses **Smart Contracts** as central element **for trusted and traceable data sharing**. A general description of the Smart Contract technology was already given in previous deliverables, this document focuses on the detailed application for the project design. The following subsections conclude the research and present the final decisions as basis for the implementation tasks.

### 6.1.1 Transaction privacy

Quorum offers the optional feature of private transactions, where the content of Smart Contracts and details of function calls are kept confidential between a selectable subset of

nodes. This functionality is provided by the private transaction manager *Tessera,* which stores the private data encrypted and transmits it to the selected recipients.

However, a purely off-chain data transfer would circumvent the immutability and non-repudiation properties of the DLT. For this reason, the cryptographic hash of this encrypted data is additionally written as public transaction to the Blockchain. Authorized nodes use this hash to retrieve the corresponding raw data from the database when a transaction of this kind is received. Other nodes pay no regard to this transaction payload, ensuring only transaction validity (by checking the signature) and helping with consensus finding.

This mechanism is particularly useful when applied to Smart Contracts, because deployment and interactions are fully handled by adding a special payload to regular transactions: Authorized nodes can decrypt the bytecode and deploy the Smart Contract in a private database. Subsequent write operations are also distributed encrypted to prevent leakage of sensitive information. The function identifier and optional parameter values are read from the encrypted payload and all authorized nodes execute the request and update their local state.

From the developer's perspective, the previously described process is mostly transparent and does not require changes in the workflow. Quorum handles the interaction with Tessera internally, the application can simply use the provided API to deploy and access private Smart Contracts.

### 6.1.2 Smart contract functionalities

As described above and in previous deliverables, **Blockchain and Smart Contracts are a key building block** for the operation of the DataVaults infrastructure. Many of the operations are happening through the execution of such contracts, and many of the different components need to execute and retrieve information from such Smart Contracts, executing either on the private- or on the public ledger.

In general, Smart Contracts in DataVaults will be used for facilitating certain operations that have to do with the recording and retrieval of information. Records should be saved in an immutable manner in the system for providing the necessary trust guarantees, and of course for supporting the compensation mechanisms that will be deployed. As such, the main functionalities of the smart contract will be:

- Storing in the ledger the **sharing configuration** relevant to any data sharing activity from the side of the Data provider/Individual
- Storing in the ledger the attributes and outputs of any **data acquisition transaction** that happens between the **Data Seeker and the DataVaults platform** and transferring the respective value between the Data Seeker's and the DataVaults Platform's wallets.
- Storing in the ledger the attributes and outputs of any **data acquisition transaction** that happens between the **DataVaults platform and the Data Provider** and transferring the respective value between the DataVaults Platform's and Data Provider's wallets.
- Enforcing the proper access control and decryption options for any data asset by providing to the relevant engines the corresponding information as stored in the ledger

- Providing a **complete log of data shared** by individuals, contributing in that manner as an input source to the risk exposure management monitor features offered to Data providers

The following table provides a revised view on what conceptual types of smart contracts will be available over the DLT infrastructure in DataVaults

**Table 4. DataVaults Activities Relevant to Smart Contracts.**

| ***Component*** | ***What activity is being carried out in the system*** | ***Type of smart contract involved*** | ***Components Involved*** |
|---|---|---|---|
| **DataVaults Personal App**<br><br>Collecting personal data, configuring, and setting sharing parameters for those data | The DataVaults Personal App offers capabilities to the Individual user to manage data access policies and data sharing configurations. | SC relevant to data sharing configurations | • *Sharing Configurator*<br>• *Private Ledger* |
| | The DataVaults Personal App offers capabilities to the Individual user to remain aware of privacy exposure. | SC for reading the executed data sharing configurations | • *Sharing Configurator*<br>• *Risk Management Monitor*<br>• *Private Ledger* |
| | Receives compensation for the data assets they place at the disposal of third parties. | SC for transferring the value from the platform's wallet to the wallet of the data provider | • *Private Ledger*<br>• *Public Ledger*<br>• *Platform's Wallet*<br>• *Individual's Wallet* |
| **DataVaults Cloud Platform**<br><br>Supporting Data Seeker connects to explore, acquire and analyse data coming from Individuals | The DataVaults Cloud Platform allows Data Seekers to search through data of Individuals and express their interest to acquire them. | SC for reading access and decryption policies that have been set during the sharing configuration phase | • *Public Ledger* |
| | The DataVaults Cloud Platform allows Data Seekers to create data request Contracts and execute data acquisition transactions. | SC for initiating and executing a data acquisition transaction, and transfer the relevant compensation | • *Public Ledger*<br>• *Data Seekers Wallet*<br>• *DataVaults Platform Wallet* |
| | The DataVaults Cloud Platform is used to transfer compensation to the Data Providers once a data acquisition transaction is validated. | SC for transferring the relevant compensation to the Data Provider's wallet | • *Private Ledger*<br>• *Data Seekers Wallet*<br>• *DataVaults Platform Wallet* |

### 6.1.3 Access Policy Contracts

The access policies are elements created using the Access Policy Editor that can be accessed from the Sharing configurator at the Personal DataVaults App side.

The information gathered there is related to the conditions that the users or data owners want to apply when sharing their data. The policy and other indications related to the data sharing configuration are stored as a sharing contract through the execution of a smart contract.

The structure of the access policy part of the contract follows the model defined in the D1.2

DataVaults core data model [2], based on the Open Digital Rights Language (ODRL) data model for defining policies. Each policy is composed by a set of rules:

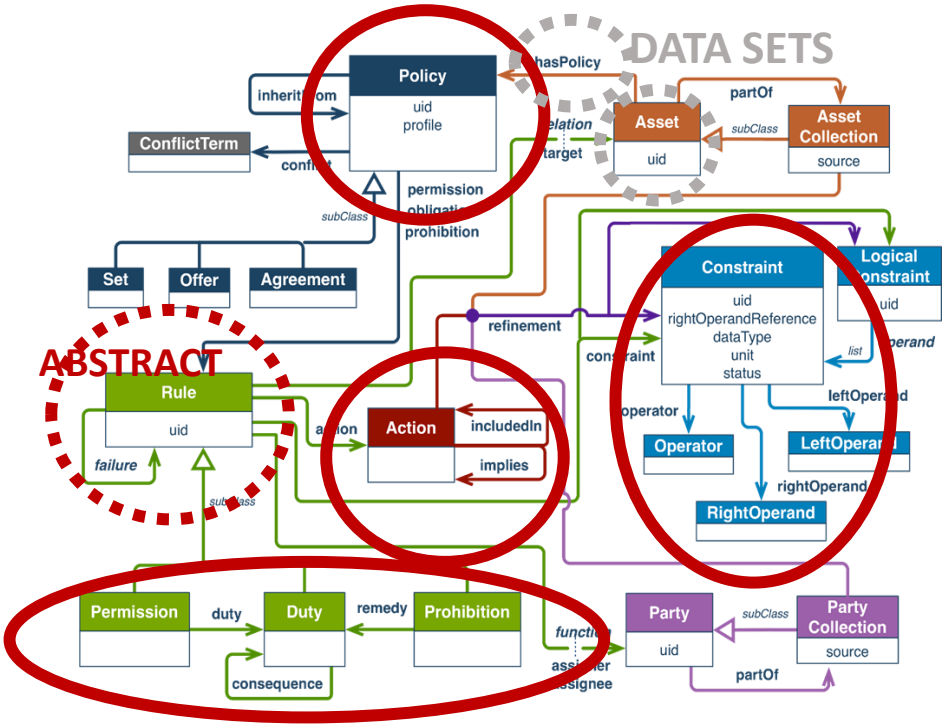


Figure 8 – DataVaults access control model

- **Policy:** DataAsset is linked to a Policy (or more-TBD) over hasPolicy property
- **Rule:** A policy can consist of several rules or one rule (it is a decision to take if each new element is a rule or a policy). It is an abstract element, that means that it must not be used directly, and use Permission and Duty instead
- **Action:** USE is the action that contains the rest of them (distribute, compensate, write, grant_use, encrypt, notify, aggregate_by_consumer, anonymize)
- **Permission:** Permission allows access and/or usage of a dataset with further specifications in the form of Constraints or connected Duties. At least one permission is needed in order to allow usage of a data asset.
- **Duties:** Duties frame any type of obligations connected to the usage of a dataset

- **Constraints:** serve as conditions, which must be satisfied before a Rule is active

  leftOperand: The attribute you want to evaluate

  rightOperand: The value that makes the evaluation true

  Operator: The way the evaluation will be made

In the case of allowing access of data in DataVaults, the type of rule will be “permission” and the criteria to allow that access would be defined with the constraint element. These constraints check the real values of the attributes of the seeker against the allowed values in order to inform that the access is allowed.

The structure of a contract can be as follows:

- Individual_ID
- Dataset_ID
- Sharing configuration aspects (Prize, Anonymization needs, etc.). These aspects will be defined within the “Sharing configurator” design process.
- Access Policy:

```
"@context": "http://www.w3.org/ns/odrl.jsonld",
"profile": "datavaults_profile",
"permission": [{
       "target": "dataSetURI",
       "action": "Access",
       "constraint": [{
           "leftOperand": "seeker_attribute_location",
           "operator": "EQ",
           "rightOperand":"Europe"
       }
,{
       "constraint": [{
           "leftOperand": "seeker_attribute_purpose",
           "operator": "isPartOf",
           "rightOperand":"list_of_allowed_values"
       }
]
```

The process for storing the contract consists of calling the corresponding function exposed by the Smart Contract passing the individual_ID and the Dataset_ID as parameters, in order to make a backup of the conditions set by the individuals for sharing their data. Private transactions will be recorded in the private state of the DLT.

The Policy Engine will call the function for retrieving the policies associated to datasets from this DLT and taking into account the values of the attributes informed by the seeker requesting access.

The public part of the DLT will register the conditions once the seeker requests a particular access and it has been allowed.

The types of constraints envisioned for managing the sharing conditions from the individuals are associated to the seeker attributes and the Engine has to read those attributes and execute the comparisons defined by the policies.

The correspondence among those attributes and the type of standard policies are:

**Table 5. Attributes and types of policies**

| Type of policy | Seeker attribute |
|---|---|
| Interval-restricted | n/a |
| Duration-restricted | n/a |
| Restricted Number of usages | n/a |
| Location-restricted | Countries |
| Event/processing-restricted | n/a |
| Purpose-restricted | Sector/Industry group |
| User-role restricted | n/a |
| TBD | Organization Type |
| TBD | Reputation Score |

**n/a: There is no attribute of the seeker for this type of policy. It should be analyzed taking into account the objectives of the project for further versions of the Editor.*

**TBD: There is no type of policy covering this attribute. It should be analyzed taking into account the objectives of the project for further versions of the Editor.*

### 6.1.4 ABE/SSE Contracts

As identified in the DataVaults architecture, and in the related user stories and feature, DataVaults will explore security options that have to do with encryption of data and metadata towards defining a highly flexible framework for enabling data sharing but at the same time guarantee to Data owners the confidentiality of their data. As such, two distinct mechanisms will be put in place, in order to make sure that only authorised actors are able to get access to the unencrypted data, while at the same time allow for all other actors to perform queries over encrypted data, without however disclosing the real data itself.

#### *6.1.4.1 ABE Engine:*

For the first group of features, DataVaults is going to work towards the design of an Attribute Based Encryption Engine, which will be utilising information stored in the smart contracts to operate. This ABE engine, which is built over the Ciphertext Policy (CP-ABE) pillar, manages access to cyphered data by applying access policies in the form of Boolean expressions, which must be satisfied by a decryption key. In the context of DataVaults, individuals are offered to apply an extra security layer by protecting the access to and/or search of their data with Searchable Symmetric Encryption (SSE) and ABE encryption schemes.

ABE schemes are built on the premises of an attributes' ownership game. This is, the subject to protect is cyphered by applying an access policy which must be satisfied by a decryption key in order to decrypt the data. This decryption key must contain a set of attributes that fulfil the access policies.

The ABE scheme candidates to be used in the project present some limitations to build Boolean expressions:

- Policies can only contain *AND* and *OR* gates.
- Policies do not evaluate a parameter with a value, instead they evaluate the presence of an attribute into the policy.
- Depending on the implementation of the encryption scheme attributes can be natural numbers or strings.

Now we provide some definitions to explain the behaviour of the ABE engine:

- Condition: It is a **Boolean expression** built by a set of mandatory attributes. It takes the form of a AND chain:
  - Attribute_1 AND Attribute_2 AND…
- Policy: It is a **set of Conditions** where the fulfilment of any of the conditions result on the fulfilment of the policy:
  - Condition_1 OR Condition_2 OR…
- Pattern: It is a **set of Policies** and identifiers where an identifier points to a subset into the data and the policy is the one to be applied at the encryption process. These are examples using JSON paths and XML paths.
  - {[$.activity.path.geoLocation,Policy_1],[$.activity.health.heart_rate, Policy_2]}
  - {[/activity/path.geoLocation,Polliciy_1],[/activity.health.heart_rate, Policy_2]}

As starting point, the ABE engine will take as attributes the list described in the previous section, this list will be enlarged later to enhance the capability to build policies.

Individuals will take combinations of attributes to build policies that will be used to encrypt data following patterns. On the other side, seekers must present, in a secure and trusted way, a set of attributes they own. These attributes will be embedded into a decryption key. Therefore, seekers will only be able to decrypt those cyphertexts encrypted with a policy that can be fulfilled by their keys.

The ABE engine manages three different types of keys: Master Secret Key (MK), Public key (PK), and Secret keys (SK) which are generated from MK for a given set of attributes. While the pair PK/MK are created at the instantiation of the scheme, SK is generated on demand when it is requested by the seeker and provides link to the authorization service to collect her attributes.

Beyond attributes used to generate SK, which are the ones owned by the seeker who has request the key, the ABE engine can also attach a set of parameters to manage the life cycle and usage of SK: validity period of the key, number of usages, and Dataset ID in the case SK was generated for a specific dataset. These parameters are attached in a secure way to SK and evaluated at decryption time. DataVaults will use ABE encryption schemes developed in FENTEC project[28].

#### *6.1.4.2 SSE Engine*

This type of encryption utilizes symmetric keys, which allow for a faster execution of decryption tasks, however they suffer from inherited key revocation issues, Therefore, applying SSE on the whole payload of a dataset is rather inefficient at this point, and as the indexing on the DataVaults platform will be done primarily on a metadata level, the same shall apply also for the SSE to offer to users a unified querying experience.

Using Symmetric Searchable Encryption allows DataVaults to maintain a higher degree of privacy and security guaranteed for the different Data Owners, as they can make their assets discoverable and available to Data Seeker, but at the same time lock their access based on encryption properties, selecting to whom to hand over the different decryption keys for accessing the content of those data. This type of encryption utilizes symmetric keys, which allow for a faster execution of decryption tasks, however they suffer from inherited key revocation issues. Therefore, the approach taken in DataVaults, as this will be described later, includes the utilization of this scheme alongside with the ABE scheme in an effort to protect the keys as much as possible.

DataVaults will work on the SSE Scheme that has been devised by the ASCLEPIOS project[29], which is used to encrypt metadata of files and then allow users to search on top of those encrypted metadata. This done because applying SSE on the whole payload of a dataset is rather inefficient at this point, and as the indexing on the DataVaults platform will be done primarily on a metadata level, the same shall apply also for the SSE to offer to users a unified querying experience.

#### *6.1.4.3 SSE/ABE integration*

There are many options to combine these two engines in the literature. In DataVaults, given the use of different authorization mechanism and to avoid overlapping, we envisage two options:

[28] https://fentec.eu/
[29] https://www.asclepios-project.eu/

1. Search protection: This option is based on encrypting data with SSE scheme and protect SSE keys with ABE, in this way seekers will be only permitted to perform those searches they are enabled by their attributes, through ABE decryption keys. Once the seeker obtains a list of data sets of their interest, they can request these data sets through the Access Policy Engine. This option enables to manage searches each seeker is permitted to perform.
2. Layered data protection: This option, addressed to structured data, is based on encrypting data with ABE, applying a different policy to each piece of data and encrypt metadata with SSE. In this case, the seeker can discover those data sets of interest by querying for metadata. Once a list of relevant data sets is obtained, access can be requested through the Access Policy Engine. The ABE decryption key is then used to gain access to the data sets she is permitted to see. This option enables the application of more restricted policies and differentiate among the different pieces of information into a data set.

At the moment, the second option seems more preferable for DataVaults, as it allows metadata to be discoverable by much more users.

## 6.2 COMPENSATION SCHEMES

### 6.2.1 Micropayments

The state-of-the-art analysis in Delivery 2.2 listed the common technologies for micropayments and transaction aggregation. Those schemes are used in public Blockchain networks to prevent disproportionate transaction costs when calling a Smart Contract function or sending a small amount of currency.

Based on the technology evaluation and proceeding architecture definitions, DataVaults will use *Quorum* for private consortium Blockchains. All transactions, which are triggered by functions offered by the platform, will be processed by the DataVaults Backend and therefore not impose any cost to the user. **Fehler! Verweisquelle konnte nicht gefunden werden.** compares the transaction throughput of *Quorum* to Bitcoin and Ethereum. The provided performance exceeds the expected workload of the demonstrators, avoiding the need for transaction aggregation schemes.

**Table 6. Comparison of transaction throughput and latency**

| | Tx/second | Max. latency |
|---|---|---|
| **Quorum public transaction [2]** | 2100 | 2 seconds |
| **Quorum private transaction [2]** | 900 | 4 seconds |
| **Bitcoin** | 7 | 10 minutes |
| **Ethereum** | 15 | 13 seconds |

The compensation for a data sharing activity is internally transferred to the Wallet of the Data Owner. A more detailed description of the Wallet is given in the next section, but it essentially allows the transfer and storage of currency in a privacy-preserving manner.

The last step in the compensation flow is the exchange of the digital coins to a fiat currency, coupons or goods from a merchant. This work package focuses on the internal value flow, supported external compensation types are investigated in the business value exploration and exploitation activities.

Independent from the compensation type, the platform will provide a flexible interface to support the exchange of the internal currency to external compensations. Multiple types with different exchange rates could potentially be supported in parallel. The platform may also mandate a minimum quantity and deduct a processing fee for some exchange operations.

### 6.2.2 Personal Wallet

Personal Wallet in DataVaults is the tool Data Owners use to manage compensations they receive by sharing data. It consists of an electronic wallet, or eWallet, following the Blind Signature Specification for Untraceable Payments from David Chaum [4]. It describes a system based on an exchange service, which by means of a **blind signature cryptographic scheme** creates and validates electronic representations of coins in an anonymous way.

In this system, customer users transfer funds to an Exchange Server and can create and use electronic representations of their funds, while any merchant user who receive one of these representations as payment can check their authenticity **without learning identities** nor any other private data of customers. The main requirements of the system are that users must be adhered to the system and that it is an on-line system, this is, users must be connected to the system at the time of carrying out any transaction.

The basic flow in this kind of platform can be described in the following steps:

0. Establishment of the system: A user, through an electronic wallet adheres to the system and creates an account in the Exchange Server.
1. Transfer: The customer transfers funds to her account in the exchange.
2. Withdrawal:
    a. Hashing: the wallet of the customer generates a random number named *Hash*, obfuscates or blinds it using the public key of the Exchange Server and request a signature for the blinded hash to the Exchange Server.
    b. The Exchange signs the blinded hash
    c. The wallet applies the inverse blinding function to the signature. With this procedure, the wallet obtains a valid signature for the original hash (not blinded), that it is not known by the Exchange Server, but which can be verified with the public key of the Exchange server. This set, hash plus signature, is the most basic form of electronic coin.
    d. The exchange transfers the correspondent founds from the customer account to the Exchange bag.
3. Payment: Once the coin has been built, the wallet of the customer can pay to a merchant. The most basic procedure to pay is to send the coin as it is to the wallet of the merchant. The wallet of the merchant can verify the authenticity of the coin by checking the signature with the public key of the Exchange Server.

4. At this point, the merchant can request to redeem the electronic coin or use it in other transaction.

This scheme was initially designed for daily micropayments and has been deeply studied to expose its shortcomings and provide improvements; additionally, it was already widely exploited.

In DataVaults, we integrate a payment platform based on this scheme that has been developed in the FENTEC project [5]. It is a one-way payment platform, once a merchant accepts an electronic coin as payment, she can only redeem it by querying the Exchange Server to transfer the correspondent funds from its bag to the account of the merchant. It also introduces a new player, the Trusted Authority that contributes to provide a privacy preserving linkability functionality. The Trusted Authority participates in the creation of the eCoin in a way that prevent any of the players of the system to learn the identity of customers, but in case of need, enable to establish a link between an electronic coin and the customer who created it by collaboration between Exchange Server and Trusted Authority.

In this type of systems, an electronic coin is a construction formed by a set of cryptographic keys and hashes. They can be interchangeable by any other like in the case of electronic coins of the same value, point in a loyalty program, etc. or be unique like a ticket for a specific product or service. This quality assists on the definition of different types of compensations for each pilot in DataVaults.

The integration of the wallet and the surrounding platform into DataVaults is performed through the private DLT. In this case, instead of the customer transferring funds to her account in the Exchange, a compensation is transferred by the Trusted DLT Engine to the private DLT. Then, the wallet receives a notification regarding the availability of a new compensation and initiates the process to create a token which represents the compensation. This process is similar to the withdrawal step described above, but the currency exchange is additionally recorded on the ledger.

At the payment step, there are several possible flows depending on the nature of the compensation. If the case of compensations of monetary nature, once the wallet of the merchant receives the token, it will request to redeem the token. The value will be transferred to the account of the merchant in the exchange. In the case of non-monetary compensations, for example a ticket for a specific service, the flow of value between the seeker transferring value to DataVaults platform, and the service provider accepting the token as payment, entails a parallel process of value transfer between the seeker and the service provider.

The spending activity is also recorded on the ledger to act as immutable record of the exchange.

This parallel process must be taken into account as a risk for the privacy of the customer, the individual, as personalized compensations or a low number of individuals sharing data to the seeker could easily drive to reveal the identity of the individual.

#### *6.2.2.1 Integration of the Blockchain Security 2Go Starter Kit*

Functionalities, which are responsible for compensation and the transfer of monetary values, are always a target for exploits and fraud. Secured key storage to prevent unauthorized access

and mechanisms to ensure the non-repudiation of transactions are essential to protect both the users and the platform operator.

The application of the *Blockchain Security 2Go Starter Kit* for secure key management and signing of Blockchain transactions was already described in earlier deliverables. It is a smart card, which generates and securely stores key pairs for the usage with ECDSA. The DataVaults project will also investigate new applications, where the functionalities are utilized aside from the default use case.

In the Personal Wallet, each token holds a key pair for signing the invoices when they are spent. A valid signature ensures that the corresponding coin was deliberately used by the legitimate owner. The key pairs can optionally be managed by the smart card to provide the highest level of key security. Because the private key is securely stored on the smart card, one disadvantage of the hardware solution is the limited size of the secure storage, providing only 255 unique key slots. The demo implementation will thus use a key rotation, were keys from already spent tokens will be re-used when necessary.

A new revision of the Starter Kit is currently being developed. In contrast to the earlier smart card, it is produced in the form factor of a chip with contact-based bus interface. DataVaults will investigate the feasibility of integrating this new product into a USB-based security key.

### 6.2.3 Findings from the consultation with Data Seekers

The Consortium conducted a survey to explore for gathering the view of stakeholders potentially interested in our research activities and outcomes and/or operating in the data economy.

The survey was directed to data seekers, which are “the stakeholders that are on the other side of the data subjects, asking for their personal data. They could be considered a recipient or a processor, depending on if they are going to just use the data (recipients that receive the data) or if they are going to process those data” (DataVaults D2.1 "Security, Privacy and GDPR Compliance for Personal Data Sharing", 2020).

Besides the DataVaults demonstrators, the respondents include other potential data seekers in their respective ecosystems, such as public institutions working in key sectors, such as mobility and culture (including two museums), manufacturing companies, service companies in the energy sector and others.

The survey, though addressing several topics relevant to a better understanding of the factors hampering the emergence and consolidation of a strong data economy in Europe in view of fully setting, sustaining and mobilizing an ever-growing ecosystem for personal data and insights sharing, is especially focused on the compensation mechanisms and on the most suitable approach for rejuvenating the personal data value chain, giving rise to a multi-sided and multi-tier ecosystem governed and regulated by smart contracts.

It comprises the following six categories of questions:

1- **Attitude towards compensation**, aimed at exploring the stakeholders’ attitude towards providing a compensation to the individuals providing their personal, as well as their availability to pay such a compensation, their perception on the adequateness

of existing business models in the data economy in terms of advantages for all the involved actors, on the most appropriate type of compensation (money, tokens, services, coupons…);

2- **Compensation modalities**, investigating the preferred modalities for the payment of the compensation to individuals providing their personal data, including for instance mechanisms such as "moneybag" as an alternative to single payments for any access to the data, as well as exploring the first reaction to the idea of using cryptocurrency as the payment method;
3- **Pricing parameters**, addressing the possible proper types of parameter of the provided data to be used to quantify the compensation amount (storage, data quality, data type, a mix of all) and deepening also some specific aspect, such as the attitude to pay more depending on data type / source (such as higher payment in case of provision of sensitive data, like health data) or depending to their quality, paying more for verifiably, integer and authentic data, or depending on the nature of actor providing data (public authorities and private sector);
4- **Fee for brokering services**, investigating the usefulness of brokering platforms like DataVaults to manage the negotiation between data seekers and data owners, the willingness to pay a subscription fee for such a brokering platform to get personal data for their business/institutional activity, as well as to pay a monthly member subscription fee to be a member that could buy data assets over the platform. It also examines the suitable percentage as a brokerage fee to pay to the platform for each data purchase transaction.
5- **Data and filters**, which lingers over the features of the data which make them more interesting, under different points of view, in particular distinguishing: i) eponymous data, anonymous data and data associated with given user categories; ii) their nature (social, health-related, activity, etc.); iii) the typology in terms of raw data and metadata, processed data, analysis and charts. Also the type of filters is taken into account (i.e. age groups, type of activity, etc.), as well as the type of information considered as more valuable (Financial, Health, Geolocation, Hobbies and interests,etc.) and the interest in metadata associated to processed data. Two more specific questions are also included, respectively addressing the potential added value to get information about new trends in swarm behavior in the transport sector and the potential new areas of interest in data due to COVID-19;
6- **Confidentiality**, where the perception and feelings in relation to confidential information are explored, in particular as regards the provision of information about the data seeker's data purchase transactions to the data owner and to all other entities that are operating over the platform.

This survey was not aimed to be representative, considering the limited number of data seekers to which it was circulated. Rather, it was aimed to provide useful insights and hints on the addressed topics, following the qualitative, explorative research methodology with a small sample of well-identified respondents. We sought to understand the given topic from their perspective and to gather information about the "human" side of it, also by identifying intangible factors which may not be readily apparent, in order to gain a rich and complex understanding of compensation mechanisms and related aspects relevant to DataVaults future development.

The outcomes and analysis of the survey are reported hereunder and they will drive the further design and development of DataVaults technology.

| Attitude towards data sharing | | |
|---|---|---|
| **Q1** | **Do you think that the business models in the data economy are correctly balanced in terms of advantages for all the involved actors? Please motivate your answer.** | Despite nowadays data is one of the most valuable resources in the world economy, it was unanimously recognized by the stakeholders that existing business models in the data economy are not correctly balanced in terms of advantages for all the involved actors. All the stakeholders agree on the fact that the business models are unbalanced: data providers are not appropriately gaining benefits in the actual business models, whilst the companies, especially big players, use the data to obtain a profit.<br><br>The most common business model, in fact, provides a service in exchange of data. It means that the<br><br>service has to be attractive for data owners to convince them to upload their data.<br><br>If the service is really attractive, owners tend to give too much data as they don't have a clue on what's the true value of such kind of data. In this scenario, data seekers get more from the owners.<br><br>On the other hand, a limited number of data seekers are struggling to get data from users because they don't offer enough added-value to convince owners to share data or the platform is not trusted by owners. This very unbalanced business model is perceived as detrimental for the data owners and possible free services do not compensate such deficit.<br><br>There is a lack of knowledge in the operators, besides the lack of transparency between all the parties involved: more specifically, the data providers normally do not know the final use of their data and it's not clear which data is captured and for which purpose. In this context, people are passive data donors, with neither economic advantage nor awareness on how data are used, whilst the platforms collecting data are often connected to other platforms which are unknown.<br><br>Some of them also underlined that the individuals as data owners in this way are not motivated to share their data, but if there were an attractive – trusted platform, with benefits for them, it would be easier for data seekers to attract more users.<br><br>Current policies are perceived as not effective in this regard. |

| | | |
|---|---|---|
| **Q2** | **Do you think that individuals providing their personal data should get a compensation for that? Please motivate your answer.** | Almost all the stakeholders perceive compensation as fair and agree on the opportunity to get compensation to the individuals for their personal data sharing, at least in case of data seekers that are profit entities like companies. Some of them pointed out that, on the contrary, in case of no profit data seekers (public bodies, association or other no profit organization), there is no reason to claim a compensation. Another viewpoint remarked that it depends on the type of personal data shared.<br><br>It was pointed out that individuals need an incentive to share personal data and, depending on the objective of this sharing, the rewarding mechanism can change: If data are used to improve a data seeker's service (provide custom insights to the individual himself/herself, comparison with others,etc.), data is not mandatory but it helps get better results out of the feature. In this case, the individuals already get some kind of compensation out of it by having more accurate insights, so more compensation is facultative (depends on how much data is needed to make the feature good enough). If data is needed and the added-value for the individual is not as visible (for example in a clinical trial), compensation can really help gathering enough information.<br><br>It was also stressed that, considering that individual do not know the real value of their data, they must have support and training. Other stakeholders argued that providing a fair compensation could represent a competitive advantage for the platform and, thereby, it could be an incentive to personal data sharing: if individuals should get a compensation for their data sharing, they would share their experience with more people, and it can increase the brand awareness and the positive impressions related the platform.<br><br>User databases are perceived as fundamental and the economic value of data is clear: as a consequence, it is seen as fair to reward data providers, although -at the moment- it is difficult to estimate their data value.<br><br>Another respondent argued that, rather than a compensation for the individual, it would be useful to get a compensation in terms of service for the society/community, like for example data availability for community needs. |
| **Q3** | **Which type of compensation do you think would be more** | Though there is not a consensus on the most appropriate type of compensation among stakeholders, most of them underline the suitability of providing service with rewarding function and |

| | | |
|---|---|---|
| | **appropriate (money, tokens, services, coupons…)?** | it is also common the viewpoint that economic compensation is not appropriate in all the circumstances and contexts.<br><br>As regards the monetary compensation for individual, one doubt is related to taxation because it is not clear if this kind of compensation is an economic income to be declared.<br><br>Some other clarified that it should depend on the preferences of each user: though money is a good "standard" and to some extent it could be considered as appropriate for all type of users, sometimes there are more interesting and appealing forms of compensation like services or coupons.<br><br>As for the medical field, though it is difficult to have only one answer due to the existence of different scenarios, most of the time, services related to shared data will be the most appropriate and money can also be appropriate in some use cases (like clinical trials).<br><br>In the sport sector, services and coupons are considered as more appropriate.<br><br>In relation to the public sector, discount to municipal services (theatres/museums) or access to information derived by the platform are good option. Instead of monetary schemes, which in general are not seen as feasible in this domain, the preference is to give back services (even as an alternative to shows currently provided), coupon (fidelity programme) and gadgets. On the other hand, it is also underlined that services or coupon could also be perceived more as a sop, while a monetary compensation would be very democratic and would better acknowledge for data value, despite it could be more complex to implement. It might be conceived as a final goal in a path.<br><br>Sometimes the respondent clarified that it depends on the subject acquiring the data and also on the market that might arise from the compensation approach: a given type of compensation could be appreciated for some time and then it could be necessary to change it, to ensure data owner's fidelity. |
| **Q4** | **Will you agree in paying a compensation to individuals to get their personal data?** | The unanimous answer is positive. Some stakeholders added explanations and clarifications, for instance that it is necessary in any case to have the individual's consent to the use of his/her data and that the willingness to pay depends on the usefulness of the data for them, such as the possibility to sell again the data later on or to use them. It is necessary that any |

| | | |
|---|---|---|
| | | limitation for the future use is communicated in advance to the purchase. The type of compensation depends on the kind of data and in some cases the willingness to pay regards only to the services/discount coupon or some sort of credit. In the public sector, it would be hard to provide monetary compensation due to internal/governmental procedures.<br><br>It is underlined that in any case data access has a cost and this should be compensated both to databases' managers and to data owners.<br><br>Andaman7 clarified that, as a data seeker, they will most likely use data collected from DataVaults to improve their service and provide new features, so other types of compensation will be limited. But in case they also act as an intermediary for partners that want to collect all kinds of personal data, an additional compensation can be paid and it depends on the partner. |
| **Compensation modalities** | | |
| **Q5** | **Would you prefer to pay directly to the individuals for their personal data or agree on a different payment mechanism? (e.g., flat rate, pay-as-you-go, etc.)** | The answers to this question were diversified, though the majority prefers to adopt different payment mechanisms, sometimes mentioning also loyalty programs and some other underlying that such preference is subject to the condition that the alternative payment mechanism is trustable and representative of all stakeholders' interests.<br><br>Some respondents don't have strong preferences on payment mechanisms and underlines that, as usage of the service may fluctuate depending on the usage of our platform and partnership, it is important that the amount paid match this usage and that limitations (if any) can be changed anytime and easily.<br><br>Others prefer to pay through a platform: some of them mentioned that it has to be capable of certifying the transaction, with individual payment for bought, whilst others clarified that it would be easier to pay through the platform, provided it's clear who operates it and who are those using the data (regulation). It was also remarked that, despite a direct contact might build a greater trust between the data seeker and the data owner, on the other hand, it might be difficult from the point of view of the privacy legislation, therefore the approach of different payment methods appears more feasible. |

| | | |
|---|---|---|
| | | On the contrary, some data seekers would opt to pay directly to individuals and one of them mentioned that this method looks like the easiest way and another underlined that in this way the transactions and payment flow would be transparent. |
| **Q6** | **Do you prefer to have a "moneybag" and movements in that bag or pay for every single access to the data?** | The majority of respondents prefer to use a “moneybag” and certain details were provided, such as the opportunity to have in it a carnet of possibilities, the comparison with loyalty programs and the usefulness of having a history in order to view the state of the moneybag in time, as well as the data purchased, also in view of correlating both.<br><br>A limited number of stakeholders has no a preference, because both of them are fine, though some of them mentioned that in case of several transactions or of a fixed budget is planned, it would be more useful to use the moneybag. Some others stated that both are relevant depending on the use case. If a data seeker should implement a new feature that collects data regularly and should not be limited in time/access/budget, single data access payment would be the most appropriate one. This can also be done through moneybag but will be more restrictive because we need to have alerts to fill moneybag if it becomes empty to avoid interruption of the service.<br><br>When payment depends on connected partners, moneybag seems to be the most relevant one.<br><br>Most projects are limited in time and budget (for example clinical trials) so having this budget limitation directly in DataVaults is a plus.<br><br>One data seeker prefers to pay for every single purchase of data, motivating with the need to have a control over the transactions. |
| **Q7** | **What is your first reaction to the idea of using cryptocurrency as the payment method? (Very positive, somewhat positive, neutral, somewhat negative, very negative)** | The majority of respondent is neutral on this regard, though someone underlined that at the moment there is not a good control by the government of the banks over this kind of payments and some other do not know the need to use such payment method.<br><br>It is important to mention that the reaction of some stakeholders is very negative and in some cases it was motivated with the fact that there are many cryptocurrencies, nowadays they are very volatile and some users don't understand them at all: in case of use of such kind of payment, the target of users would be limited and unbalanced rewards |

| | | |
|---|---|---|
| | | would be possible, depending on the fluctuation of the currency. It was also underlined that at the moment there is not a good control by the government of the banks over this kind of payments.<br><br>A couple of respondents were only "somewhat negative", remarking that cryptocurrencies are not stable in their value and can be risky. Only one stakeholder is positive, though it is perceived as an issue to be able to set the data price, due to the volatility of almost all the cryptocurrencies. |
| | **Pricing parameters** | |
| **Q8** | **Which parameter of the provided data would you use to quantify the compensation amount (storage, data quality, data type, a mix of all)?** | The preferred solution was the mix of different parameters, also because the estimated value can largely vary and this multiple reference could help to define and objective price for a piece of data.<br><br>An additional parameter could be the frequency of data update.<br><br>It was also pointed out that it depends on the use, though a mix of all parameters should be at least considered. As regards the relationship between quality and quantity, it should be preferred the quality in case of specific actions considered, whilst the quantity in case of general analyses.<br><br>On the other hand, many stakeholders recognized the importance of data quality and some also of transparency in order to have traceability. The quality is seen as a prerequisite of the usefulness of the data and, once achieved this, the compensation can be based on the data type and on the amount of the data. In addition, the data type is relevant. |
| **Q9** | **Do you agree to pay a different price depending on data type / source? (e.g. sensitive data is more expensive)** | All the respondents agree to pay a different price depending on data type /source. Some of them added as explanation that some data are more valuable than others. |
| **Q10** | **Would you pay more for data if they are verifiably, integer and authentic (e.g. from a certified source)?** | All the data seekers except one confirmed that they would pay more, though some of them pointed out that the platform should guarantee the veracity of the sources and that a minimum data quality should be guaranteed to been able to sell the data, so it should be given for granted that the data source is verified. Data sources are seen as one of the most important parameters to set the price. |

| | | |
|---|---|---|
| **Q11** | **Would you pay more for data from public authorities over data from the private sector?** | Though there were some confirmations, based on a sort of presumption of veracity of the data coming from the public sector, the opinion of the respondents is not unanimous. Most of them, however, are negative regarding this possibility. They underline that data sources should be associated with a certain level of certification that, despite potentially influencing the price of data, does not mean that data from the private sector are automatically less valuable than data from public authorities. What really might have an impact is the quality of the data and the certified sources, not the provenance from the public sector, which does not guarantee a higher value automatically. Others clarified that, in case of certified sources from the private sector, these would be preferable in respect to those from the public sector. |

## Fee for brokering services

| | | |
|---|---|---|
| **Q12** | **Do you think a brokering platform like DataVaults could be a useful tool to manage the negotiation between data seekers and data owners? Please motivate your answer.** | All the data seekers except one confirmed the usefulness of a platform like DataVaults. Some of them mentioned that the platform should manage the searches and purchase in an easy way and provide a common repository, so to concentrate the data in it, unifying the process if different types of data are required or they come from different sources. Another stated that all the brokering platforms are useful, but the success of each of them is determined by the functionalities and fees. It was also stressed as a positive factor the fair relationship between the data seekers, receiving useful data from the platform, and the data owners, receiving a compensation for sharing data. The relevance of the user-friendliness (both for data owners and for data seekers) and the existence of clear and transparent terms of use were mentioned as well, together with the importance of the reliability of the third party managing the platform and the simplification of data searching. It is valuable to have a unique contact point to get high quality certified data, capable of facilitating the data economy process and giving guarantee, besides offering the possibility of standardizing data value.<br><br>One of the respondents specified that such platform is useful for all data seekers that:<br><br>- require limited amount of data and/or<br>- need data for a short period of time and/or<br>- need data fast and/or<br>- don’t have enough resources to build their own cloud-based data collection platform |

| | | |
|---|---|---|
| | | - and/or<br>- need specific pieces/type of data from various sources<br><br>One of the stakeholders, on the other hand, was reticent to use the platform and underlined the legal basis should be guaranteed and the country where it hosts the server and the storage and operations should be national or European. |
| **Q13** | **Would you pay a fee to subscribe in a brokering platform like DataVaults to get personal data for your business/institutional activity?** | Though the confirmation seems slightly prevailing (sometimes underlying that usually in this kind of platform functionalities and services to improve the business model are provided), there were several clarifications about the conditions/limitations. For instance, it was mentioned that it would depend on the purpose, the budget planned to be spent in the platform and the profitability. Others mentioned that they would pay only in case the functionalities of the platform are useful.<br><br>There are also data seekers who would prefer to pay only when they use the platform.<br><br>As for the amount, some respondent declared maximum 3%-5% of the total spent, some other 0,5-0,7% of the total spent, another one stated that it depends also on expected results and that, generally speaking, they imagine a fee of 5.000 – 10.000 euros per year. Some respondents are favourable to pay a fee under 5.000/6.000 euros and one under 10.000 euros/year.<br><br>There were also some data seekers not willing to pay: A7 clarified that they would not pay a monthly membership subscription, whilst a brokerage fee or combination of both seems more appropriate for a cloud-based storage/computing platform. They believe customers prefer paying only when actually getting data, with a percentage going to the brokering platform for the service. Andaman7 as such will probably not buy data. A7 position themselves as a way to get in touch with patients, who will then share their data via our platform. Also Olympiacos as data seeker would probably not pay a fee for registration, but all partners – companies who need these data- would pay a small amount. |
| **Q14** | **Do you agree to pay a monthly member subscription fee to be a member that could buy** | The answers were diversified. Some respondents are positive in this regard (also underlying that a monthly fee could be a useful entry approach, to check the real advantage the platform can provide), others confirmed what mentioned in relation to the previous question, others were negative and specified that they would pay per use and based on their |

| | | |
|---|---|---|
| | **data assets over the platform?** | customers' requests. If use is very frequent, then a monthly fee could reduce costs.<br><br>A yearly fee seems to be slightly preferred, instead of a monthly fee, though it was also mentioned that it would be more reasonable to pay according to needs.<br><br>As for the amount, there were only two specific answers, which are really different: about 25 euros/month and 250 euros/month. |
| **Q15** | **How big (%) of a brokerage fee are you willing to pay to the platform for each data purchase transaction?** | The respondents were unanimous in replying yes, though the amount suggested range from 1% to 20%. It was specified by one stakeholder that it would be useful to have the brokerage fee have a lower and an upper limit. Another clarification, in line with this, pertains to the fact that the fee could be variable, with bounds: 1% to 3% seems reasonable, but in case of a big amount, there should be a maximum in order to avoid that the profit of the platform is too high. One respondent clarified that it is better to compensate the platform rather than data providers (90% and 10%). |
| | **Data and Filters** | |
| **Q16** | **Which form of personal data are you most interested in (eponymous data, anonymous data, data associated with given user categories, all)?** | It was often mentioned that all forms are interesting, depending on the use case and of the needs. Some respondents gave a preference to eponymous data, others to data associated with user categories in order to know the behaviour of groups and and data aggregated by category. Also, anonymous data are perceived as relevant. |
| **Q17** | **What type of personal data are you more interested in (e.g. social, health-related, activity, etc.)?** | The given answers were mainly related to the area of activity of the data seeker and so they are very different and depend on the type of analysis needed. They range from health related and activity tracking data, to data related to energy consumption data or in order to offer additional services, users' activities and social information. For instance, the municipalities stated that any information derived from any kind of data is potentially useful, especially social, transport, tourism related data, data associated with city mobility.<br><br>Some respondents perceive genetics, health and financial data as more interesting, others mentioned also social data and activity data. Social data, data on cultural activities and personal interests, data related with people's life habit, data on cultural interests, age, education, residence and economic data are other types identified as potentially relevant. The |

| | | |
|---|---|---|
| | | same occurs with social profile, education degrees, earnings, family composition, disabilities, energy consumption and vital statistics. |
| **Q18** | **What type of filters you need to get the data you would like to access (i.e. age groups, type of activity, etc.)?** | The replies were diversified, though many respondents identified the type of activity and the age as useful filters.<br><br>Also, the financial information, such as income, were widely mentioned. The location, gender and level of studies, as well as occupation were identified by some respondents, as well as the type of cultural interests and the geographic area of provenance.<br><br>A7 underlined that groups should correspond to the health profile of the user, based on age, demographic data, ethnicity, one or more condition(s) they are suffering from, symptoms and similar aspects. |
| **Q19** | **What data are you interested in: Raw data and metadata, processed data, analysis and charts?** | For some of the respondents all types of data are interesting, whilst others are particularly interested respectively on processed data, data analysis, charts, raw data and metadata (mentioning that the reason is that they can be analysed further). As for the the medical field, the respondent underlined that all are interesting: processing of data can be complex, therefore it would be an advantage if the tool can provide some basics. In addition, some services may directly use processed data, analysis and charts but some others (for example clinical trials) will need raw data anyway. |
| **Q20** | **What information do you consider more valuable (Financial, Health, Geolocation, Hobbies and interests,…)?** | It was pointed out that the value depends on what the data seeker has to do, though in general the more difficult data to obtain might be considered more valuable.<br><br>Geo-localization data are widely mentioned (and someone clarified that residence is relevant above all, rather than having a continuous data flow), as well as hobbies/interests and health data. Other choices were financial information (for claims), tourist information, cultural information and life habits. |
| **Q21** | **Are you interested in metadata associated to processed data (e.g. number of data collected, number of data owners)?** | Most of the respondents are interests in metadata associated to the processed data, in order to understand the context: such kind of metadata can provide very useful information on what was processed and can lead to further and more precise requests. One of the respondents explained that they are interested only in the big numbers (number of data collected and the one that matched the filters). |

| | | |
|---|---|---|
| **Q22** | **Would it be an added value to get information about new trends in swarm behavior in the transport sector?** | The majority of respondent stated that it is not interesting in their case and for their potential partners. One of them specified that there might be interest in local behaviours like routes for electric vehicle. |
| **Q23** | **Do you see new areas of interest in data due to COVID-19 (e.g. motion tracking profiles)?** | The majority foresees that there will be new areas of interest in data due to the pandemic, especially in the future. Health and tracking are increasingly interesting.<br><br>It was explained that for the moment there are country specific applications that can track contact and notify users of potential exposure. When travelling will be widely allowed, those apps may not talk to each other, and the tracking will not be relevant: having an European / worldwide database tracking close contact could help getting to the next step of COVID-19 prevention. It could be also very useful to gather vaccination status.<br><br>Some data seekers also perceived that new opportunities might arise from the increasing time spent at home by people, for instance data could be gathered on people's behaviour at home, as well as analysis of changes in consumers' behaviour or in changes in transfer modalities. There is also the opportunity of improving the knowledge of citizens' behaviour and cultural interests, while too often this is focused on tourists. Other possible areas of interest identified by the stakeholders include the analysis of changes in the people's work habit and the analysis of air quality data from domestic sensors. |
| | **Confidentiality** | |
| **Q24** | **Are you ok with the platform revealing your data purchase transactions to the owners of those data?** | Most of the data seekers are favourable, also motivating with the fact that it is a prerequisite for trust building with the data owners and to promote their data sharing.<br><br>Only two respondents were negative on this regard. |
| **Q25** | **Are you ok with the platform revealing your data purchase transactions info (without revealing the amount paid) to all other entities that are** | There isn't a common view of this aspect. Around half of the respondents are against this disclosure of information, because the other companies could detect their commercial plans and because this kind of information has an important value for possible competitors. One is against unless such information is anonymously grouped by type of data seeker. |

| | | |
|---|---|---|
| | **operating over the platform?** | The other half of respondent is favourable but does not elaborate on the reasons. |

## 6.3 DATA VALUE FLOWS

In DataVaults, a wide variety of data flows are envisaged, as the overall infrastructure is facilitating the collection, transformation, sharing and analysis of data. Therefore, one can expect that data is the core element that is flowing from each component to another.

In parallel, this flow of data is signaling also a flow of value, between the main stakeholders that are part of the main data transactions; the Data Owners on the one side, and the Data Seekers on the other. All this data value flow is facilitated by the different components and governed by the introduction of distributed ledger technologies and smart contracts, which are not only used for auditing reasons regarding the activities that are performed over the platform, but also for the flow of compensation between those two different stakeholders. In this flow, DataVaults plays a central role, as it does not only act as a trusted authority over which the transactions are executed (e.g., there are no peer-to-peer transactions for various privacy and trust issues), but also assumed the role of a broker, which can demand a stake of the overall value, as compensation for the service it provides.

The following section provides the main flow of the data as well as of the value in a transaction, in an end-to-end scenario executed over DataVaults. For a visual representation of these flows, the core architectural diagram is presented below, illustrated using with red arrows for the flow of payload of data, with orange the flow of information relevant to how the data asset should be handled (e.g. From a security and privacy perspective (e.g. access policies, smart contract information, etc.), and blue arrows and badges for the flow of value. It needs to be noted that two distinct value flows are identified. The first one (Value Flow #1) has to do with acquiring data that is already shared over the platform and is readily available to be purchased. and has already a fixed price attached to it. In such a case, the Data Seeker designs a contract which is automatically executed so that the respective compensation can flow back to the Data Owner. The other value flow identified (Value Flow #2) concerns the case where a Data Seeker makes a custom request to a Data Owner for some data. In such a case, the final compensation contract is executed upon the Data Owner's acceptance and execution (or rejection) of a sharing configuration proposal.

**Table 7. Data and Value Flows**

| Actor | Data Flow | Value Flow #1 | Value Flow #2 |
|---|---|---|---|
| Data Owner | 1.a A Data Asset is collected from external data sources and stored in the Personal App | | |

| Actor | Data Flow | Value Flow #1 | Value Flow #2 |
|---|---|---|---|
| Data Owner | 1.b Data Asset is collected as part of a questionnaire (go to step 4) | | |
| Data Owner | 2. Data Asset can be analysed for a new data asset to be generated | | |
| Data Owner | 3 Data Asset is configured for sharing | | |
| Data Owner | 4 Data Asset is shared and stored in DataVaults<br><br>*Security Options:*<br>4a – Data is anonymised if the user chooses so<br><br>4b – Access Policies are applied on the Data<br><br>4c – Data is merged in a persona if the user chooses so | | |
| Data Owner | 4.1 The Sharing Configuration provided to the DLT Engine to create the necessary smart contracts | | 2.b The contract is executed in the DLT Engine |
| Data Owner | 4.2 The Sharing Configuration is stored in the Private DLT | | |
| Data Owner | 4.3 The Sharing Configuration, minus personal information is stored in the Public DLT | | |
| Data Seeker | 5 The Data Seeker retrieves the information about a Shared Asset from the Query Builder | | |
| Data Seeker | 5.1 The relevant configuration parameters of that file are retrieved from the public ledger | | |
| Data Seeker | 5.2 The relevant access policies are applied on the file based on the saved configuration parameters | | |
| Data Seeker | 5.3 The relevant decryption methods are performed based on the saved configuration parameters | | |
| Data Seeker | 6 A contract is designed so that the Data Seeker can acquire the data asset | 1. A contract is designed. | 1. A contract is designed |

| Actor | Data Flow | Value Flow #1 | Value Flow #2 |
|---|---|---|---|
| Data Seeker | 7 The contract is executed in the DLT Engine | 2. The contract is signed and the Data Seeker pays out the compensation from his wallet to the DataVaults platform wallet | 2a. The offered compensation is moved and blocked in the DataVaults Platform.<br>(go to 2.b) |
| Data Seeker | 7.1 The contract is written in the Public ledger | 4. The Data Seeker pays out the compensation from his wallet to the DataVaults platform wallet | 3. The Data Seeker pays out the compensation from his wallet to the DataVaults platform wallet |
| Data Seeker | 7.2 The contract is written in the Private ledger | 4. The DataVaults platform, pays out the compensation (minus a fee) from the DataVaults wallet to the Data Owner's wallet | 4. The DataVaults platform, pays out the compensation (minus a fee) from the DataVaults' wallet to the Data Owner's wallet |
| Data Seeker | 8. The Data Asset is made available in the Data Seekers Vault | | |
| Data Seeker | 9. The Data Asset can be used as input for Analytics | | |

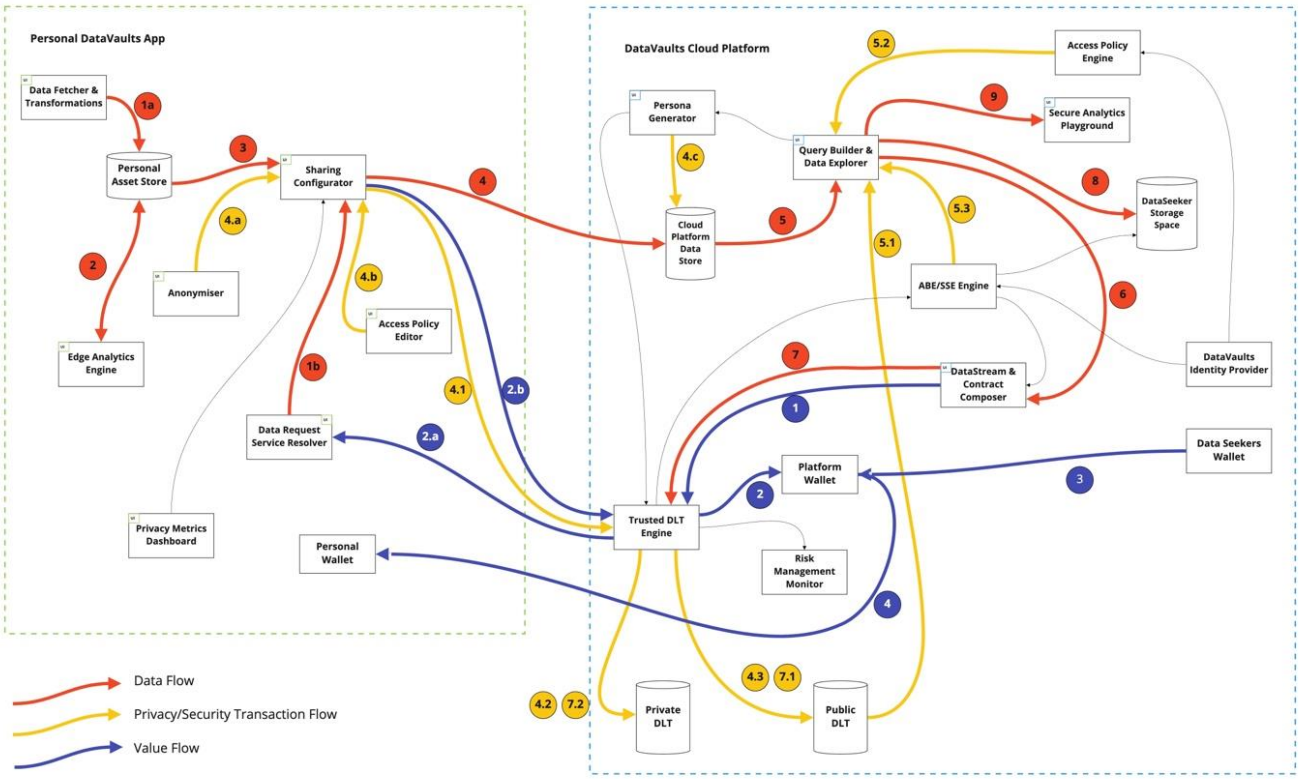


**Figure 9 – DataVaults Data and Value Flows**

# 7 CONCLUSIONS

This document outlines the key outcomes of WP2 "Security Aspects, Privacy Considerations, Value Generation and Commercialisation Outlines in Personal Data Management" and sets the basis of the holistic DataVaults Data Security and Privacy Framework driving further project progress within WP3 " Bundles for Secure Data Sharing and Access, Privacy and Trust Preservation and IPRs Management", WP4 "Multitude Trusted Intelligence Bundles for Personal Data Insights Generation" and WP5 "DataVaults Platform Continuous Integration".

The set of legal, ethical, security, privacy and trust requirement for DataVaults cloud-based platform and Personal App has been finalized, relying both on the enriched regulatory framework (both already applicable and under development) and on the analysis of the relevant features of the technology under development.

Concerning the regulatory landscape, the deliverable contains references to several important concepts and principles retrieved in numerous pieces of legislation which have to be considered in the course of the project and development (and/or future uptake) of DataVaults solutions, though the main regulatory source for DataVaults remains the data protection regime. In particular, the key instrument is the GDPR and its basic concepts, such as the transparency principle, the lawfulness of the processing which requires that the processing activities must be legitimate and rely on a valid legal basis, such as the individual's consent, the accountability principle and the risk-based paradigm.

The future regulatory development under development, as planned by the European Strategy for Data, such as the Data Governance Act and the Digital Service Act, should be monitored,

also in order to align the final results of the project to them: this is expected to facilitate a wider adoption and sustainability of DataVaults technology.

As underlined in D2.1, the common ground of the technological choices and requirements setting is the strong commitment to operationalize the “sharing the wealth” paradigm and to contribute to move ahead in the direction of a win-win data sharing ecosystem in view of unlocking the social value of personal data and fostering individual human flourishing, together with their business and economic value and the same is at the core of DataVaults vision. This approach is strongly consistent with the vision behind the European Strategy for Data and the 2030 Digital Compass Communication. These legal instruments are setting the scene for the advancement of the data economy and for the growth of the data sharing environments across Europe: the recent regulatory reforms which are expected to have an impact on the future development and uptake of DataVaults, such as the Data Governance Act and the Digital Service Act, have been launched under the umbrella of such legal sources, going beyond GDPR compliance towards a value-driven and human-centric data-driven ecosystem: DataVaults technological-empowered balancing operations and the underlying multi-layer approach in data sharing, within the boundaries of the rule of law and the ethical ground, are contributing to properly face with the management of privacy / utility trade-offs. Personal data can be widely accessed and shared for the benefit of the overall society and of the European undertakings, despite still remaining in full control of their owners.

On this regard, it has to be remarked that DataVaults also employs strong crypto primitives towards the secure and privacy-preserving platform authentication as well as the secure and anonymized access control and interaction with the underlying DLT infrastructure. In this context, this deliverable presented the remote attestation mechanisms (Configuration Integrity Verification and Direct Anonymous Attestation) that are leveraged and the offered capabilities for the establishment of a secure communication channel between the Data Owners and the DataVaults platform itself; such advanced attestation variants are used in DataVaults for enabling Data Owners to both authenticate their platforms in a privacy-preserving manner but also to share their data in an anonymous way by leveraging group-based pseudonyms.

Furthermore, trusted and secure data sharing is enabled by storing the sharing configuration and access policies in immutable smart contracts. The combination of ABE and SSE, as well as the compensation flow through the private wallet, ensure data confidentiality and privacy of the Data Owners. The results from the consultation with the Data Seekers validate the goals of the project and give a good indication about future strategies for marketing and exploitation.

# 8 REFERENCES


[1] B. Larsen, H. Bergsson Debes e T. Giannetsos, «CloudVaults: Integrating Trust Extensions into System Integrity Verification for Cloud-Based Environments,» 2020.

[2] Datavaults Consortium, «D1.2 Datavaults core datamodel,» 2020.

[3] A. Baliga, I. Subhod, P. Kamat e S. Chatterjee, «Performance Evaluation of the Quorum Blockchain Platform,» 2018.

[4] D. Chaum, «Blind Signatures for Untraceable Payments,» in *Advances in Cryptology*, Boston, 1983.

[5] «FENTEC Project Website,» [Online]. Available: http://fentec.eu/.

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

[7] The DataVaults Consortium, "D2.1 - Security, Privacy and GDPR Compliance for Personal Data management", 2020.

[8] The DataVaults Consortium, "D2.2 - Personal Data Market Design, Contracts and Rules", 2021.

[9] Benjamin Larsen, Heini Bergsson Debes, and Thanassis Giannetsos. Cloudvaults: Integrating trust extensions into system integrity verification for cloud-based environments. In European Symposium on Research in Computer Security, pages 197–220. Springer, 2020.

[10] J. Whitefield, L. Chen, T. Giannetsos, S. Schneider and H. Treharne, “Privacy-Enhanced Capabilities for VANETs using Direct Anonymous Attestation,” in IEEE Vehicular Networking Conference (VNC), 2017.

[11] B. Larsen, T. Giannetsos, I. Krontiris, K. Goldman, ”Direct Anonymous Attestation on the Road: Efficient and Privacy-Preserving Revocation in C-ITS”, In WiSec 2021.

[12] N. Koutroumpouchos, C. Ntantogian, S. Menesidou, K. Liang, P. Gouvas, C. Xenakis, and T. Giannetsos, “Secure edge computing with lightweight control-flow property-based attestation,” in 2019 IEEE Conference on Network Softwarization (NetSoft), 2019, pp. 84–92.

[13] Thanassis Giannetsos, Tassos Dimitriou, Ioannis Krontiris, and Neeli R. Prasad. Arbitrary code injection through self-propagating worms in von neumann architecture devices. Comput. J., 53(10):1576–1593, December 2010.

[14] Thanassis Giannetsos and Tassos Dimitriou. Spy-sense: Spyware tool for executing stealthy exploits against sensor networks. In Proceedings of the 2Nd ACM Workshop on Hot Topics on Wireless Network Security and Privacy, HotWiSec '13, pages 7–12, 2013.

[15] Cucurull, J and Guasch, S.: "Virtual TPM for a secure cloud: fallacy or reality?" Scytl Secure Electronic Voting, 2014.

[16] TCG. 2016. Trusted Platform Module 2.0, Part 1: Architecture. Rev 1.38. Trusted Computing Group

[17] Marcos Allende Lopez, "Self-sovreign identity. The future of Identity: self-sovreignity, Digital Wallet, Blockchain", 2020

[18] Domingo, Ignacio Alamillo. 'SSI EIDAS Legal Report - How EIDAS Can Legally Support Digital Identity and Trustworthy DLT-Based Transactions in the Digital Single Market', April 2020

[19] Sovrin, Sovrin Trust Assurance Framework, 2019. Retrieved from https:// sovrin.org/wp-content/uploads/Sovrin-Trust-Assurance- Framework-V1.pdf

[20] Smart Contract Alliance, "Smart Contracts: is the Law Ready?", 2018

[21] Communication of the Commission of 11 April 2018—A New Deal for Consumers, (COM)2018, 183 final

[22] COM(2021) 118 final. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, "2030 Digital Compass: the European way for the Digital Decade"

[23] COM (2020) 67 final "Shaping Europe's digital future"

[24] COM(2020) 825 final. Proposal for a Regulation of the European Parliament and the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC

[25] COM(2020) 790 final on the "On the European democracy action plan"

[26] COM(2020) 842 final. Proposal for a Regulation of the European Parliament and the Council on contestable and fair markets in the digital sector (Digital Markets Act)

[27] COM/2017/010 final. Proposal for a Regulation of the European Parliament and the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC

[28] European Data Protection Supervisor, “Opinion 5/2021 on the Cybersecurity Strategy and the NIS 2.0 Directive”, 2021

[29] EDPB-EDPS Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act) - Version 1.1, 2021

[30] EC, “Impact Assessment on enhancing the use of data in Europe. Report on Task 1 – Data Governance”, 2020

[31] BDVA/DAIRO “Towards a European-governed data sharing space. Enabling data exchange and unlocking AI potential” Position Paper v2, 2020