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

D2.1
Security, Privacy and GDPR Compliance for Personal Data Management

Editor(s)	Marina Da Bormida (ETA), Thanassis Giannetsos (DTU)
Lead Beneficiary	DTU
Status	Draft
Version	1.0
Due Date	31/03/2020
Delivery Date	06/03/2020
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	D2.1 – Security, Privacy and GDPR Compliance for Personal Data Management
Editor(s)	Marina Da Bormida (ETA), Thanassis Giannetsos (DTU)
Contributor(s)	Sotiris Koussouris (SUITE5), Miguel Angel Mateo Montero (ATOS), Maria Jose Lopez Osa (TECNALIA), IFAG, Christina Tsilihkiri (OLYMPIACOS), Michail Bourmpos (PIRAEUS), Sébastien Hannay (ANDAMAN7), Ramon Ruiz (MIWENERGIA), Elena Palmisano, Paolo Boscolo (PRATO)
Reviewer(s)	ATOS, Stefanidis Kyriakos (FRAUNHOFER)

Abstract	Legal and ethical requirements, focusing on GDPR, related to personal data management, and DataVaults planned activities and tools for addressing the related challenges. SotA analysis of anonymization/pseudonymization methods, techniques and algorithms, TPM technology approaches to security, authentication and attestation. Preliminary insight for approaches DTL and smart contracts for personal data management.

Executive Summary

The vision of DataVaults is to provide a secure, trusted, auditable and privacy-preserving platform for data sharing economies that complements existing ICT deployments through the use of Blockchain/Distributed Ledger Technologies (DLT). This will enable enhanced data privacy and ownership safeguarding (privacy by design) and data provenance and sovereignty checking mechanisms, whilst respecting prevailing GDPR legislation.

This deliverable outlines the main findings in terms of analysis of the legal and ethical requirements that DataVaults needs to adhere to, focusing on GDPR and the other privacy and data protection legislation. This analysis relies on the initial snapshot of the privacy-relevant properties and personal data collection, processing and sharing in each service and tool, as well as details on the data categories, data sources and purposes of processing.

On the basis of both the analysis of the regulatory landscape and the factual description, the legal and ethical requirements are elicited, via a systematic and holistic approach, driven by Fairness & Privacy-by-Design-and-by-Default enriched with the Protection Goals method.
The fulfilment of these requirements will ensure that the research activities, results and validation activities are legally compliant and ethically sound. Furthermore, recommendations and preliminary insights on how to face with the identified boundaries and constrains are provided. In the next deliverables of this WP, if necessary, the legal analysis will be integrated with potentially further applicable sources, such as the Law on trust services and electronic identification, and the list of requirements consequently updated.

Reflecting on DataVault's work and data flow and how data security, privacy, sharing and management services are to be engrained in a policy-compliant Blockchain structure, this document puts also forth the technical security, privacy and trust requirements for the DataVaults platform. This includes an analysis of how crypto algorithms (focusing on the data anonymization and pseudonymization techniques, the data protection algorithms, as well as on the encryption and authentication techniques) and trusted computing technologies can be enhanced towards achieving the vision of DataVaults. Requirements, with a view on an enhanced (holistic) data sharing solution have been categorized as *mandatory* and *desirable*.

All the requirements elicited in the document, which will be considered during the design of the platform and app, represent a first version and might be updated according to project's progress, once its services, solutions and demonstrators are better shaped, and to the extent this will be permitted by the legislation and technical advancements.

Furthermore, preliminary considerations on Distributed Ledger Technologies (DLTs), smart contracts for access and usage policy management (when it comes to data sharing) and compensation services for fair and secure personal data sharing and management of transactions in DataVaults are outlined.

The overall purpose of this deliverable is to provide a reference document for the DataVaults project to be used as input to the platform's architecture definition, the functionality of the platform's sub-components and the further investigation, design and development of the

core security, privacy and trust services towards the support of enhanced data sharing economies. Considering the initial stage of development of the project, this document, which is strictly interrelated with WP3, 4, and 5 might be updated, integrated and refined in D2.2, according to the future project's progress.

Table of Contents

List of Figures

List of Tables

Terms and Abbreviations

ABE	Attribute-based Encryption
DAA	Direct Anonymous Attestation
DoA	Description of the Action
SoTA	State of the Art
GDPR	General Data Protection Regulation
DFD	Data Flows Diagram
DLT	Distributed Ledger Technology
DPIA	Data Protection Impact Assessment
IoT	Internet of Things
WP	Work- Package
DTL	Distributed Ledger Technologies

BDVA	Big Data Value Association
EDPS	European Data Protection Supervisor
PCRs	Platform Configuration Registers
PID	Personal Information Diagram
PET	Privacy-enhancing Technologies
SC	Smart Contract
TPM	Trusted Platform Module

1 INTRODUCTION

This deliverable has a threefold objective, aimed at reporting the work and findings for:

- identifying, depicting and analysing the regulatory framework relevant to the project, with a special focus on the General Data Protection Regulation and how to address the related challenges and requirements for personal data management;
- outlining the SoTA analysis for the selection of the data anonymization models and the data privacy technologies for the definition of the process of enabling end-to-end security, privacy and intelligent handling of personal information, towards the definition of the holistic DataVaults Data Security and Privacy Framework. In particular the SotA addresses:
    - data anonymization and pseudonymization techniques;
    - data protection algorithms;
    - encryption and authentication techniques;
- providing first insights on possible useful approaches, technologies, tools and frameworks for smart contracts and DLT for fair and secure personal data sharing and management of transactions.

All the requirements elicited in the document, which will be considered during the design of the platform and app, represent a first version and might be updated according to project's progress, once its services, solutions and demonstrators are better shaped, and to the extent this will be permitted by the legislation and technical advancements.

The document, and the related research activities, are interrelated with most of the WPs and tasks, and in particular with:

- WP3 "Bundles for Secure Data Sharing and Access, Privacy and Trust Preservation and IPRs Management" and WP4 "Multitude Trusted Intelligence Bundles for Personal Data Insights Generation", because in them key bundles will be implemented, such as:
    - the security modules, assuring trusted and secure communication between the Personal DataVault and the DataVaults cloud-based engine, and the bundles to undertake attribute-based data asset and analytics access policies (WP3);
    - the multitude trusted intelligence bundles for personal data insights generation (WP4).

  The design and development of such and other bundles will be driven by the legal and ethical requirements, as well as by the security, privacy and trust requirements, as depicted in this deliverable;
- WP5 "DataVaults Platform Continuous Integration", because the findings and requirements set in this document (likely to be updated in the course of the WP2 activities) will be reflected in the definition of the DataVaults platform architecture, as well as in the the platform integration and testing;
- WP6 "Multi-Layer Demonstrators Setup, Operation and Business Value Exploration", because it is necessary to take into account the outcomes of this deliverable in the set up and execution of the different DataVaults demonstrators' cases, as well as, on

the other side, it is important that demonstrators' assessment and lessons learnt also cover human well-being and empowerment.

- T9.3 "Ethics Requirements and Project Data Management" and WP10 "Ethics Requirements", because they refer to ethics requirements, are also aimed at exploring the societal consequences of DataVaults and consider ethical issues in a more comprehensive manner, in line with the Fairness & Privacy-by-Design-and-by-Default approach, enriched with the Protection Goals method, identified by this deliverable.

1.1 DOCUMENT STRUCTURE

The document is structured as follows:

- **Section 2** provides an analysis of the regulatory and ethical instruments relevant to DataVault personal data management in terms of data collection, sharing and processing, both during the project development phase (including research, demonstration activities and results), and after the end of the project, in the exploitation phase. The main sources of legislation that were analysed and deepened are mentioned, in particular in the field of Privacy and Data Protection Law, Human Rights Law and Ethics and Soft Law, as well as the national data protection regulatory framework applicable to the demonstrators. In the next deliverables of this WP, if relevant considering the main technical choices that will be taken, other possible sources could be studied for eliciting further requirements, such as the Law on trust services and electronic identification;
- **Section 3** reports the description of the facts and aspects of the project relevant in order to provide the legal analysis and to elicit the legal and ethical requirements. The description includes a first snapshot of the privacy-relevant properties and personal data collection, processing and sharing in each service and tool, as well as details on the data categories, data sources and purposes of processing. This factual description will be updated according to the project's progress;
- **Section 4** firstly sets out the list the legal and ethical requirements to DataVaults design, development and operation and then proceeds to layout the security, privacy and trust requirements elicited from a technical point of view;
- **Section 5** complements the previous ones with a SoTA for the selection of the data anonymization models and the data privacy technologies for the definition of the process of enabling end-to-end security, privacy and intelligent handling of personal information, towards the definition of the holistic DataVaults Data Security and Privacy Framework. In particular the SotA addresses:
    - data anonymization and pseudonymization techniques;
    - data protection algorithms;
    - encryption and authentication techniques;
- **Section 6** outlines the preliminary considerations in relation to the DataVaults Distributed Ledger technologies (to be considered in detail in WP4) and smart contracts used for policy management (when it comes to data sharing) and compensation services for fair and secure personal data sharing and management of

transactions in DataVaults, in order to strengthen trust – to be aligned with the security, privacy and trust requirements described in Section 4. The detailed study and SotA will be part of D2.2;
Section 7 draws conclusions.

2 THE REGULATORY AND ETHICAL REFERENCE FRAMEWORK

The following paragraphs provide a snapshot of the regulatory and ethical framework relevant to DataVaults, considering the applicable instruments in a systematic way. Such instruments have been analysed carefully to elicit the legal and ethical requirements are described in section 4.1. The purpose is to deliver an ethical, privacy and fairness-friendly framework and platform, that is at the same time compliant with the legislation, and where individuals are enabled to take ownership and control of their data and share them at will, whilst value is properly attributed to all the entities involved in generating the same.

This section doesn't present a comprehensive analysis of the European regulatory framework, that would fall outside the scope of this document. It indicates the main instruments that are functional to the objective mentioned above, focusing on the privacy and data protection legislation and soft law, bearing in mind that down the line as the project progresses, other areas of law, such as Telecommunication law & IT-security law, IP law and Law on trust services and electronic identification, might need to be investigated for eventually eliciting additional requirements

2.1 PRIVACY AND DATA PROTECTION LAW

As regards this area of law, the two main sources that have been analysed (and will be further investigate throughout the development of the project) are:

GDPR, "General Regulation on data protection"

The first piece of legislation to mention is the GDPR, "General Regulation on data protection 2016/679, of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data"[1]. It repealed the Directive 95/46/EC (General Data Protection Regulation), providing a comprehensive reform of data protection rules in the EU, establishing common European rules to ensure that personal data enjoys a high standard of protection everywhere in the EU.

One of the main objectives of the reform is to give individuals back control over of their personal data, thus acting as key enabler of the Digital Single Market: personal data can only be gathered and handled legally under strict conditions and for a legitimate purpose. The individuals or organisations collecting or managing personal information have to protect it from misuse and have to respect data subject's rights, the data subject is enabled to complain and obtain redress if his/her data is misused. GDPR outlines key definitions relevant to DataVaults (such as "personal data" and "processing") and its articles are key for identifying the legal constraints that have to comply with.

The main findings from the analysis of GDPR provisions in relation to DataVaults research, demonstration and uptake will be described in the Legal and Ethical Requirement list, as well as will be addressed in other parts of this document (such as Section 3.1.4).

Directive 2002/58/EC "ePrivacy Directive"

[1] It can be retrieved for instance at this link: https://eur-lex.europa.eu/eli/reg/2016/679/oj

The second instrument relevant to DataVaults is the "ePrivacy Directive" (Directive 2002/58/EC on privacy and electronic communications[2]), which replaced the Directive 97/66/EC and was partially amended by Directive 2009/136/EC.
It pertains to the processing of personal data and the protection of privacy in the sector of electronic communications, telecommunications networks and internet services, transposing in the telecommunications sector, which is a "sensitive" area from a privacy perspective, the main principles and rules of the GDPR, aiming at particularising and complementing the former (for instance as regards the consent to the use of cookies and opt-outs) in case electronic communications data are personal data.
Several provisions are relevant in relation to DataVaults, such as Article 4, on the obligation of adopting security measures appropriated to the risk presented, Article 5, dwelling on the protection to confidentiality of the communications among individuals, Article 2, on the traffic data and location data, Article 6 on user's consent, Art. 15 on data retention, and others.
The ePrivacy Directive is expected to be repealed by the e-Privacy Regulation: the European Commission adopted a proposal, which is currently under discussion in the European Parliament and the Council of the European Union. The current draft is likely to be subject to further changes due to the concerns expressed by relevant stakeholders (such as the Article 29 Working Party, the European Data Protection Supervisor Giovanni Buttarelli).

2.2 HUMAN RIGHTS LAW

Among other sources, attention is focused especially on the European Convention of Human Rights[3] and Charter of Fundamental Rights of the European Union[4].
Both of them acknowledge privacy and data protection as fundamental human rights in Europe. From an international perspective, the Universal Declaration of Human Rights (1948) recognises the privacy as a fundamental human right by protecting territorial and communications privacy.
Article 8 of the European Convention for the protection of Human Rights deals with private and family life, home and correspondence of the citizen. Since then, more enforceable European tools surpassed its application in the field of data privacy.
The European Court of Human Rights' jurisprudence pointed out that private life concept, quoted by it, extends to aspects relating to personal identity and that therefore, the right to privacy established by this provision refers also to identity and personal development and interaction, as well as to the right to establish, maintain and develop relationships with other human beings. This case-law interpretation has to be taken into account in future project progress.

[2] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). It can be retrieved, for instance, at the following link: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:l24120&from=EN

[3] The European Convention on Human Rights, adopted in 1950 and entered into force in 1953. The Convention and its Protocols can be retrieved at the following link:https://www.coe.int/en/web/conventions/search-on-treaties/-/conventions/treaty/results/subject/3

[4] Charter of Fundamental Rights of the European Union, 2016/C 202/02. It can be retrieved at the following link: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12016P/TXT&from=EN

Article 8.2 states the lawfulness criterion, in the meaning of rule of law.
As for the Chapter of Fundamental Rights of the European Union, legally binding in the EU Member States since the adoption of the Treaty of Lisbon on 1 December 2009, it refers to both the right to privacy and the right to data protection, setting forth an explicit right to respect for privacy (Article 7) and an explicit right to protection in case of personal data processing (Article 8). Also, in this case, the European Court of Human Rights case law is an essential factor supporting the application of these articles in DataVaults and the elicitation of legal and ethical requirements.
The fundamental rights impact assessment is a useful mechanism that can be used to make sure that a system does not hamper EU fundamental rights, allowing for assessment and feedback on any potential risks or infringement of such rights. This impact assessment is additional to the Data Protection Impact Assessment (DPIA) regulated by the GDPR, and the EC prepared a set of indications adopted on it by the EC.

2.3 ETHICS & SOFT LAW

The composite regulatory system applicable to DataVaults is completed by the soft law (quasi-legal instruments), which may not have any legally binding force, such as European Courts' case law. This sort of instruments is helpful in so far they serve to fill in gaps, identify safeguards, boundaries and obligations to ensure the legitimacy and fairness of technologies like DataVaults, and, at the same time, contributing to find out, on a case-by-case basis, a balance between competing interests.
Soft law has an array of possible benefits and usually runs within the boundaries set by its interplay with the traditional legal instruments, in a landscape of increasingly dynamic cross-fertilization of regulations and technology.
It should receive the appropriate consideration when determining DataVaults technology design and deployment, especially due to the rapidly developing field of data sharing ecosystems: thanks to its flexible nature, that let it be quickly adapted to future technological progress, soft law could provide useful insights, recommendations and indications and support in identify the adequate safeguards and mechanisms in relation to transparency and accountability.
Among the other sources, in DataVaults the following have been considered for the first elaboration of the ethics and legal requirements:

- EC's Communications "AI for Europe" (25 April 2018) and "Building Trust in Human-Centric AI" (8 April 2019);
- "Data Protection in the era of Artificial Intelligence. Trends, existing solutions and recommendations for privacy-preserving technologies" (BDVA[5], October 2019);
- "Meeting the challenge of Big Data. A call for transparency, user control, data protection by design and accountability" (Opinion 7/2015, European Data Protection Supervisor, 2015)

2.4 REGULATORY FRAMEWORK IN THE SELECTED JURISDICTIONS

All of the pilots undertaken in the project will also be subject to the national regulatory landscape, such as national privacy and data protection legislation relevant to each of the

[5] Big Data Value Association

piloting operations and use cases. It is important to identify the national data protection authority and check if any notification/authorization is necessary with respect to the planned activities, both from the perspective of research with humans and data protection. The national privacy and data protection legislations mainly include:

- Demonstrator #1 – Sports and Activity Personal Data and Demonstrator #2 – Strengthening Entrepreneurship and Mobility: Greek Law 4624/2019 on the protection of natural persons with regard to the processing of personal data
- Demonstrator #3 – Healthcare Data Retention and Sharing: Belgian Law of 30 July 2018 on the protection of natural persons with regard to the processing of personal data ("Law of 30 July 2018") entered into force on 5 September 2018.
- Demonstrator #4 – Smarthome Personal Energy Data: Organic Law 3/2018 of 5 December on the Protection of Personal Data and the Guarantee of Digital Rights.
- Demonstrator #5 – Personal Data for Municipal Services and the Tourism Industry: Legislative Decree n° 101 of 10th August 2018 adjusting the Italian personal data protection code (Legislative Decree no. 196 of 30th June 2003) to the provisions of (EU) Regulation 2016/679.

More details on such sources, as well as the identification of the national data protection authority, can be retrieved in Annex 1 of this document. These will be further deepened in the next months, once the planned activities to be implemented in each local context will be better shaped, also in view of checking if a notification/authorization is necessary, both for the research with humans and for data protection issues. In case of need, the pilot partner concerned will apply accordingly.

3 FACTUAL BASIS FOR THE LEGAL AND ETHICAL ANALYSIS AND FOR THE REQUIREMENTS ELICITATION

3.1 DATAVAULTS DATA MANAGEMENT AND ANALYTICS CLOUD BASED PLATFORM AS A SERVICE & PERSONAL DATA APP

This section provides the description of the facts and aspects of the project relevant in order 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. This factual description is mainly based on the DoA and, if opportune, will be updated in D2.2, according to project's progress.

3.1.1 Overall reference architecture, services and components

The overall conceptional architecture of DataVaults has been devised with the security of data in mind, in order to increase the feeling of trust of the platform's users towards achieving the development of a personal data platform that can be trusted by its users for handling personal and confidential data based on the users' needs and preferences.

The security by design architectural blueprint of the overall DataVaults platform is presented in the following figure, where it is evident that various components are placed at strategic points that facilitates encryption of data as soon as it enters the overall platform, while users are offered with various options on how their data is used over the platform, always adhering to their own commands and preferences.

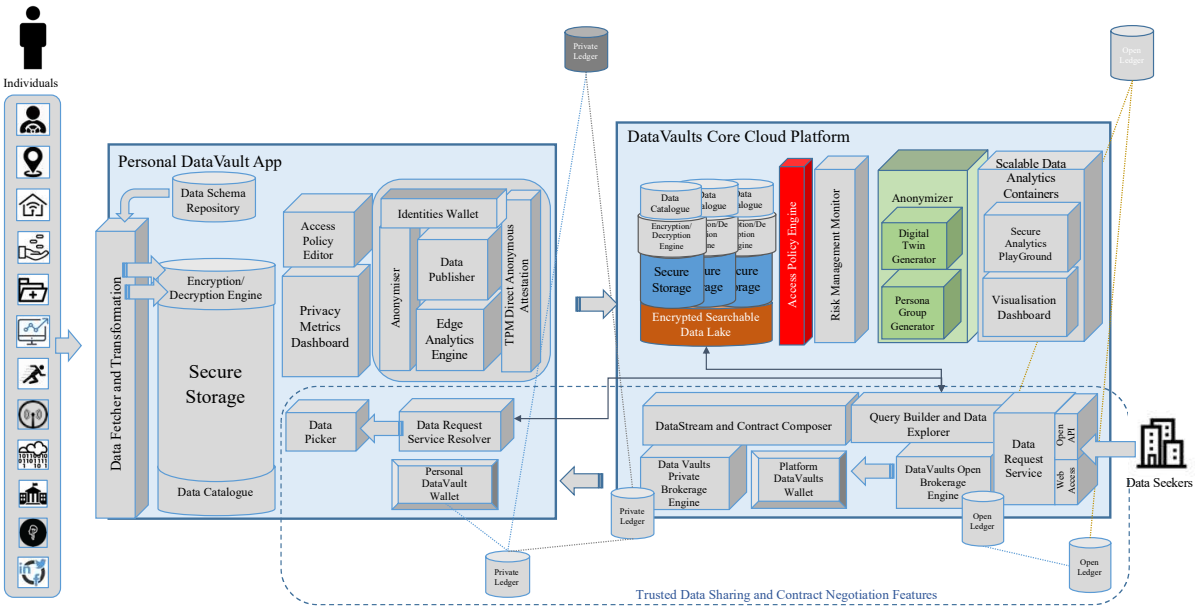

Figure 1 - Security by design architectural blueprint of the overall DataVaults platform.

As shown in the annotated version of the preliminary architectural figure above, the components that are relevant to the handling of personal data at the Core DataVaults

Platform (e.g. excluding the Personal Data App – see next section), and the services to be offered are the following:

- **The Secure Storage Containers**. These are secure storage facilities, where each one corresponds to a single individual user. Within this container, data can be encrypted or not, however access to these is only allowed to the platform, by using the Access Policy Engine (see below)
- **The Encrypted Searchable Data Lake**. This is a data infrastructure where pointers (keywords) to specific contents of each secure storage container are stored and that allow to search over the encrypted data (based on the keywords selected for each container
- **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.
- **The Anonymizer Engine**. A component that is used for manipulating data that should not be shared in their original format, in order to guarantee the privacy of the data owners and non-traceability, by actions on the stored data to anonymise/pseudonymise them, or to merge together data from similar data owners, to generate personas.

It is noted that the above-mentioned component and services are derived out of the original architecture diagrams of the project's proposal stage, which might be amended based on the work that is currently performed in the project. The first version of the DataVaults architecture is expected to surface at M13 of the project (Deliverable 5.2), nevertheless the main concepts and services as envisaged in the proposal's architecture will likely remain similar.

3.1.2 Personal Data App, services and components

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. In this context, the Personal Data App can be seen both as the personal data harvester component of a DataVaults users, as well as the control interface which dictates how the data to be collected is shared and used over the core DataVaults platform.

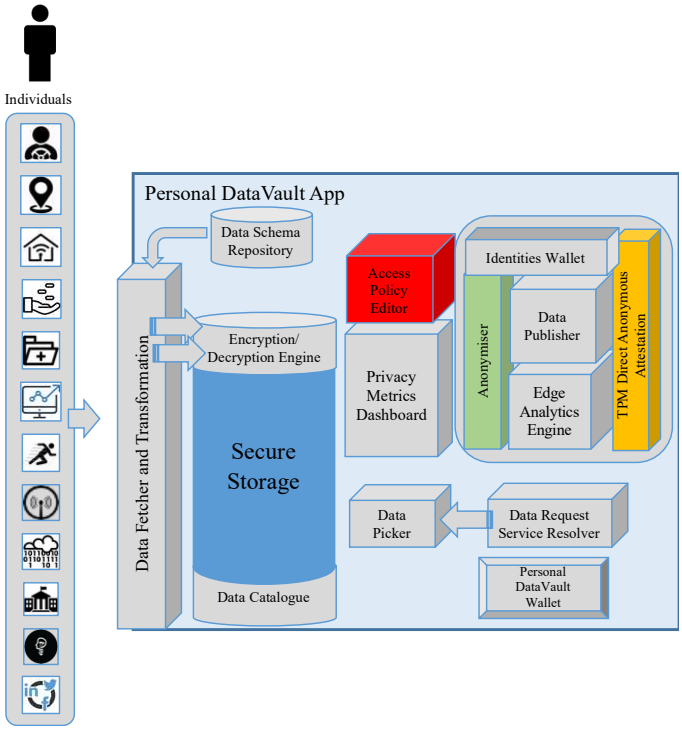

Figure 2 - Personal DataVaults App.

The Personal Data App, at this early stage of design, includes services and components that are similar to the Core DataVaults platform, though most of them are in smaller scale in order to meet the computational resource constrains of personal devices which will be considered as host environments for this App. These include:

- **The Secure Storage facility**. This is the local storage container for the Personal DataVault App where data resides, in an encrypted or not state. Access to this data, for further pushing it to the Core DataVaults platform is only provided using the Access Policy Engine (see below)
- **The Access Policy Engine**. This 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, or on preferences defined by data owners on how they wish their data to be uploaded to the Core DataVaults platform
- **The Anonymizer Engine**. A component used for manipulating data at the user's side, for uploading anonymous data to the core data platform
- **The TPM DAA module.** An infrastructure, based on TPM technology that allows the Personal DataVault App to be attested and trusted by the Core DataVaults platform, allowing the uploading of data and keeping the identity of the data owner hidden in case the latter chooses to be in incognito mode (especially needed when uploading anonymous data).

Similar to the Core DataVaults Platform, the above described services and components, coming out of the Personal Data App preliminary architecture might be amended based on the work that is currently performed in the project.

3.1.3 High-Level Data in Data vaults

3.1.3.1 Data sources

We can consider that a data source is any place from which we obtain data, be it data that is being generated at the moment or stored data. This would already allow a first division to be made about the nature of the sources.

In a first group would be any device capable of producing new data and offering it on demand: sensors, sensor networks, smart phones, wearable devices, etc. and that can also be divided into three subgroups:

- sources that do not actively provide the data and should be queried.
- sources that actively and periodically communicate the data.
- sources that allow an exchange of information and adaptation to the needs of the requester.

The second group includes all the data repositories that temporarily store data, such as databases, document repositories, etc. This group is fed with data from sources of the first group, with data provided by users through forms and also with the result of analysing other data sets.

3.1.3.2 Types of personal data

Another way to characterize data is by referring to its nature in terms of identifying people. In this regard, the GDPR establishes that personal data is:

> *"Any information relating to an identified or identifiable natural person (' data subject '); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. "*

This description is extensive and includes not only the data that directly identifies a person or is associated with an identified person, but also includes those datasets, which, through their analysis, allow identifying a person.

In this way, we can establish that regarding personal data we can have at least four types of datasets:

- **Raw data**: this is data collected from sources. It can be directly associated to a person through an identity or not, and in the case of being associated, this identification can be removed or anonymized.

- **Processed data**: is the result of analysing raw data to look for valuable information. In this case the analysis can be performed over one or several sets of raw data of a single person or several people and can derive on identifying a person.

- **Identification data**: this is a sensitive type of data set aimed at identifying a person with a clear objective. it can contain name, legal information like passport number, fiscal information, banking data, etc.

- **Contracts**: this is a special case which is aimed at regulating the relations in between the three former types, and between their owners and their users. They may contain several aspects such as:
    - who is the owner of a dataset
    - who can analyse a dataset
    - who is the owner of the result of the analysis
    - the type of analysis it can be performed
    - what kind of information can be included in the result of the analysis
    - who can use the result of data analysis
    - how data owners can execute their rights regarding GDPR
    - compensations for data owners who let third parties use their data.
    - etc.

  These datasets will include identification of parts signing the contract.

3.1.3.3 Format

Another aspect to consider is the format of the data, since it affects the space required for communicating and storage needs. Three major groups of formats can be distinguished:

Unstructured data: is any dataset without a reliable structure from which we can extract other data of our interest. Images, texts, etc. belongs to this group. They usually take the format of documents and are stored in document repositories.

- **Structured data**: is the data that has fixed and well-known format and organization that allows relationships to be established. This is the case of relational databases and spread sheets.
- **Semi-structured data**: is data that has a fixed format but with a non-strict organization, this is the case of mark-up languages such as Extensible Markup Language (XML)6 and JavaScript Object Notation (JSON)7. These formats arose from the need to send data between systems in a versatile way that would serve in all contexts.

We can find various standards that use these data formats and that are dedicated to specific fields, for example in health. HL7[8] is a set of standards dedicated to the exchange of clinical and administrative data between different health service providers.

[6] https://www.w3.org/XML/
[7] https://www.json.org/json-en.html
[8] https://www.hl7.org/implement/standards/

In other fields such as energy, we find initiatives such as the CIM[9] standard for the exchange of information in electrical networks, or the Energy@home data model[10] which aims to create a standard to connects smart energy devices in home to the Smart Grid.

3.1.4 Data Subjects and other actors

The General Data Protection Regulation (GDPR) is the toughest privacy and security law in the world. Though it was drafted and passed by the European Union (EU), it imposes obligations onto organizations anywhere, so long as they target or collect data related to people in the EU.

Since the project DataVaults aspires to become "one of the flagship personal data platforms which are fully compatible with GDPR" [1] it is mandatory that the definitions and the actors involved in the project should be designed according its specifications.

The GDPR in its 4th article [2] lists a group of "Definitions" among which we can find the most significant actors related to the data protection and regulation:

- **Data subject:** the identified or identifiable natural person whose data are processed. The identification can be directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.
  The processing of the data must be lawful, fair, and transparent to the data subject and according to the legitimate purposes specified explicitly to the data subject when collected.
- **Data controller:** The natural or legal person, public authority, agency or other body which, alone or jointly with others who determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law.
  The data controller establishes the purposes for which personal data is used and what privacy protection should be implemented. Each controller shall maintain a record of processing activities under its responsibility, that shall contain information such as the name and contact details of the controller, the purposes of the processing, a description of the categories of data subjects and of the categories of personal data, the categories of recipients to whom the personal data have been or will be disclosed, and more information detailed in the 30th article of the GDPR.
- **Data processor:** A natural or legal person, public authority, agency or other body party that processes personal data on behalf of a data controller.
  That processing is described in the 28th article of the GDPR and shall be governed by a contract or other legal act under Union or Member State law, that is binding on the processor with regard to the controller and that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of

[9] https://www.dmtf.org/standards/cim
[10] http://www.energy-home.it/Documents/Technical%20Specifications/E@h_data_model_v2.1.pdf

personal data and categories of data subjects and the obligations and rights of the controller.

- **Data recipient:** A natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not. Public authorities can be an exception when receiving personal data in some specific cases but in any case, the processing of the data should follow the GDPR.
  The information about recipients or categories of recipients of the personal data shall be provided to the data subject by the controller, if personal data have not been obtained from the data subject.
- **Third party:** A natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorized to process personal data. Sometimes third parties can act as processors, but usually are vendors and other outside stakeholders which if performing any processing of personal data, it shall be governed by a binding contract.
- **Representative:** A natural or legal person established in the Union who, designated by the controller or processor, represents the controller or processor regarding to their respective obligations under the GDPR. More details in Article 27.

The DataVaults project, and each individual demonstrator, should determine who is who in each scenario. The main entities involved in the solutions will take a role according to those actors described above, in order to be complying with the GDPR. In a high-level vision and being aware of the necessity of a deep analysis of the scenarios, not performed due to the current early stage of the project, a first approach of the main roles in the different demonstrator can be as indicated in Table 1.

Demonstrator	Data subjects	Controllers	Processors	Recipients
Sports and activity personal data	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
Strengthening entrepreneurship and mobility	Citizens, visitors.	Local 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
Healthcare data retention and sharing	patients.	Andaman7	doctors, hospital	third parties in the health sector (e.g.: clinical trial, research)

Demonstrator	Data subjects	Controllers	Processors	Recipients
Smarthome personal energy data	users/customers	MIWENERGIA	other companies	other companies to offer services
Personal data for municipal services and the tourism industry	users/customers, citizens of Prato	The Municipality, the registry office	The event organization of Cultural and tourist institutions in the city	Cultural and tourist operators

Table 1. First approach in Demonstrators and actors.

From the point of view of the high-level architecture that will be followed in Data Vaults, these actors can be identified, at a general level:

- Individuals as data subjects. Identifiable natural persons who will take ownership and control of their data and decide to share them with other entities.
- The "Personal DataVaults" component as part of data subjects. It can be considered as a tool that they use to manage their data, edit them, define the permissions, define the policies that indicate how the data can be used.
- The "DataVaults Cloud Platform" component that acts as a broker between individuals or data subjects and organizations, can be considered as a controller, taking into account that this platform is going to be used by an entity (company, club or administration), responsible of the management of the data. If the entity using this platform hires the services of another company for managing some data or any process, then this entity and subsequently this platform would act as a processor.
- Data seekers 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.

The demonstrators will determine the specific roles that each of the components of the architecture will take, depending on the scenarios that are going to be developed

3.1.5 Data Life Cycle: collection, processing, storage, sharing personal data and derivatives

The data lifecycle of DataVaults starts with an individual that decides to collect its personal data and may go until the point where this data is shared and reused by other parties. However, the data lifecycle may end at any point of this process, and this is to be decided by the data owner, at any given time, respecting in any case the data contracts that may have been signed between a data owner and a data consumer (e.g. in case access to and usage of past data has provided unconditionally, this cannot be revoked by the user, but access to future data can be prohibited).

The following workflow, taken out of the DoA provides a high-level overview of the data life cycle within DataVaults, which will be further detailed in the next stages of the project and

will be used to drive the definition of the activities of the MVP that refer to the sharing of the data.

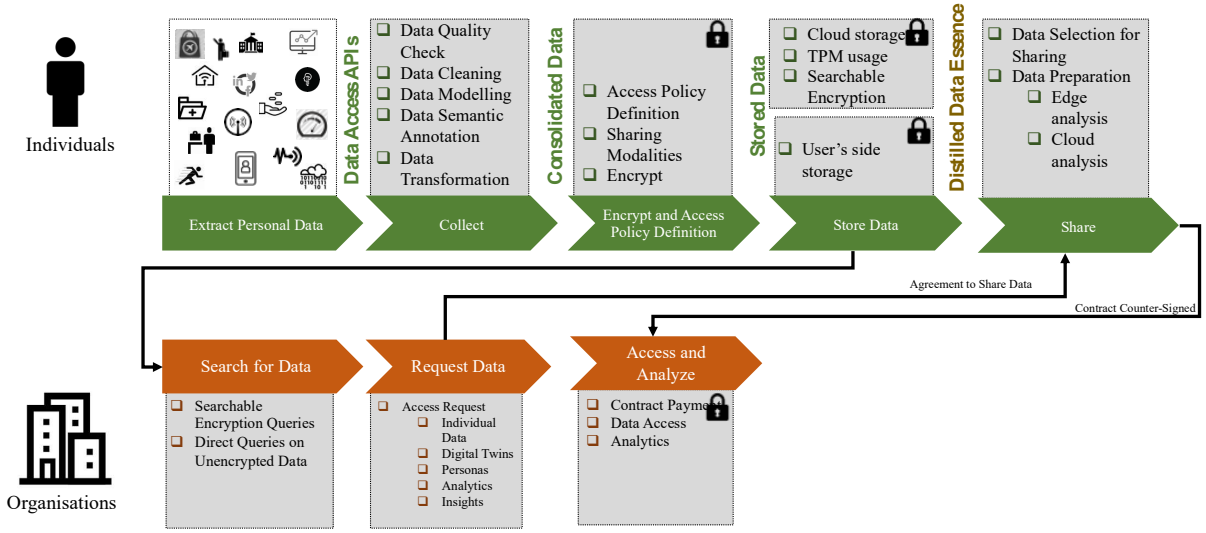

Figure 3 - High-level overview of the data life cycle within DataVaults

As shown in the figure above, the lifecycle of data sharing starts with data that are generated at the user side from various sensors or APIs and these are then "**Collected**" by the user. At that stage data preparation activities are performed, that have to do with data quality checking, data cleaning, etc, in order to transform the data to the common schema of DataVaults (to be defined in WP1). Following this, the "**Encrypt and Access Policy Definition**" step is performed, where the data owner chooses if/how to encrypt the data and decides the access policies that are to be applied on the selected data. Following this, the owner is able to "**Store** the **Data**", which can be done on the user's side, or on the Core DataVaults Platform, or even in both locations, depending on the user's preference. At that point, data is securely stored in the repositories and resides there until a sharing request emerges and is of course accepted. To arrive at such a situation, an external Organisation performs a "**Search for Data**" step, which allows it to query the data stored on the Core platform and identify if he would like to proceed to request it. In case that he proceeds, the "**Request Data**" step is triggered, where the data seekers define the type of data (actual data/analytics/insights) to retrieve and the nature of it (original data/digital twin data/personas data) and at that point DataVaults performs internal operations to identify, access, gather and define the value of the data necessary towards constructing a smart contract that has to be signed by the data-seeker. Upon acceptance of the contract by other parties (e.g. the data owner and the data seeker), control is moved to the platform which executes the "**Share**" stage, where the data is bundled together and released to data seeker. At this stage, also depending on the type of data requested, analytics may run either on the user's side or on the cloud platform, to provide to the data seeker the relevant information. Finally, the data is provided to the data seekers where during the "**Access and Analyze**" step it is able to perform operations on the retrieved data and its derivatives. It needs to be noted that the retrieved data and its derivatives may be used by a data seeker for further operations with other collaborators, however these steps do not concern DataVaults as a platform and cannot be controlled/imposed by the platform as a technical infrastructure.

Nevertheless, it is envisaged that legal clauses that will be part of the smart contracts in order to control (offline) the way data can be re-used by any engaged party.

3.2 DEMONSTRATORS AND USE CASES: FACTUAL BASIS FOR INITIAL ETHICS AND DATA PROTECTION INSIGHTS

The scenarios description and the data sources involved in each of them is part of WP1 work and their initial description can be found in D1.1 "DataVaults Data Value Chain Definition". Here we extract a summary for each of the scenarios to provide some context before going into detail to provide initial insights for legal analysis, in particular regarding data protection and ethics.

3.2.1 Demonstrator #1 – Sports and Activity Personal Data

The planned use cases described in D1.1, are the following ones:

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.

3.2.2 Demonstrator #2 – Strengthening Entrepreneurship and Mobility

The planned use cases described in D1.1, are the following ones:

Scenario A. Smart Mobility Services for Individuals. This scenario will engage both OLYMPIACOS and PIRAEUS and will use the data shared by the interested 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 specific area of interest during the course of the project 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 Personalised cultural and touristic experiences. This scenario will build on data analysed form the profiles and preferences of the DataVaults app users, in order to create services that target tourists and citizens visiting the city of Piraeus. During this scenario, the data to be analysed 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 organisations, etc.) data sources.

As for privacy and data protection initial insights, the participation of the Municipality of Piraeus goes beyond its role as a public service authority and therefore for all scenarios and for all obtained personal information, the participants must sign a consent form assuring they acknowledge how the data are going to be processed and their corresponding rights. The Greek Law 4624/2019 must be respected for all Greek users. Regarding privacy and data protection the users should always maintain the ownership of the data and the right to decide about sharing them. They should be able to perform all their rights as derived from GDPR, more specifically:

- The right to information
- The right to access
- The right to rectification
- The right to erasure ('right to be forgotten')
- The right to restriction of processing
- The right to data portability
- The right to object

Third parties accessing personal information must respect the data protection law as well as the user's rights.

3.2.3 Demonstrator #3 – Healthcare Data Retention and Sharing

The planned use cases described in D1.1., are the following ones:

Scenario A. Current users of Andaman7 will be able to connect to DataVaults to store all or part of their health data (after explicit consent). This storage can be used as a backup to retrieve data when lost. This can also be used by third parties in the health sector (e.g.: clinical trial, research). Users will be able select what services they want to subscribe to and what kind of data will be shared.

Scenario B. Current users of Andaman7 will be able to connect to DataVaults to collect their data (coming from various sources) and store them in Andaman7 on their smartphone. This will make the data available to patients for reviewing, learning, using in other set ups (e.g. share additional data with their doctors, hospital, etc.). Data will mostly be raw personal data but also aggregated data (e.g. result of a clinical trial, comparison to a specific group, …)

Concerning privacy and data protection, it has to be underlined that health data is a special category of data, also referred to as sensitive data. From a technical perspective, we need robust data protection safeguards in order to maintain the trust and confidence of individuals in the rules designed to protect their data. In addition, explicit consent of the source should be asked and stored to exchange, store or process such data.

According to the Belgian Law of 30 July 2018, 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
- make sure that the designated persons are subject to a legal, statutory or equal contractual obligation to ensure the confidential character of the personal data.

3.2.4 Demonstrator #4 – Smarthome Personal Energy Data

MIWenergia is an electricity retailer with a database of around 3.000 customers. The company collects energy consumption data from clients, and also gather personal data when they sign the contract for billing purposes (name, address, bank account, contact details, etc.).

The planned use cases described in D1.1., are the following ones:

- PV installation design for self-consumption: using energy data consumption with extra information provided by the user about their building such as available space, kind of roof (slope or flat), location, our company could design the installation of a self-consumption PV plant.
- Improve profiling of clients to enhance energy efficiency: using energy data along with additional information such as area, number of people living in the dwelling, our company can profile the user for their efficiency and offer them services related to energy savings.
- Energy consumption patterns with personal preferences: getting personal data regarding likes and dislikes, hobbies and typical schedule combined with their electricity consumption, patterns of the users can be created and used by third companies to offer different services.

Energy consumption data would be provided by MIWenergia but always with the user acceptance. The other data needed in each case regarding to the building information or personal data would be provided by the users through DataVaults platform. Data providers would need to give specific permissions for each kind of information and use of data. In all the scenarios described in D1.1 related with smarthome personal energy data, the approval and consent of DataVaults user´s is required.

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. The Spanish Organic Law 3/2018 of 5 December must be respect for the Spanish users. If third parties are involved, users must be informed. Third parties must respect the data protection law as well as the user's rights.

3.2.5 Demonstrator #5 – Personal Data for Municipal Services and the Tourism Industry

Case 1: Customer satisfaction analysis for the administration services

The user installs the Prato app on his/her smartphone and accesses own data registered in the municipal registry service (in the case of a resident citizen), in order to check and update them if necessary. In this case, the updated data on DataVaults are automatically used also for the registry services, which interface with DataVaults. In case the user is not a resident in Prato (and therefore not present in the registry), through the app he/she can still enter his contact details.

Alternatively, the citizen residing in Prato goes to a registry office where he/she updates his/her data: at the counter he/she is suggested to download the Prato app to help the municipality in keeping data updated. The updating of the data made at the counter automatically feeds the DataVaults database, as the registry database interfaces with the platform.

By using the app, the user helps to provide data continuously in his/her DataVault repository at least in the following ways:

1. form for voluntary updating of personal data (name, address, telephone number, email, interests, etc.),
2. location information at intervals through the smartphone,
3. detection of the email accounts configured on the smartphone.

The service operator who must carry out a customer satisfaction survey uses the web interface of his service to extract a large list of potential interviewees, submits the list to the DataVaults platform through the appropriate web interface, also composing the form with the questions to be administered. The users of the list will receive an invitation to participate in the survey and, in case of consent, will fill in the relative form.

The user included in the sample list receives on the app a notification of the proposal of a contract that regulates the use of his/her data by the Municipality, through which he/she will obtain a fair remuneration. Through the app, the user accepts the proposal and receives a survey questionnaire for customer satisfaction from the Municipality. Once the questionnaire is completed, the user receives the equivalent agreed in the contract in his/her digital wallet on DataVaults.

Case 2: approval and use of cultural and tourist services in the city.

Cultural and tourist operators access the web interface of the DataVaults platform to set up news relating to the various cultural events they organize, scheduled in the city.

The user installs the DataVaults app on his/her smartphone and with the daily use of the app he/she voluntarily provides data in at least the following ways:

1. form for voluntary updating of personal data
2. location information at intervals
3. detection of the addresses corresponding to the mail accounts configured on the phone.

Cultural and tourist operators can use the platform in the following ways:

- targeted information on scheduled events: the operator sends the users who meet certain selection criteria based on the data present on the DataVaults platform (e.g. position, personal data, movements in the city, etc.) information banners relating to the scheduled events;
- request for feedback on attended events: always on the basis of the above selection criteria, the operator can request the expression of an opinion/approval for a given cultural event attended by the user;
- data analysis: the operator can extract from the DataVaults platform analytical information relating to, for example, typical itineraries of tourists in the city, statistics on the attenders of an event, correlations in the attendance of events, satisfaction, etc.
- furthermore, through the interaction between the DataVaults app and the other apps on the user's smartphone (e.g. social networks), the cultural/tourist operator can invite the user to post comments with suggested content on his/her own social networks.

Concerning privacy and data protection, we plan to follow well-defined consent procedures: before gathering information, the volunteers must be duly informed and sign a consent form. All the rights of the users as described by GDPR will be respected.

4 LEGAL, ETHICAL, SECURITY, PRIVACY AND TRUST REQUIREMENTS

4.1 LEGAL AND ETHICAL REQUIREMENTS

4.1.1 Requirements list

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 an initial insight, taking so far input mainly from the DOA and the literature. Therefore, the list has been drawn at a higher level of abstraction, to cover various possible future technological choices. It may be updated according to project's progress, once its services, solutions and demonstrators are better shaped, till their final fashion. In a later stage of the project, and in particular in the future deliverable of this WP, if the case we will refine or revise it.

The requirements have been elicited adopting a systematic and holistic approach, driven by Fairness & Privacy-by-Design-and-by-Default enriched with the Protection Goals method[11]. Such elicitation relies on the analysis of the regulatory landscape and the factual analysis of the privacy-relevant properties and personal data collection, processing and sharing in each service and tool, including details on the data categories, data sources and purposes of processing.

These 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.

[11] 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
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.	The whole system and app	All	PDPL, HRs, ESL
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 Managemenr 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	Core DataVaults platform: Secure Storage Containers Personal Data App: Secure	All	PDPL, ESL, ITSL

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		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.	Storage facility, Data Feeder and Transformation, TPM DAA module		
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 the data subject is fairly considered as one of the fundamental rules of a lawful personal data processing	Core DataVaults platform: Access Policy Engine. Personal Data App: Access Policy Engine, TPM DAA module	All	PDPL, ESL, RFSJ

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		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).			
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	The whole system	All	PDPL, HRs, ESL

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		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).			
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 individual's over their personal data, taking into account 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.	The Personal data app	All	PDPL, ESL, HRs
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)	The Personal data app	D	PDPL, HRs, ESL, RFSJ

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		volunteers' dignity has to be safeguard and direct/indirect incentives for participation must not affect it.			
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: - Transparent communication (Art. 12 GDPR); - 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); - Right of access (Art. 15 GDPR); - Right to rectification of inaccurate data (Art. 16 GDPR); - Right to erasure, 'right to be forgotten' (Art. 17 GDPR); - Right to restriction of processing (Art. 18 GDPR); - Right to receive a notification from the controller regarding rectification or erasure of personal data or restriction of processing (Art. 19 GDPR); - Right to data portability (Art. 20 GDPR);	Personal data app: Privacy Metrics Dashboard, Access Policy Editor, Identities Wallet, Data Request Resolver, Data Picker Core DataVaults platform: Access Policy Engine, DataVaults Private Brokerage Engine	All	PDPL, ESL

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		- Right to object (Art. 21 GDPR); - Protection against automated decision-making, including profiling (Art. 22 GDPR).			
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 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.).	Core DataVaults platform: Open Ledge, DataVaults Open Brokerage Engine, Contract Composer, Private Ledger, DataVaults Private Brokerage Engine, Access Policy Engine	All	PDPL, ITSL, RFSJ
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.	The whole system	All	ESL

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		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.			
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). 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.	The Personal data app	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	The Personal data app	All	PDPL, ESL

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		authorities (with undue delay, and, if feasible, within 72 hours). As for the data owners, 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. 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**	In case it is likely that the data processing in DataVaults	Core DataVaults platform and	D (and	PDPL, ESL,

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
	Impact Assessment	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. 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. DataVaults is strongly committed to operationalize the risk-based approach encouraged by the GDPR and specific tools and services will be devoted to this.	personal data app: Risk Management Service and Risk Exposure Dashboard	potentially R)	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	N/A	D	PDPL, RFSJ

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		demonstrators have to take this clarification into account and consult their 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. 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). 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	The whole system and in particular: Secure Storage Containers,	All	PDPL, ITSL, RFSJ

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		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.	Encryption/Decryption Engine, Access Policy Engine		
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
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	Personal Data App: Privacy Metrics Dashboard, Anonymiser, Identities Wallet	All	PDPL, ESL, RFSJ

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		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".	Core DataVaults platform: Anonymiser		
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	N/A	All, but especially D	PDPL

N.	Short name	Description	Supporting DataVaults Tool	Phase	Nature
		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.			
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

Table 2. Legal and Ethical Requirements.

4.1.2 Additional notes, recommendations and guidelines for the requirement operationalization

The following remarks are relevant for the operationalization of the requirements, listed in the previous paragraph, in DataVaults research.

R1. Fairness and Lawfulness

Though the usual legal basis is the data subject's consent, when consent is not applicable or may be disproportionate for the individuals, the data processing can rely on further different legal bases (Art. 6 GDPR), namely:

- the necessity to perform a contract with the data subject or take steps prior to entering into such contract;
- the necessity to comply with a legal obligation;
- the necessity to protect a vital interest of the data subject or another individual;
- the necessity to perform a task in the public interest;
- the necessity for the purposes of a legitimate interest pursued by the controller.

When the pursuit of legitimate interests is invoked as legal basis, the necessity of processing for each data category has to be assessed, as well as the documentation of purposes and underlying interests of data processing, considering the concrete situation of data processing.

The collection and processing of special categories of data, such as health data and biometric data, require for a separate legal basis.

The correct choice of a legal bases for the purposes of the personal data processing activities involved is a key element for compliance in DataVaults. A straightforward requirement is that the controller must not only carefully choose the most appropriate legal basis, but event justify this choice in the information notices and the records of processing activities created.

This is a central requirement, closely interrelated to the purpose limitation principle (Art. 5 GDPR)

It demands that the controller must define the purposes before any personal data collection or processing, as well as identify the appropriate legal basis for each of the purposes.

The description of the legal basis must be in a plain and understandable form and, at least for more complex legal grounds, rather than just citing the legal norm, the controller has to substantiate how the norm covers the envisaged processing procedures. This might be challenging in DataVaults uptake, in case the purposes were not previously defined, and it is recommended that adequate specifications for communication by automated means between controllers and data subjects provide a viable solution for data controllers.

R2. Purpose limitation and legitimate aim

This principle is relevant for determining the different data processing activities that are performed by a controller. The processing purpose: i) determines the number and kind of processing activities to be carried out, in the sense that they depend on the reasons for which personal data are processed ii) has an impact on the specific law provisions to be

complied with iii) is strictly interrelated with other basic principles, namely the transparency principle, the data minimization and the data retention principle.

This principle is strictly interrelated with: i) the transparency principle, in the sense that the processing purposes should always comply with what specified by the controller in the privacy notice ii) the data minimization principle, since personal data must be processed to the extent strictly necessary and proportionate for the purposes established. The personal data which, when assessed towards the processing purpose, is deemed redundant or unnecessary, can't be collected or used iii) the data retention principle, in the case where data is no longer necessary for achieving a specific processing purpose, for instance because the said purpose has been achieved, then these data, as soon as they become unnecessary, should be promptly either deleted or anonymised.

As for the DataVaults demonstrator, a key dimension of the purpose of the processing is the research itself.

Recital 159 of the GDPR lingers over the processing for scientific research purposes, specifying that "where personal data are processed for scientific research purposes, this Regulation should also apply to that processing. For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research".

Art. 89 GDPR requires that, in the case of processing for scientific research purposes (and others, not relevant in DataVaults context), appropriate safeguards for the rights and freedoms of the data subject pursuant to GDPR have to be taken, ensuring that technical and organisational measures are in place in order to ensure, in particular, the principle of data minimization. This provision is specified by Recital 156.

R3. Data minimisation

The interpretation of the minimisation principle needs to be adapted to the context of DataVaults, which has personal data, coming from diverse sources, in its centre. DataVaults project is precisely aimed at developing a personal data platform to set, sustain and mobilize an ever-growing ecosystem for personal data and insights sharing. The data collection and processing – also in terms of data minimization - need to adhere to the users' needs, commands and preferences, in order to let this personal data platform be trusted by its users for handling personal and confidential data.

Appropriate anonymisation techniques should not consist just in stripping a dataset of some directly identifying attributes, especially in case of bigger and comprehensive personal data collection (like the one foreseen by DataVaults), where there is the possibility to identify the individuals whom the data relates to, especially when data is retained for longer periods of time and/or shared, could potentially be higher. The use of such techniques should be careful and combined with other effective safeguards, such as access control.

For this purpose, privacy and anonymisation service in DataVaults is expected to capitalize on advanced anonymisation tools, such as the Privacy Preservation Anonymizer developed in the AEGIS project, as well as to make possible a number of combinations of data sharing

modalities also as regards anonymisation, ranging from sharing individuals' actual data of individuals, to the creation of anonymised "digital twins" of individuals (using anonymisation and data obfuscation mechanisms), or sharing anonymous data grouped under "personas" containing data from similar individuals.

In DataVaults, this principle has to be read in conjunction with the storage limitation principle and purpose limitation principle, as well as with the principle of data quality. It is essential to determine to what extent, considering the processing purpose, keeping personal data instead of anonymised data is preferable (with the consent of the data subject).

R5. Integrity and Confidentiality

The GDPR provides specific rules for data security (Art. 32 GDPR), relevant to DataVaults. As for the demonstrators, the researchers in charge of collecting, using or accessing personal data in each of them must be subject to an enforceable duty to keep them confidential and secure. Therefore, a confidentiality clause or agreement should be concluded by all research staff that will be having access to personal data in DataVaults. It is also recommended to establish a closed user group, composed of only authorised persons (contractually obliged to keep confidentiality and meet data security rules), as well as to make use of an authentication and authorisation mechanism in each of them.

Besides the technical and organisational measures to be taken in view of ensuring confidentiality, also publication of project result will not reveal the data subjects

This principle has to be applied according to a risk-based approach, which is elaborated somewhat in Recitals 75 and 76 GDPR.

R7. Transparency

In DataVaults, transparency requires documentation of the processing activities (which will be necessary anyway under Art. 30 GDPR), including, especially with regard to the data subject, a description in easy language of how personal data are processed, of the main features and conditions of the processing and of what are the potential risks of such processing in respect to privacy. The transparency principle needs to ensure in DataVaults technology and its future operation that at any time it is possible to understand and reconstruct the data capturing, processing, and use, both actual and planned.

Information must be provided in form and extent adequate to its recipient: different ways of information (channels, granularity, language, etc.) could be advisable in relation to different user groups.

Where there are several exemptions to requirements, such as obtaining consent as legal basis to processing, the exemptions and limitations to information requirement, instead, are very circumscribed, both at a European and at national level.

In the DataVaults environment, both during project's implementation and in the exploitation stage after the end of the project, users may struggle to receive meaningful and complete information on the data processing activities. This might happen, for instance, in relation to IoT-related data capturing and processing, where there might not be a comprehensive information channel through which users can understand how data are collected and processed and by whom and, therefore, they might not have an effective user knowledge

and/or understanding of the data processing. Furthermore, the principle of transparency might be significantly challenged in case the IoT devices inadvertently collect also personal data of data subjects who did not provided consent to the data processing, such as visitors of smart homes/offices. These concerns have to be taken into account in the design and development of the system and app. A layered approach could be explored for this purpose: further indications on it will be provided under “Informed Consent” and “Privacy Notice” requirements.

R8. Privacy and Data Protection by Design and Privacy by Default

In a flavour consistent with the data minimisation requirement and the accountability principle, DataVaults needs to put the Privacy-by-Design-and-by Default approach at the core of its design and development efforts, both at the time of the determination of the means for processing and at the time of the processing itself (including in the post-project phase). In the project, Privacy and Data Protection by Design also means that new ways to be informed about what happens to one’s data, and to exercise control over one’s data must be offered to individuals through innovative, responsible and privacy-friendly engineering (besides privacy-friendly organisational arrangements and business practices), capable of facilitating, among others, the exercise of individuals' rights of access, objection, opt-out, correction and data portability.

The principles of “Data protection by design and by default” relies on the idea that there is the need of putting privacy principles into the design process of data processing systems since the very beginning and should encompass not only the strictly technological dimension of the system design, but covering the envisaged business processes. As for the data protection by default, is consistent with the opt-in approach to data-based services, strongly promoted in the GDPR.

For the operationalization of this principle, as for the accountability principle, it is necessary to adopt a risk-based approach.

In DataVaults design, the Privacy-by-Design-and-by-Default approach should be combined with the Security-by-Design-and-by-Default one, in order to minimise the risks of compromising privacy, as well as with the Privacy Protection Goals method, in order to use a systematic approach to determine technical and organisational measures. This approach relies on the protection goals as central element for deriving requirements to be complied with in system design, as well as for identifying risks, safeguards and countermeasures. Besides the well-known security protection goals named “Classic CIA Triad” (consisting of confidentiality, integrity, and availability), three further specific privacy protection goals are encompassed: unlinkability (enriched with data minimization), transparency and intervenability. Protection goals promote the balance of these privacy and security requirements against other protection goals and take the fundamental rights perspective more into account.

This model might serve in DataVaults as a tool to translate sometimes rather abstract legal and ethical requirements into concrete functional and organisational requirements, through the application of the protection goals to data, systems and processes, alongside with a determination of required level of protection for the personal information involved. A

relevant example of Privacy Protection Goals method, merged with the Privacy-by-Design-and-by-Default approach, is the Standard Data Protection Model elaborated by the National Data Supervisory Authorities in Germany and recommend for use in Germany: https://www.datenschutzzentrum.de/sdm/.

This requirement is imposed not only by the GDPR, but also by the upcoming ePrivacy Regulation. The European Data Protection Supervisor's Opinion 05/2018 on privacy by design needs to be followed as well.

In addition, it is important to refer to relevant standardization initiatives on this topic, such as the ISO 29100 - Privacy framework, ISO 27550 - Privacy engineering services, CEN-CENELEC/JWG 8 'Privacy management in products and the IEEE P7002™ addressing Data Privacy Processes and Methodologies. Examples for technologies supporting privacy by design and by default that may be considered in DataVaults are, besides anonymization and pseudonymization (expressly mentioned by the GDPR), the sticky policies, the automated procedures for obtaining informed consent in user-friendly manner and the provision of functionalities to manage own personal information.

R10. Informed Consent

Consent has to be given for the processing of personal data for one or more specific purposes. In case of new purposes, the controller needs to either get fresh consent specifically covering such new purpose or find a different legal basis for the new purpose.

As for DataVaults, it has to be remarked that, even when expressed through electronic means, the consent of data subject should be preventive and unambiguous.
It requires a statement or clear affirmative action of the data subject. For instance, these actions can consist of ticking a box in an online environment, the choice of technical settings for information society services, and any other statement or conduct clearly indicating the data subject's acceptance of the data processing activities.
DataVaults technology must also ensure that, where consent is obtained through use of a service-specific user interface (for example, within the DataVault personal data app or the interface of an IoT device), the data subject must be able to withdraw consent through the same electronic interface with undue effort and without detriment.
The EDPS Opinion 7/2015[12] outlines challenge relevant to DataVaults and that need to be addressed. It clarifies that in many Big Data environments "individuals cannot efficiently exercise control over their data and provide meaningful consent in cases where such consent is required. This is all the more so as the precise future purposes of any secondary use of the data may not be known when data is obtained: in this situation, controllers may be unable or reluctant to tell individuals what is likely to happen to their data and to obtain their consent when required".

The data collection and processing in DataVaults might be intended for multiple purposes (considering also the whole DataVaults ecosystem and value chain, as expected to evolve

[12] "Meeting the challenges of big data. A call for transparency, user control, data protection by design and accountability", 19 November 2015

after project's end) and it is necessary to ensure the consent for all of these purposes (Recital 32 GDPR).

Recital 43 GDPR casts doubt on an approach based on one single consent form, broadly formulated as pre-emptively covering different future business models of the data controller.

Globalized, generic consent for multiple vague purposes risk to be assumed as not freely given and the question that arises is whether separate consent and the need for several, broken down consent requests are appropriate. This needs to be explored in the context of DataVaults, but also reflecting on the need to avoid 'consent-fatigue' of a data subject.
As acknowledged by the Article 29 working Party[13], a layered approach could be a possible solution, still providing all necessary information step by step and providing balancing means of user control, whilst being substantially different by the mere use of pre-ticked boxes: it is not necessary that the first layer of information is completely in-depth about the details of the processing. It should be explored if, for most of the cases (though not applicable to the special categories of personal data of Art. 9 GDPR), an implicit consent (such as a shade going away after a few seconds and assumes "yes") could work, after the first general consent during the installation of DataVaults personal data app. It should be likewise investigating which information needs to be given to the data subject in which layer.

Useful indications for DataVaults can be retrieved in the following GDPR Recitals:

1) Recitals 32, which clarifies that it can be a written statement, including by electronic means, or an oral statement, if the data subject's behaviour clearly indicates his/her acceptance of the data processing. It is recommended that if the data subject's consent is to be given following a request by electronic means, such a request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided;

2) Recital 33, which states that, being often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection, data subjects should be allowed to give their consent to certain areas of scientific research (or parts of research projects) when in keeping with recognised ethical standards for scientific research. "Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose". This is especially relevant for DataVaults demonstrators and needs to be implemented, in particular in the recruitment and consent procedures of volunteers.

3) Recital 42, which states that "…For consent to be informed, the data subject should be aware at least of the identity of the controller and the purposes of the

[13] Article 29 Working Party "Guidelines on consent under Regulation 2016/679", adopted on 28 November 2017 and revised and adopted on 10 April 2018

processing for which the personal data are intended. Consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment".

DataVaults consent management policies need to ensure that the consent is:

1) "granular", capable of providing distinct consent options for distinct processing operations;
2) specific to "one or more specific" purposes, ensuring that the data subject has a choice in relation to each of them;
3) freely given, in the sense that the data subject should be able to exercise a real choice, without risk of deception, coercion, intimidation, or significant negative consequences if he/she does not consent;
4) informed, being the provision of information to data subjects prior to obtaining their consent necessary to enable them to understand what they are agreeing to, make informed decisions, and exercise control and, in general, their rights (including to withdraw their consent). As noted, a layered approach could help in this regard;
5) separate from other terms and conditions;
6) "explicit", in case of processing of special categories of data, profiling activities or cross-border data transfers. Though in many cases the term "explicit" could be interpreted as given in writing with a hand-written signature, in digital or online context like DataVaults, a data subject may be able to issue the required statement with other modalities (such as by filling in an electronic form, or by using an electronic signature).

In case of public authorities, there might be a clear imbalance of power in the relationship between the controller and the data subject and other lawful bases for the processing could be, in principle, more appropriate. This has to be taken into consideration, for instance, for the Demonstrator #5, – Personal Data for Municipal Services and the Tourism Industry.
Pursuant to the accountability principle, the existence of valid consent must be demonstrable by the data controller (accountability).

R12. User Control

DataVaults personal data app dashboard needs to provide certain functionalities of user control, in addition to the information that must be given according to the GDPR.

Being DataVaults a blockchain-based system, it has the potential to allow individuals to retain and exercise such control and even to understand, in a transparent manner, who has access to their information, but the system must be specifically crafted applying the Data Protection and Privacy by Design paradigm and by the fairness by design approach, to ensure that individuals' privacy and real control over their data is afforded to the system itself.

New, user-friendly ways should be developed and offered to individuals to better exercise control over data in DataVaults, for example for letting them effortlessly switch on and off the tracking or information sharing of the devices and applications, as well as for facilitating the correction, update or delete of data, for modifying who is entitled to have access to it

and for monitoring who actually accessed it, and for what purposes

R13. Data subject's rights

In DataVaults it is recommended that the right to object to processing is valorised as a powerful tool in the hand of the individuals. This requires well-designed and workable mechanism for opt-out, for effectively exercising an unconditional, 'no questions-asked' opt-out right.

Special attention should be given also to the right to access and correct one's personal data, as a precondition for empowering individuals to better detect unfair biases and challenge mistakes arising from algorithm-based solutions to determine assumptions and predictions.

DataVaults should concretely give the individuals the ability to use their personal data to benefit from them in a tangible way, thus moving towards the "featurization" of data protection, instead of an administrative burden.

As for data portability, it is relevant for DataVaults, since it can help in giving more control to individuals and in sharing with them the benefits of data processing and insights, letting individuals benefit from the value created by the use of their personal data. In this way the benefits of big data would be maximized in a more balanced and transparent way, helping to redress the economic imbalance between controllers and individuals and, at the same time, minimising unfair or discriminatory practices, whilst reducing the risks of using inaccurate data for decision-making purposes.

Considering the right to data portability, it is recommended that DataVaults tools are designed and developed taking into account, for instance, that individuals should be provided with access to their own data in portable, interoperable and machine-readable (usable and reusable) format; they should be allowed to delete, modify, transfer, or otherwise further process their own data, as well as to switch providers and take advantage of other third party applications to analyse their own data and draw useful conclusions (such as get personalized health care services or switch to a cheaper electricity provider).

Furthermore, the data portability-driven tools in DataVaults should be conceived also to allow individuals to use the data for their own purposes, besides for licensing the data for further use to third parties (in exchange of additional services or for cash value).

R14. Enforcement

This enforcement, potentially based on blockchain technology itself, should be conceived as complementary to the penalties and administrative fines imposed for any infringement of the GDPR. In other words, DataVaults is required to seek and deploy technological solutions that enable the enforcement of GDPR rules (including data subjects' rights) in a smart flavour, besides the other rules for data sharing, preventing the misuse of data. This will represent a key step in view of the fair attribution of value represented in data creation and sharing, effectively taking into account the multiple, competing interests at stake. Therefore, in DataVaults the concept of enforcement should be intended rather than limited to data

breaches and privacy violation, above all as a holistic process-wide approach of control and GDPR compliant access and usage control enforcement.

The DataVaults solutions regarding enforcement could be inspired by the data sovereignty model, promoted by the IDSA[14] ecosystem in relation to proprietary data. Whether IDSA's data sovereignty paradigm is mainly directed to safeguard data producers' control over data generated, DataVaults model could extend or adapt the same approach for ensuring data subjects' control over their data (and related enforcement), specifically elaborating access and usage policies and protocols for the purpose. In this sense, DataVaults technological layer should investigate enabling technologies to implement and enforce not only the terms and conditions set forth by the data sharing agreements, but also GDPR rules. Technologies for enabling usage control and enforcement that could be explored include, for instance, besides blockchain, DLT and smart contracts (where the efficiency of the enforcement is enhanced by automated execution of the provisions), also the sticky policies, digital rights management technologies and the APIs.

In this direction, DataVaults could take advantage of the findings of other EU projects in the field, such as BPR4GDPR "Business Process Re-engineering and functional toolkit for GDPR compliance". The achievements of this project, aimed at providing "a holistic framework able to support end-to-end GDPR-compliant intra- and interorganisational ICT-enabled processes at various scales[15], provide useful insights for a rule-based access and usage control framework capable of formalizing policies and thus filling the gap between the legal and the technical work as for GDPR compliance. In other words, its results can be explored in DataVaults for developing the appropriate formalisations in order to capture the legal concepts for the specification of the policy framework and rules (mainly stemming from the GDPR), in the context of the future elaboration of the DataVaults rule-based access and usage control.

R15. Fairness by Design

The substantive dimension of Fairness by Design implies efforts directed to develop tools capable of equal and just distribution of benefits and costs, without unfair bias, discrimination and stigmatization for individuals and groups. The DataVaults system should also move towards societal fairness, fostering equal opportunities and avoiding a situation where people are deceived or unjustifiably impaired in their freedom of choice. This is especially relevant in relation to the compensation schemes and mechanisms to be established in DataVaults. Besides data monetization approaches, other approaches need to be investigated based on other rewarding incentives, in order to avoid a situation where the poorest brackets of the population are disproportionately increasingly motivated to share data. Fairness also implies respect of the proportionality between means and balancing operations between competing interests and objectives. As for the procedural dimension of fairness, this entails the effective exercise of the data subjects' rights (rectification, erasure, object, etc.).

[14] International Data Space Association, https://www.internationaldataspaces.org/

[15] https://www.bpr4gdpr.eu/

R16. Effective "sharing the wealth" paradigm

The set of a win-win data sharing ecosystem is at the core of Big Data Value Association works and is envisaged, for instance, by the BDVA Position Paper "Towards a European Data Sharing Space - Enabling data exchange and unlocking AI potential", released in April 2019. DataVaults needs to move in this direction, also contributing to unlocking the social value of personal data. It is necessary to go beyond user consent and concretely move towards an effective "sharing the benefits" model in DataVaults, fostering individual human empowerment and flourishing and the common good of society, besides businesses' interests.

R17. Privacy Notice

Recitals 61 and 62 provides useful indications, respectively on the time of information and on the exception to the information obligation.

As indicated, as regards this requirement in DataVaults, it could be taken into account the layered approach (acknowledges by the Article 29 Working Party[16]) which was described hereabove in the Informed Consent requirement.

The exceptions to the information obligation could be relevant for DataVaults demonstrators, with possible reference to the processing carried out for scientific research purposes.

User notification in case of specific security risks could also be relevant in DataVaults, but this will be investigated at a later stage, when the project progress will allow a better understanding of the legal and ethical implications.

R19. Accountability

The principle requires both the respect of data processing principles and to be able to demonstrate compliance, with reference to, for instance: i) personal data categories and data formats intended to be used; ii) personal data sources; iii) purposes of the processing and legal ground on which the processing operation is based; iv) Technical systems involved (hardware, software and infrastructure); v) The internal organization of the processing entities involved and related human resources involved.

Tools able to prove that the core platform and the personal data app are functioning properly need to be included in the overall DataVaults platform, by creating an audit trail with logging. A data model/ontology capable of conceptually, logically and physically describing the structure and flow of the information and inferences, by enabling a clear re-traceability which specific data set is used for which analytical process and how the corresponding analytical results were generated (data, process and analytical provenance). As well, DataVaults technology should allow, for each process, to determine and prove the specific roles of involved actors, in order to allocate the legal responsibilities. To do this, a comprehensive role concept could be defined, functioning as a core precondition to identify

[16] Article 29 Working Party "Guidelines on consent under Regulation 2016/679", adopted on 28 November 2017 and revised and adopted on 10 April 2018

which of the participating organisational instances has to actively ensure the legitimacy of a data processing procedure. This is especially relevant in DataVaults, where complex structures (both at technological and at organizational/business level, potentially in a multi-stakeholder and multi-facet ecosystem) are expected to operate and therefore it is possible to classify either whole DataVaults platform-wide processes, or independent sub-processes (at component or at app level). One of the elements that, pursuant to the accountability principle, must be demonstrable by the data controller is the existence of valid consent.

The GDPR requires DataVaults to review and update, the chosen appropriate technical and organisational measures functional to ensure GDPR compliance and to be able to demonstrate that.

It is key to bear in mind in DataVaults that the accountability principle is qualified by the so-called risk-based approach, that will be described in the Data Protection Impact Assessment requirement.

R21. Data Protection Impact Assessment

The Consortium must adopt a risk-based approach and evaluate the ethics risks related to the data processing activities of the project, assessing the particular likelihood and severity of each risk to data protection, taking into account “the nature, scope, context and purposes of the processing and the sources of the risk”.

The assessment of the risk must be conducted in an objective manner to determine whether there is a "risk" or a "high risk", in order to let the data controller be particularly prudent to carefully consider their obligations when necessary. As for the risk-based approach (GDPR, Recital 75, 76), it requires consideration of what measures are appropriate in each case, depending on the scope, nature, context and purposes of the processing concerned, as well as of the risks of varying likelihood and severity for freedoms and rights of individuals.

The more severe and likely the risks from the proposed processing, the more measures will be required to counteract such risks. According to recital 75, examples of potentially risky processing relevant to DataVaults include: i) processing that may give rise to discrimination, identity theft, financial loss, reputational damage, unauthorised reversal of pseudonymisation, or any other significant economic or social disadvantage; ii) processing that might deprive data subjects of their rights and freedoms or prevent them from exercising control over their personal data; iii) processing of sensitive personal data; iv) processing for purposes of profiling; v) processing of personal data of vulnerable natural persons; Vi) processing involving a large amount of personal data and affecting a large number of data subjects.
The risk must be assessed in an objective manner to determine whether there is a "risk" or a "high risk".

For operationalizing the risk-based approach, DataVaults has in mind to develop/use specific tools and services. The Risk Exposure Dashboard will facilitate the calculation of the privacy risk exposure based on previous knowledge, depending on the data already available and shared and specific metrics, and will make it possible to notify individuals of their risk exposure. The Risk Management Service could be valorised as a high-value powerful accountability tool for the fulfilment of the informed consent requirement, but even for the

risk assessment component of the DPIA. This could strengthen the market exploitation of DataVaults technology.

In case of need of a DPIA, it shall contain at least: i) "a systematic description of the envisaged processing operations and the purposes of the processing"; ii) "an assessment of the necessity and proportionality of the processing operations in relation to the purposes"; iii) an assessment of the risks to the data subjects' rights and freedoms; and iv) the mitigating measures to be taken, including "safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate" GDPR compliance.

R23. International Data Transfer

Tools should be provided to address 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).

Special attention should be given to Art. 44 and Art. 46 GDPR, respectively setting forth the general principle for transfers and the transfers subject to appropriate safeguards. Also, Recital 101 should be addressed.

R24. Technical and organizational measures

The appropriateness of the implemented measures is to be determined taking into account a set of factors, respectively listed in each of the two provisions.

It is important to mention here again the risk-based approach adopted by the legislator (and elaborated somewhat in Recitals 75 and 76), that requires to consider the risk of varying likelihood and severity for the rights and freedoms of natural person.

Recital 78, on the other hand, provides some clarifications on the appropriateness of the measures.

Considering DataVaults framework, the following measure indicated by the legislator can be recognized as appropriate, inter alia: i) the pseudonymisation and encryption of personal data; ii) the ability to ensure the continuous confidentiality, integrity, availability and resilience of processing systems and services; iii) "minimising the processing of personal data, pseudonymising personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing".

R26. User and data protection friendly User Interface

An important element to consider is the wide range of data sources and to pay special attention in case where it includes sensitive information in the sense of Art. 9 GDPR.

A filter on those data categories could allow the UI to distinguish between consent requests on "normal" personal data and those involving sensitive data. It could be investigating whether introducing functionalities for automatically detecting when sensitive data (or particular subset of sensitive data, for instance in the health care demonstrator) is collected, using machine learning techniques or other techniques and filtering such data.

The following challenges could occur and need to be addressed:

- managing consent in a fine-grained way (including, for instance, partial granting or withdrawal of consent in some circumstances);
- managing the own data and exercise data subject's rights in an easy way, for instance as regard adding, deleting and rectifying personal data, and including also the possibility to access additional information in case of a data breach;
- switching back and forth between different consent modalities, such as always requiring explicit consent for personal data sharing in some situations and opting for convenient assumption of implicit consent in other;
- ensuring data portability and exporting the own personal information (for instance in an RDF format).

R27. Measures in case of profiling
This aspect is particularly relevant for the future uptake of the project's results, whose contexts could be very different. Recital 71 clarifies that profiling consists "of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her".

4.2 SECURITY, PRIVACY AND TRUST REQUIREMENTS

In the following subsection, 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, it is the mandatory requirements that will drive the design and development of the core security, privacy and trust services of DataVaults platform (in the context of WPs 2, 3 and 4) and the initial investigation presented in Sections 5 and 6.**

4.2.1 Requirements List

N.	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.	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	M	DataVaults Identity Access Management

N.	Short name	Description	ID	Supporting DataVaults Tool
		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 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 (their devices and their actions) should not be identifiable without breaching the non-repudation 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-repudation 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-repudations requirement of their transactions (SR3). Non-repudation 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 data stored. This: (i) should guarantee the protection of	M	DataVaults Crypto Suite (i.e., DAA,

N.	Short name	Description	ID	Supporting DataVaults Tool
		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 (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.	M	Trusted Platform Module as hardware-based Root of Trust
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	M	DataVaults TPM-enabled Blockchain computation and verification functionalities

N.	Short name	Description	ID	Supporting DataVaults Tool
		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.		
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

Table 3. Security, Privacy and Trust Requirements.

5 USER AND DATA SECURITY, PRIVACY AND TRUST SERVICES - SOTA

The vision of DataVaults is to provide a **secure, trusted, auditable and privacy-preserving platform** for data sharing economies that complements existing ICT deployments through the use of Blockchain (Section 6). This will enable enhanced **data privacy** and ownership safeguarding (**privacy by design**) and data provenance and sovereignty checking mechanisms, while respecting prevailing GDPR legislation. As will be described in Section 6, the platform will use Blockchains for enhanced data and transaction security. Blockchain is one of the most disruptive technologies related to data security today, but beyond the inherently sensitive nature of various personal and commercial data are the persistent challenges of interoperability, data matching, and data information processing, sharing and exchange. To this end, DataVaults will protect data and resources against leak or improper modifications, while at the same time will ensure data availability to legitimate users. Internal storage and ledger infrastructures, handling personal and/or corporate data, can track its provenance and are regularly audited to comply with **specified security and privacy policies and regulations**. This way users are in control of their own privacy and that of their devices, applications and services. For the former, users will be able to participate in the specification of privacy-related policies, which will then be codified in smart contracts, following the principle of **user privacy empowerment**. Depending on the selected privacy level, **privacy enhancement is achieved through the use of trusted computing technologies** (i.e., TPMs) as a central building block towards the provision of privacy-preserving signature schemes (e.g., Direct Anonymous Attestation (DAA)).

In this context, DataVaults-enabled applications might involve security- and privacy-sensitive data (as has also been envisioned by the formulated use cases defined in D1.1 which require strong **confidentiality, security and privacy assurances** at different levels to be supported by the system, which are further mandated by regulatory compliance requirements. Towards this direction, DataVaults will provide enhanced **information protection and secure data management, over the entire data lifecycle,** ranging from data generation, collection and storage to data search and deletion. Within all these operations, DataVaults will integrate advanced crypto primitives towards data security, user privacy and secure access control as holistic services to allow the trusted data movement between different entities and data infrastructures.

In what follows, we will present and assess the most suitable and robust **encryption technologies** needed to secure different types of information, while still allowing advanced knowledge discovery through the provision of **enhanced data search services** (i.e., Searchable Encryption – Section 6.2.1.2), and **advanced security and privacy-preserving primitives** (i.e., data anonymization and pseudonymization techniques) for authentication, authorization, attestation and verification through the use of trusted computing technologies. Such an analysis will serve as the basis (and provide valuable insights) on the identification of the most appropriate security technologies to be further investigated in WP2.

5.1 State of the art And Key Technology Axes

This section is devoted to discuss the state-of-the-art of the key technology elements (e.g., Trusted Platform Module (TPM)) and authentication, authorization, attestation and verification algorithms that will constitute the basic security and trust building blocks leveraged by the DataVaults framework. In relation to data sharing environments, this section provides a reference guide to the specific technologies that are embraced by the communities targeted by the DataVaults project. We can classify these elements into four different domains:

- the **Trusted Platform Module (TPM)** and the **host** user device containing the TPM;
- the **Cryptography subsystem** responsible for providing the crypto stack (i.e., encryption, signature, HMACs, etc.) to be leveraged by the DataVaults platform;
- the **Authorization, Attestation and Privacy** components providing user data privacy and information security services;
- the **Security Policy Enforcement** component.

5.1.1 Towards Decentralized Security- and Privacy-Enhanced Solutions

The advent of the Internet and the next-generation smart connectivity networks have been some of the most important technological changes experienced by the society in the last decades. They have a profound impact in the way we communicate, conduct business and socialize. As expected, such a major transformation brought quite a lot of difficult challenges to tackle, for example, in the fields of **regulation, standardization, privacy, ethics or economics**. Security (and privacy), as an essential factor for the development of any digital technology, soon became one of the fields where the scientific community devoted much of their research efforts. It is widely believed that the next major technological transformation that we will experience as a society will come in the next few years with the introduction of the Internet of Things (IoT). The interconnection, through the Internet, of users with everyday devices, home appliances and other items embedded with inexpensive electronics (sensors, actuators, connectivity endpoints) will result in the generation of an unprecedented amount of data coupled with strict data processing and sharing requirements that will undoubtedly also increase the amount of privacy and security challenges, in addition to the already existing ones in the Internet.

Towards overcoming these challenges, over the recent years, emphasis in data security and user privacy research has converged on the use of **Public Key Infrastructures (PKIs)** [80] **for credential management and privacy-friendly authentication services** through the use of short-term anonymous credentials, i.e., *pseudonyms*. The common denominator in such architectures is the existence of trusted (centralized) infrastructure entities for the support of services such as authenticated registration, pseudonym provision, revocation, etc for either systems users or Cyber-Physical Systems (CPSoS). While intensive research efforts have proven the security and privacy guarantees provided in PKIs, there are still a number of challenges to be conquered [81, 82].

Firstly, it is essential to provide efficient, reliable and in timely and privacy-preserving communications to all users and devices. The reliance on infrastructure entities within the overall architecture for such services raises questions towards a system's availability and

scalability in the case of a technical fault or attack. Secondly, many researchers have demonstrated the privacy weaknesses of varying pseudonym re-usage policies; even in the case of unconditional anonymity where frequently changing pseudonyms has been proposed to avoid user service linkability, it has been shown to be ineffective due to the timing information of changing pseudonyms [83]. Thirdly, in the context of revocation policies for removing misbehaving users from the network, this can only be achieved when the employed pseudonym scheme supports the resolution of participants' long-term identities from their pseudonyms [84, 85]. In this case, information about the revocation of a user's long-term credentials, is disseminated to other participants through Certificate Revocation Lists (CRLs) or other means. Besides being computationally intensive, this is harmful to the protection of their privacy [86].

If we are to fruitfully benefit from the evolution of such technologies and the advantages they bring in data availability, we need to enable a smooth transition from traditional centralized datacentre-based environments, which are vulnerable to a variety of security threats, to more decentralized data sharing networks capable of offering **data security and privacy management through policy compliant Blockchain structures**, where trust is shifted from the back end infrastructure to the users (**user empowerment**).

Towards this direction, **Trusted Computing** is a core pillar: it provides confidence in a system, especially if the system's behaviour isn't fully secure or might become insecure. It establishes whether a system is the intended system, i.e., whether it is doing what it is designed to do, and provides controlled access to keys and secrets that depend on the system's current behaviour. It also allows a compromised system to be restored by installing and replacing software.

The Trusted Computing Group (TCG) is a not-for-profit industry consortium that aims to provide standards for Trusted Computing technologies and to promote their usage. The TCG-defined methodology relies on the concept of Root of Trust (ROT). A Root of Trust is a component of a system on which all other trust is based, and which must be blindly trusted. It is worth to note that a Root of Trust is inherently unverifiable: if a system has a proposed Root of Trust and relies on another component to verify it, then this second component becomes, implicitly, the actual Root of Trust. While it is not possible to determine if a Root of Trust is behaving properly, there are defined certification procedures that allow manufacturers to provide assurance that a root has been implemented in a way that renders it trustworthily.

The most prominent ROT technologies nowadays are based on the use of Trusted Platform Modules. A TPM is a highly secure hardware component that, together with the BIOS, can serve as a root of trust; a hardware anchor on which secure systems could be built. Its main goals are to provide a protected space for key operations and other security critical tasks, cryptographic functions, measure and report the behaviour of computing platforms, store data securely and perform secure authorization.

It is designed to enhance **platform security** beyond the capabilities of software and shield **cryptographic and sensitive material from software-based attacks**. Moreover, augmenting

computers with these hardware modules adds powerful functionality in distributed settings, allowing us to reason about the security of these systems in new ways.

The first widely adopted version TPM 1.1b was released in 2003, which was subsequently revised to version 1.2, published in about 2005, and was later standardized by the ISO/IEC in 2009 [87]. The last major revision of TPM is version 2, released on 2014 and standardized in 2015. The last updates to the TPM 2.0 specification have been made early in 2018. TPM 2.0 has been designed with a "library" approach. This allows vendors to choose TPM functionality for different implementation levels and platforms. Also, new features and functions were added, such as the ability to implement new cryptographic algorithms as needed. This flexibility allows the latest TPMs to be used for many embedded applications, including but not limited to, automotive, industrial, cloud computing and IoT.

In the context of DataVaults, **TPMs will be used as the main ROT in the user devices**. The platform will investigate their use towards the design and development of advanced Blockchain-control services (Section 6) through the specification of novel TPM-based security and privacy-preserving protocols for advancing the state-of-the-art in scalability and computational efficiency when securing different levels of user privacy.

5.1.1.1 Trusted Platform Module Architecture

The aforementioned Trusted Platform Modules (TPMs) will be investigated as a central building block of DataVaults and form the basis for enhanced security, privacy and operational assurance guarantees both when it comes to user and data security and privacy but also for ledger management and maintenance. The smart integration of the TPM technology (through Infineon's Blockchain 2Go Starter Kit) will allow DataVaults to develop new Blockchain verification and operation methods. The DataVaults framework **will not only use TPMs for user authentication and access authorization or to build secure blockchain wallets, but also continuously attest and assess the security of involved devices in a privacy-preserving way and use TPM features to build efficient alternatives to rather inefficient or biased mining procedures**.

With the idea of becoming a hardware security anchor in mind, the TPM was developed to address a number of issues in the fields of security and privacy. The latest revision of the TPM2.0 specification dates from September 2016 (and this is the one that will be used as the baseline in the context of DataVaults). The main specification is structured into a set of four (extremely) detailed documents:

- **Part 1 – Architecture** [88]: This document describes the TPM operations in detail (e.g., how to create sessions, and all its variation types), and the rationale behind the TPM design.
- **Part 2 – Structures** [89]: This document presents a description of the data types, constants, and command returning values and error definitions.
- **Part 3 – Commands** [90]: This document presents the commands of the TPM and error conditions in detail.
- **Part 4 – Supporting routines** [91]: This document contains code for the supporting routines used in Part 3.

An overview of the conventional TPM architecture is given in Figure 4. In a nutshell, its architecture consists of: (*i*) a cryptographic module (including private/public key encryption, digital signatures, hash functions and MACs), (*ii*) a random number generator, (*iii*) a protected storage region (volatile and non-volatile), (*iv*) a management component, and (*v*) an authorization component. A TPM is conceived as a System-on-a-Chip, hence all security-sensitive services are executed within a closed system. This, together with a cryptographic key storage, allows for the realization of the protected capabilities, as documented in the Trusted Computing Groupg (TCG) Specification [88]. The command execution engine is a microcontroller which forms the basis of the TPM realization. This execution engine processes the incoming commands and controls the cryptographic engines accordingly. It is also intended for the realization of the cryptographic communication protocols.

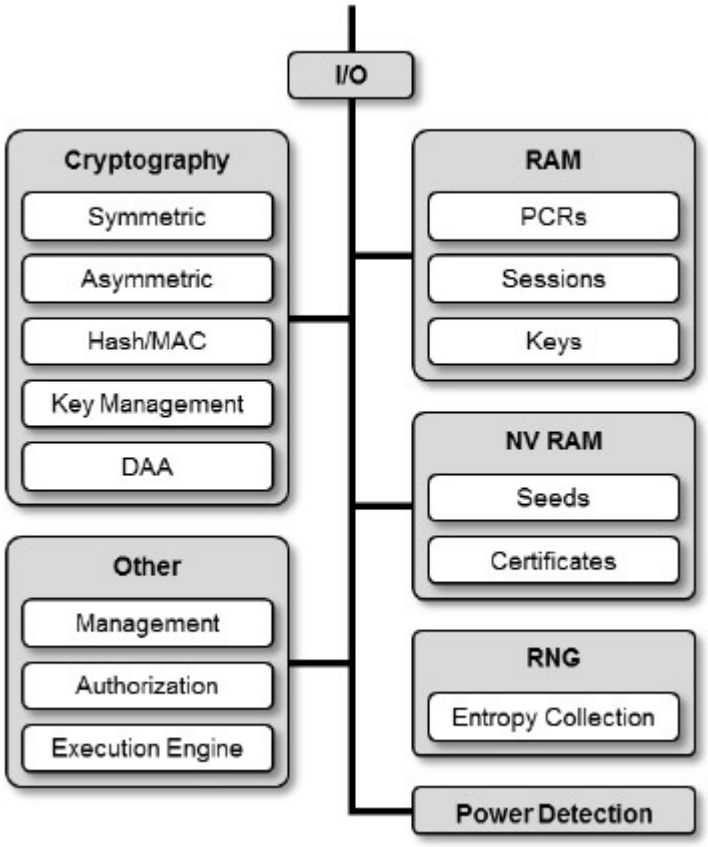

Figure 4 - Conventional TPM Architecture.

The **TPM Software Stack (TSS)** is a software specification that provides a standard API for accessing the functions of the TPM. In addition to the TCG's specifications cited above, we also refer the reader to [92], [93] and [94] for a more detailed discussion of TPM 2.0, and Trusted Computing in general.

In the following subsections, we describe in more detail the main functionalities of TPM2.0 that will form the basis for the enhanced user privacy and data security functionalities of the DataVaults framework.

5.1.1.2 TPM-based Authorization

This concerns how platform software can prove its authorization to call functions of the internal TPM. The TPM specification defines several types of authorizations, some of which maintain session state, to access the different entities and commands within the TPM. There are currently three authorization mechanisms:

- *Password*. A password is sent in clear with every command. No session is maintained in this authorization mechanism.

- *HMAC*. The password is set up at the beginning of the session. An HMAC is calculated on each command and responses received to determine its trustworthiness. HMAC authorizations maintain a session state.
- *Enhanced Authorization.* Are built on top of HMAC authorization sessions, and besides being based on a password, this kind of authorization also depends on TPM state including PCR values, external devices such as fingerprint readers or smart cards. Authorization conditions can be combined together to make complex authorization trees. This authorization mechanism also maintains state.

Also, the authorization mechanism defines roles, which control who can run a given command under what circumstances. There are three roles defined: user, administrator and DUP role. The DUP role is only focused to duplicate cryptographic material.

5.1.1.3 TPM-based Attestation

Attestation is one of the crucial services of a TPM. **It is the process by which a platform reports in a trusted way the current status of its configuration.** The basis of the attestation are the measurements recorded in PCRs (Section 5.1.1.5). They can then be read to know the current status of platform and be also signed to provide a secure report. The signed message can then be sent to the client. 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, or remote clients ask for an attestation of some value, and later they evaluate its trustworthiness. Attestation enables such clients to confirm whether the platform has been compromised. Additionally, the TPM offers means of certifying and auditing the properties of keys and data that cross the TPM boundary.

5.1.1.4 TPM-based Privacy

Every TPM has a unique public/private certified key-pair known as the Endorsement Key (EK). This can be used to prove to third parties (verifiers) that they are communicating with a genuine TPM. However, if this key was used to sign objects, it would enable verifiers to uniquely identify this TPM and link all transactions it makes. The current specification solves this issue by providing the TPM with the ability to create as many Attestation Identity Keys (AIKs) as the user wishes. AIKs act as pseudo-identity keys for the platform and can be used for different purposes and scenarios. To certify that these AIKs come from a genuine TPM, there is the need to create a new trusted entity, the Privacy Certification Authority. This approach, however, requires that the Privacy CA be highly available, as it is involved in every transaction for the creation of the AIKs. Moreover, if the Privacy CA and the verifier collude, the verifier will be able to uniquely identify a TPM.

In order to solve these issues, as well as the protocol involving the Privacy CA, there is also a protocol called **Direct Anonymous Attestation** (DAA) (Section 5.1.4), based on group signatures. This protocol does not require a highly available Privacy CA. It can transform the original credential into new unforgeable credentials that “look like” fresh credentials, while allowing that the different AIKs cannot be linked neither to the associated EK nor between them. DAA allows user-controlled linkability. That is, using the basename field of the data it provides for DAA, it can control whether a verifier can link two signatures or not.

One of the main drawbacks of the TPM privacy features comes when the EK of a TPM is compromised. Some concession on anonymity must be made to allow compromised and fraudulent TPM keys to be detected.

5.1.1.5 TPM Platform Configuration Registers (PCRs)

Platform Configuration Registers (PCRs) are one of the essential features of a TPM that allows it to act as an RTR. 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.

PCRs cannot be modified arbitrarily. When an entry is appended to a log of measurements, the TPM receives a copy of such entry (or a digest of the data described by the log), and the data sent to the TPM is employed to update the value of the PCR to its next value. The TPM can provide an attestation of the value of a PCR (or a set of PCRs) upon request, corresponding to the cumulative value of log measurements up to that point. This is used to verify the contents of the log. This attested measurement allows an independent entity to determine whether the platform has been compromised or not.

PCRs can also be used in conjunction with the authorization mechanisms to restrict access to a TPM-protected object, e.g., a decryption key. If certain PCRs do not have the required values, then the TPM will not allow access to that object. A well-known example of this use case is Microsoft BitLocker full disk encryption. BitLocker is used in conjunction with a TPM to ensure that the integrity of the trusted boot path of a platform (e.g. BIOS and boot sector) is in a trustworthy state, in order to prevent most offline physical attacks and boot sector malware. That is, full-disk decryption keys are only released if the PCRs report an expected set of measurements after platform boot.

Theoretically, a platform requires only a single PCR to record its entire history of measurements. However, this would make difficult to evaluate the platform state at different stages, and typically several PCRs are allocated to the various software layers. For example, some PCRs measure the booting environment, others record the OS environment, yet others are devoted to application measurements. Therefore, an individual that cared about which OS was loaded, but not what the OS had done since it loaded, could restrict its data to the set of PCRs that represent the booting environment, and ignore the remaining PCRs.

5.1.2 Cryptography Subsystem, Keys and Key Operations

5.1.2.1 Cryptography Subsystem

In recent years there has been a great increase in data of all kinds that are generated and shared due to the emergence of new technologies with IoT being at the centre of this evolution. A set of this data falls into the classification of “personally identifiable information” and “private data”, as classified by GDPR, and therefore requires special attention when it comes to their strict security, privacy and trustworthiness requirements.

The Cryptography Subsystem is responsible for implementing all the core crypto primitives to be leveraged by the DataVaults platform. Its main components are summarized below, providing the basic set of cryptographic primitives considered, followed also by a high-level description of more advanced security and privacy-enhancing protocols.

- ***Random Number Generation:*** The hardware nature of a random number generator (e.g., through the leveraged TPM) offers a better source of entropy to create cryptographic material, compared to software-based PRNGs. It nominally consists of an entropy collector, a state register, and a mixing function (typically, a hash function). The sources of entropy can be as diverse as noise, clock variations, air movement, and other types of events.
- ***Hash Functions:*** Current specifications (ISO/IEC 9797-2) allow the computation of hash functions as a single call for small inputs or as the usual start/update/complete sequence. An HMAC is a form of symmetric signature over some data. It provides assurance that protected data was not modified and that it came from an entity with access to a key value. To have usefulness in protecting data, the key value needs to be a secret or a shared secret.
- ***Message Authentication Codes:*** The most prominent algorithm is the HMAC described in the ISO/IEC 9797-2 using, again, the 2 modes of operation described for the hash functions.
- ***Asymmetric Cryptography:*** Asymmetric algorithms are usually used for attestation, identification and secret sharing. The most prominent asymmetric algorithms are based on RSA and ECC using prime curves. The functionalities provided are signature generation and verification, encryption and decryption. Several padding schemes are permitted for the input data, e.g., PKCS#1, OAEP, and no padding.
- ***Symmetric Cryptography:*** it is usually revolved around the use of encryption to encrypt data during a number of operations such as authentication or transport sessions, and also protect data that is stored outside a ROT. The block cipher modes references in the current specification are defined in ISO/IEC 10116:2006. Any symmetric block cipher supported by a TPM may be used for parameter encryption. Weak keys are not permitted to be used (some algorithms have known weak keys, if such a key is generated, it must be discarded, and a new key generated by starting over with another iteration). When a symmetric key is used for data encryption, the encrypted data has an HMAC. This HMAC is checked before the data is decrypted.
- ***Signature Operation:*** An entity may sign using either an asymmetric or symmetric algorithm. The method of signing depends on the type of the key. For an asymmetric algorithm, the methods of signing are dependent on the algorithm (RSA or ECC). For symmetric signatures, only the HMAC signing scheme is currently defined. If a key may be used for signing, then it will have an attribute to allow it for. A key may be restricted to sign messages with specific contents.
- ***Key Generation and Key Derivation:*** There are usually two types of key generation: either from the provided random number generator or derived using a key derivation function and a seed value, depending on the application. For the purpose of key derivation (from another seed value) the KDF used is specified in SP800-108, with HMAC as the pseudo-random function, and in SP800-56A.

Data is generated and processed in very different systems or devices in terms of their capabilities. This is one of the factors to consider when speaking of encryption algorithms as they usually need high computation effort. We must differentiate between environments in which, a priori, there is no limitation of resources, such as servers, personal computers and even tablets and smartphones, which have powerful microprocessors, high memory capacities and in which energy consumption should not be a problem. And a second environment in which the processing capacity is limited, with a reduced storage capacity and where there is not always a source of energy that ensures operation over long periods of time. This second case requires an accurate definition when we talk about data sources in IoT, with sensor networks, RFID, etc.

For systems without limitations, in recent years new algorithms with a more complex mathematical base and longer data length have been developed, such as **Elliptic Curve Cryptography (ECC) or Lattice-based** cryptography algorithms that improve processing times and memory needs compared to other asymmetric schemas such as RSA. These algorithms allow both data encryption and digital signature.

New cryptographic algorithms have also been developed that allow the execution of mathematical functions on encrypted data and they obtain results without decrypting the data at any time. This is the case of **Fully Homomorphic Encryption (FHE)** which provides the result cyphered and therefore must be decrypted. A first circuit for this family of algorithms was presented in 2009 [3]. Also, **Functional Encryption (FE)** which differs from FHE on providing the result not encrypted. In addition, FE allows the generation of private keys from a Master Key, which introduces different functions to be performed, or as in the case of the **Attribute Based Encryption** algorithms, where encrypted documents are associated with a series of attributes and the keys are generated based on a set of policies that allow controlling the decryption depending on the attributes we have. FE was initially proposed in 2004 [4]. Both FHE and FE have been the object of study during latest years demonstrating significant advances. Currently the efforts focus not only in developing new schemes but also in moving the existing ones to real world applications, like the FENTEC project[17]

The cryptography functionality provided by hashing algorithms, which was introduced in 1989 with the publication of the MD2 (Message-Digest Algorithm 2) by Ronald Rivest, it has had numerous advances, such as BLAKE2 (year 2012) [5] and SHA-3 (year 2015), the latter published as standard by NIST [6]. Although this does not mean that old algorithms do not remain safe or efficient. An example of this are crypto-based Blockchain currencies, while Bitcoin, for reasons of efficiency, uses a double SHA-256 that was published in 2004. Other platforms such as ETHEREUM have opted for more modern algorithms, in this case BLAKE2b announcing greater speed than SHA-3.

At the other end, we have systems with reduced processing resources, storage (data length and amount of memory) and energy consumption, which greatly limit the use of complex algorithms. These systems or devices are widely used in IoT.

[17] http://fentec.eu

As an answer to these problems, **lightweight data encryption** technologies have emerged for the last two decades. These are technologies, which are not as secure as traditional technologies focused on the previous group, seeking a balance between their three main restrictive factors and the security they offer. The implementation of these algorithms depends on where they are going to be executed, considering, for example, if the implementation is hardware- or software-based, the size of the CPU (8 bits, 16 bits, etc.), etc. For this same reason, old fashion schemes that are designed for systems with this type of characteristics seem to be taking on some relevance again.

These restrictions mainly affect hashing algorithms, which are used to ensure the integrity of the messages. The processing of hashing algorithms requires a fairly large register length, around 2^{64} bits of the traditional algorithms that must be reduce to a range of 64 to 256 bits as occurs with SPONGENT [7] or LESAMNTA-LW [8].

Another alternative that in addition to integrity also allows to verify the authenticity are the HMAC (Hash-based Message Authentication Code) schemes. To do this, they combine a Hashing algorithm with another secret key asymmetric algorithm. We can use different hashing algorithms, whose name is added as surname to HMAC, resulting in names like HMAC-SHA256 for example.

Similarly, symmetric encryption algorithms are also affected, although to a lesser extent: algorithms that work by dividing the document to be encrypted into blocks, such as those in the AES [9], with 128-bit blocks and keys up to 256 bits, must reduce the size of the blocks. There are proposals such as PRESENT [10] with a block size of 64bits and keys of up to 128bits, or RC5 [11] that despite being designed in 1994 its ability to vary the block size from around some austere 32 bits and key from 0 to 2040bits make you a candidate to consider.

As for asymmetric schemes, this requires a large processing capacity, for example if we compare AES with Elliptic curve the order is 1 to 1000 and higher. Although progress is being made with the definition of light ECC schemes such as ELLI (Elliptic Light) standardized within ISO/IEC 29192-4:2013/Amd 1, or more specific ECC schemes for digital signatures [12].

Given the growing concern for security and the increasing use of constrained devices, widely used in IoT and data networks, Lightweight encryption is experiencing a breakthrough and is driven by relevant actors:

- ISO / SEC has revised its ISO / IEC 29192 standard on Lightweight cryptography in 2019. In this review they have added in Section 2, the specification of two suitable algorithms for applications (PRESENT [10], CLEFIA [13]), in section 3 MAC algorithms are added (LightMAC, Tsudik's keymode and Chaskey-12) for integrity of documents, and part 7 for broadcasting authentication protocols.
- In turn, NIST and the Computer Security Resource Centre (CSRC) are carrying out a project to request, evaluate and standardize cryptographic algorithms for resource-constrained devices. Which, at the time of writing this document is immersed in the second round of candidate evaluation.

5.1.2.2 Keys and Key Operations

The correct and secure handling of keys in a cryptographic system is essential for its operation. After the description of a key internal structure (public and private areas), this section will go through the lifeline of a key from its creation (generation or derivation) to its destruction. These operations are referred as key management operations and they include mechanisms to store a key inside or outside a Root of Trust (i.e., TPM).

Key Structure: A Key Object is composed of two areas: a *public* and a *sensitive* area. Values within parenthesis denote the data type, as defined in [89].

The *public* area contains the attributes of the key and a public identity, including:

- *type* (TPMI_ALG_PUBLIC): algorithm ID used to create the key.
- *nameAlg* (TPMI_ALG_HASH): algorithm ID used as hash algorithm to compute the name of the object, it may be TPM_ALG_NULL
- *objectAttributes* (TPMA_OBJECT): usage, authorization, duplication, creation, persistence
- *authPolity* (TPM2B_DIGEST): authorization policy
- [type]*parameters* (TPMU_PUBLIC_PARMS): parameters for the algorithm specified as *type* (e.g.: key size)
- [type]*unique* (TPMU_PUBLIC_ID): for asymmetric key it will be the public key, for symmetric it will be a value hashed of information in the *sensitive* area

The sensitive area contains data that are required to be encrypted, including:

- *sensitiveType* (TPMI_ALG_PUBLIC): type of object in the sensitive area, it must be equal to the *type* parameter in the *public* area
- *authValue* (TPM2B_AUTH): authorization value for the object, it's a bite array with length equal to the length of the digest produced by *nameAlg*.
- *seedValue* (TPM2B_DIGEST): it may represents the optional protection seed (for a parent key) or an obfuscation value
- [sentitiveType]*sensitive* (TPMU_SENSITIVE_COMPOSITE): parameters dependent on the *sensitiveType* (e.g.: private key for asymmetric key)

Key Generation: Keys can be generated in two different ways. The first way is to produce a key starting from a random number generator. The second way is to produce a key starting from a Primary Seed following a KDF.

A key can be generated by deriving it from another secret value. The TPM has 3 primary seeds, which are large random numbers stored persistently in the different TPM hierarchies. Generating a key using one of these seeds creates a hierarchy of keys. See Section 2.1.4 for further details.

The TPM uses two different KDF schemes: one scheme for ECDH (Elliptic curve Diffie-Hellman) and one for all other crypto operations. These schemes are based on hash-functions (Section 2.2.1.1). For ECDH the KDF is SP800-56A, for all the others it is SP800-108.

Key Management: In a TPM, the endorsement key (EK) represents the root of trust. Typically, the EK is a unique public-private key pair that is generated individually for each TPM. Once generated and stored on a specific TPM it cannot be replaced or removed anymore. Most TPM vendors offer the generation of EKs as a service. So, the manufacturer can be identified through the public key. Typically, the OEM generates a X.509 certificate with the EK [14].

The EK offers various essential TPM features. It ensures that the TPM cannot be replaced by another. Furthermore, it supports the creation of a TPM owner. Using a secret provided by the owner a Storage Root Key (SRK) is generated using the EK. However, the EK itself cannot be used by the owner to create a signature. The owner secrets and SRK are encrypted independently with the EK public key. Whenever the TPM needs to use these secrets, they are decrypted internally.

Depending on the manufacturer there is a constant or fixed key hierarchy existing in every TPM. Other keys are of the type “managed”. Thus, when designing a host system, a design engineer has to keep the key hierarchy in mind.

There are six different types of keys for different usage purposes [15]:

- TPM_KEY_SIGNING: to sign data
- TPM_KEY_STORAGE: to encrypt other keys in the hierarchy (e.g. EK and SRK)
- TCG_KEY_IDENTITY: identification of a platform (identifies one specific platform independently from the number of users)
- TCG_KEY_AUTHCHANGE: authentication
- TCG_KEY_BIND: data are bound or unbound from the TPM (encryption/decryption)

The TCG defines a minimal key-hierarchy relating to EK and the SRK. Depending on the manufacturer there is a constant or fixed key hierarchy existing in every TPM. Other keys are of the type “managed”. Thus, when designing a host system, a design engineer has to keep the key hierarchy in mind. A typical PC-based key hierarchy is shown below.

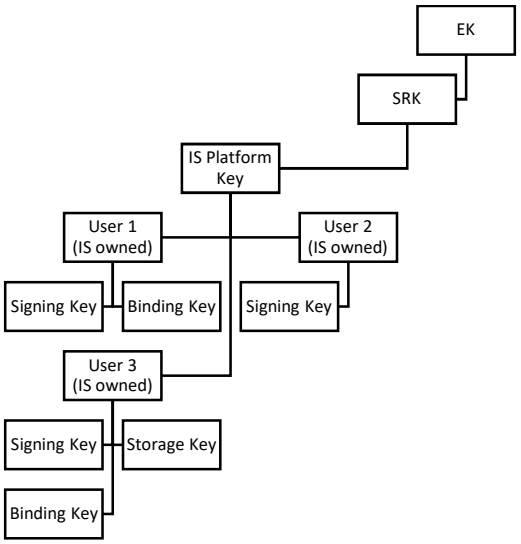

Figure 5 - A simple PC-based key hierarchy example [15].

To offer TPM operations without compromising privacy, a TPM should be able to prove that a key was created and protected internally (without being exposed to a potentially compromised host) and the recipient of that proof should not be able to infer any knowledge on the ID of the TPM that was involved in creation and protection of the key [14] (thus, achieving unlinkability). Therefore, TPMs provide an attestation protocol to make Attestation Identity Keys (AIKs), which are pseudo-identity keys for the host platform. Then,

the EK of the vendor can be used to prove that an AIK originated with a TPM without proving which TPM the AIK originated from. There is no limitation of how many AIKs can be created. This means, it is possible to destroy AIKs after crating and using them, or having multiple AIKs for different purposes.

Blockchain Security 2Go Key Management: In Blockchain systems, knowing a secret private key is directly associated with the control rights for an account. Consequently, it is important to protect the private key. The Blockchain Security 2Go cards feature hardware-based protection mechanisms to generate and store private keys in a secured way.

One Blockchain Security 2Go card can generate and store 255 private-public key pairs. Additionally, it is possible to import a key pair that is derived from a password (seed) that is provided by the user. This is achieved with the encrypted key import feature [16].

The Blockchain Security 2Go cards support on-card key generation providing highly secured private keys. This is achieved with a high entropy hardware-based random number generator.

The encrypted key import feature allows that on different Blockchain Security 2Go cards the same private key is generated. The user can provide a password (seed). From that given seed a private key is derived with the standardized key derivation function as defined in NIST SP 800-108 [17]. This allows that multiple cards can store the same private key. However, the private key itself is derived and stored on-card and is not known to the user.

5.1.3 Authentication and Authorization

Authentication consists of checking that the entity (potentially a human being) trying to access a resource is whom it claims to be. As the entity requesting a resource may potentially not be a human being but an electronic component, we consider two kinds of authentication:

- **Authentication of users**, refers to checking that a human being trying to access a resource of the system is whom he/she claims to be.
- **Authentication of devices**, refers to an electronic component, such as the network card of a server or a mobile device, is whom it claims to be.

Though it may seem that authentication of users and authentication of devices can be similar, actually there are several important differences:

- An electronic device could potentially authenticate on behalf of a human user;
- The mechanisms for checking the identity of human beings often differ from the ones used for devices. For instance, biometry can be used for authenticating a human being, but it only makes sense for devices when they are authenticating on behalf of a human user. Additionally, mechanisms for authenticating devices might be very awkward or even impractical to be used by a human being;
- The process of authenticating any entity is often associated to the process of providing information to that entity, which is name identity provision, and the information applicable to a human being is not the same as the one applicable to an electronic component.

Regarding the authentication of devices, there are several mechanisms to do it, but the most widely used are the ones (as aforementioned) based on PKI infrastructures. Among the various mechanisms used for this, there are two especially important:

- **The use of TLS for cyphering transmissions and authenticating the peers of a connection.** TLS is the transmission protocol normally used jointly with HTTP to protect transmissions carried over it. It is currently considered to be robust and it allows for authenticating both the server and the client of a connection. However, it has the drawback that the distribution and renewal of the certificates that it is based in can be quite complex
- **Self-sovereign identification** is a special kind of PKI infrastructure which, among other things, solves the problem of distributing keys associated to them, removing this way the main drawback of PKI mechanism, and it is widely used with Blockchain technologies, as it its own authentication system.

Regarding the authentication of users, it can be quite more complex than the authentication of devices because human beings are not as immutable as machines, and our characteristics are not as common as with them. For instance, even one of the most popularly accepted identifiers for people, which are fingerprints, are not guaranteed to be always available or readable. Another example could be the face, which likely to change as time passes and people grow older. Additionally, there are mechanisms for an attacker to impersonate people at a given moment. For instance, an attacker could cut a finger of its victim to pass a fingerprint control or may eavesdrop him while he is typing his keyword in the system.

For this reason, ideally a strong authentication mechanism for users should combine several measures of different nature. For instance, it could combine a password with something that the user has or the face of the user with a password. This is what is named **multi-factor authentication**. Note that combining facial recognition with fingerprint recognition would not be multi-factor authentication, because though it would involve two different authentication mechanism, they would both have the same nature (the body of the user). On the contrary, a smart PKI that required a password would be a multi-factor authentication, because it would combine two authentication mechanisms of different nature: something that the user owns with something that the user knows.

Though there were previously several different architectures for authenticating users, normally this task is currently delegated in an **Identity and Access Management (IAM) system**, which is a software that takes care of authenticating a user and even to perform the authorization of the operations issued by him. We will talk further on Authorization later on this section.

There are multiple IAM systems available in the market, which has caused the development of several standards for authentication. The main ones are:

- OpenID, which is a protocol which allows combining different authentication providers, each of them with their own authentication technologies, with a common authentication protocol. OpenID, which runs on Oauth, works with oAuth tokens, which imply that once a user has get authenticated with openId and get his oAuth token, applications may use this oAuth token once and again to avoid the user having

to authenticate for each of them. However, the drawback of openID is taht the oAuth token used by openId is valid only for those applications that recognize the oAuth server that issued it. For that reason, even though openID provides single sign on, it is limited to the set of applications that support the same oAuth provider

- SAML is a pure distributed authentication protocol from the OASIS standardization body that allows several applications to make queries named assertions to different authentication servers that do not need to be related with each other. As such, SAML is more powerful than OpenID, but this power comes at the cost of complexity. SAML is much more complex and hence more difficult both to process by a computer and to be learnt and understood by a programmer, which causes that not all IAM systems support it, which hinders its power
- Self-Sovereign Identities (SSI) are a special kind of authentication cantered on the user. The authentication process is based on PKI infrastructure. Basically, besides providing the keys for authenticating the user, it also provides keys for protecting the information associated with him through encrypting this information, so it is necessary a private key to access it. As the private key is always controlled by the user, and this key needs to be retrieved every time that the information is accessed, this grants the user the power to decide whether to grant access to this information or not.

Authorization is the process of checking that an authenticated user has permissions to resources that it is trying to access. Note that authorization is closely tied to authentication. There is little point in authorizing a non-authenticated user, because if the user is not authenticated, there is no way to know for sure who he is, and usually the resources that the users are entitled to access depends on the information that is associated to this user. For the case that we want some resources available for everybody, this is often carried out by using a generic 'guest' user, which only has access to these public resources.

There is a plethora of authorization mechanisms, which depend on how complex are the criteria for deciding whether to grant access to a given resource or not. The most important ones which will be investigate for DataVaults are:

- **Role Based Access Control (RBAC):** The user has a special attribute associated to him, which is a list of roles granted to him, and the system holds a list of the resources associated to each role. In practice, this means that a user will be granted access to the union of resources associated to each of the roles he is granted. For instance, if the user has roles r1 and r2 assigned, and r1 is assigned access to obj1 and ob2 and r2 is assigned access to obj2 and obj3, then the user will get granted access to obj1, obj2, and obj3. A simpler version of RBAC may omit the ability to combine different roles and / or to define the resources associated to each role through simply using monolithic roles
- **Attribute Based Access Control (ABAC):** While RBAC consists on deciding the access based on a single and fixed attribute of the user, which is her role, ABAC consists on deciding based on each possible attribute of the user. At first, RBAC looks as a special, simplified specific case of ABAC. However, the ability of assigning list of resources to each role is not necessarily a characterising of ABAC.

- **Policy Based Access Control (PBAC)** is the most powerful mechanism of authorization from the point of view of the ability to define complete criteria for the decision of granting access. PBAC allows defining access based on each possible attribute of the user, but goes further than ABAC, because it does not only allows to take into account the attributes of the user, but any information present in the system, and define this access criteria in separate files named access policies, which allows the access criteria to be defined independently from the code needed to evaluate them
- **Encryption mechanisms:** Unlike the previously mentioned mechanisms, encryption mechanisms are not focused on the ability to provide criteria for deciding whether to grant access to a resource, but on cyphering the resources to be protected so only those users that have the key needed to decipher them can access them. As these keys are meant to be in possession only of the owners of the information, this is a mechanism much used for user-centric systems, such as the ones using SSI.

Regarding the applicable standards for authorization, the current most important ones are:

- **XACML** is a standard for defining access policies from the OASIS standardization body. It is fully compatible with SAML and in fact SAML assertions can embed XACML policies in them. XACML is an extremely powerful language that allows defining almost any possible access criteria, but it has the drawback of being extremely difficult to learn, too;
- **Oauth 2.0** is a standard for authorization. OAuth is much used because openID is built on it. Basically, OAuth is a protocol for gathering the information required for taking the decision of whether to grant access or not to a resource, and this information may be the decision itself, and in fact that is possibly the most common way it is used.

5.1.3.1 Authorizations and Sessions in a Trusted Platform Module

Authorizations might also relate to mechanisms granting someone access to a TPM entity. The properties of that entity, often defined at creation time, determine the kind of authorization that is required by each role. The TPM 2.0 Specification considers 3 roles: the USER role is used for the normal uses of a key (e.g., signing with a signing key, or loading the child of a storage key); the ADMIN role controls the certification and the changing of the authorization value of an object; and the DUP role is only used for the duplication of keys.

Authorization may be granted by two means. The first corresponds to a proof of knowledge of an authorization value, also known as a password. This can be achieved by sending the password in the command authorization area, or via an HMAC whose behavior is determined by the password. The second means is through a policy digest, which requires that specific tests or actions are performed before an action is completed.

A session is defined to be a collection of TPM state that changes after each use. They provide means to communicate authorization data, audit a sequence of commands, build a policy digest, and encrypt command parameters.

In the TPM, there is a single, always-available password session that is used to authorize a single TPM command. Because of this, a client never needs to start a session to be granted

authorization with a password. It suffices that he passes the password in clear text format to the TPM as part of a command. This type of authorisation is of limited flexibility and presents security issues when a TPM is accessed remotely.

Sessions can be created through the TPM2_StartAuthSession command. They have associated session and HMAC keys. The values of the keys are determined not only by the authorisation value of the entity that is being accessed but can also depend on salts and on the authorisation value of another entity. When a session is started, the caller might indicate a size of nonces and an initial nonce. After initialising the session, the TPM returns a nonce generated by it. Each time the session is used for authorisation, nonces are updated, and the session and HMAC keys are updated accordingly. An entity handle might be sent with the TPM2_StartAuthSession command to indicate that that entity's authorisation value should also be included in the calculation of the session and HMAC keys of the session being initiated.

Sessions might be of 3 types: HMAC, policy or trial policy:

- When an **HMAC session** is in place, a client might compute an HMAC of the digest of the command parameters. Since the HMAC key depends on the entity's authorization value, the correct computation of the HMAC proves knowledge of the authorization value. If the entity's properties are compatible with this type of authorization and authorization role, the command will execute successfully. The command response parameters may also be HMACed, guaranteeing their integrity and authenticity.

- Access to entities might also be made dependent of a **policy session**, ensuring that a sequence of conditions have been satisfied before that entity can be accessed. A policy session is a form of enhanced authorization to allow for complex type of authorizations. It may include authorization based on TPM command sequences, TPM states or information coming from external devices (e.g., fingerprint and retina scanners, smart cards etc…). The policy is encapsulated in a value that is associated with the entity.
- A **trial policy session** provides a means to compute a policy value that can be associated with an entity. Like in a normal policy session, after the session is created, a number of commands are issued that update the trial policy session digest. In contrast to a normal policy session, all the assertions are assumed to be true, and the trial policy digest is updated accordingly. After the computation of the trial policy trial, a digest has been finalized, the policy value can be read from the TPM. Then, when creating an entity, this value can be set as the policy value associated with that entity. Trial policy sessions cannot be used to be granted access to entities.

Per-command session modifiers are available. In the case of HMAC sessions, one may encrypt the first parameter of certain commands that are sent to a TPM; or ask for a response parameter to be encrypted; or ask for commands to be audited. Similar options are available for the policy sessions, apart from the auditing. Two modes of encryption are available: CFB and XOR. The former requires both access to a block cipher and a hash function, while for the latter access to a hash function suffices. The type of encryption to be used is established at session creation time. For the CFB mode, a KDF is used to produce both the key and the Initialisation Vector (IV) from the session key and the nonces. For the

XOR mode, a one-time pad is produced with a KDF using the HMAC key and the nonces as input.

A host may maintain a record of the command and response parameters that are passed between it and a TPM. As these commands are issued, a host may furthermore request the TPM to extend the command and response parameters into an audit digest, as part of an HMAC session. An auditor can later request a signed copy of the audit digest to validate the integrity of the host's log. In addition, a host may have a single exclusive audit session, which may be used to prove to an auditor that no other commands were interleaved with the logged sequence.

5.1.4 Enhanced User and Data Privacy

As described in Sections 2 & 3, GDPR explicitly emphasizes the principles of **"privacy by design"** and **"privacy by default"** in data sharing economies. Any data sharing environment, as the one envisioned in DataVaults, should incorporate technical means to protect user and data privacy in its design. This can be translated to the *minimum disclosure, conditional anonymity, unlinkability and forward & backward privacy* requirements, described in Table 2.

Towards this direction, DataVaults will include the provision of secure, robust, and efficient attestation, verification and privacy-preservation methods to check the internal state of a user (or cyber-physical system), whose level of trust has not been verified, towards establishing its trustworthiness and privacy. This is considered as one of the main goals towards **"security and privacy by design"** solutions, including all methods, techniques, and tools that aim at enforcing security and privacy at software and user level from the conception and guaranteeing the validity of these properties. For privacy, DataVaults will leverage advanced crypto primitives, namely **Direct Anonymous Attestation (DAA)** [99], whereas for security and operational assurance, it will enable the provision of **Control Flow Attestation**.

The reason behind employing attestation mechanisms as a mean of operational assurance is twofold: First of all, one of the main challenges in managing device and network security in today's heterogeneous and scalable infrastructures is the lack of adequate containment and sufficient trust when it comes to the behaviour of a remote system that generates and processes mission-critical and/or sensitive data. An inherent property in DataVaults is the codification of trust among computing entities that potentially are composed of heterogeneous hardware and software components, are geographically and physically widely separated, and are not centrally administered or controlled. By leveraging the artefacts of traditional security infrastructure (such as digital signatures, certificates and assurance statements) coupled with advanced crypto primitives (such as run-time property-based attestation) and building upon emerging trusted computing technologies and concepts, DataVaults will convey trust evaluations and guarantees for each network entity.

This high level of trustworthiness which will not only include integrity of system hardware and software but also the correctness and integrity of the generated data flows will, in turn, reduce the overall attack vector and allow for the more effective operation of the DataVaults framework. This will allow the secure configuration, deployment and operation of distributed, scalable "Systems-of-Systems" infrastructures.

5.1.4.1 Attestation Protocols

In general, remote attestation is a mean of integrity verification of software running on a remote device. It is a mechanism, typically realized as a challenge-response protocol, which enables a trusted party (verifier) to obtain an authentic, accurate and timely report about the software state of a potentially untrusted remote device (prover). The verifier then checks whether the reported state is trustworthy, i.e., whether only benign software is loaded on the prover.

On the privacy side, **DAA** is a platform authentication mechanism that enables the provision of privacy-preserving and accountable authentication services. DAA is based on group signatures, which give strong anonymity guarantees [100]. The key security and privacy properties documented in [101], [102], [103] are:

- **User-controlled Anonymity:** Identity of user cannot be revealed from the signature.
- **User-controlled Linkability:** User controls whether signatures can be linked.
- **Non-frameability:** Adversaries cannot produce signatures originating from a valid TPM.
- **Correctness:** Valid signatures are verifiable, and linkable, where needed.

A DAA scheme considers a set of issuers, hosts, Trusted Components (TCs), and verifiers. A host and its TCs together form a Trusted Platform. An issuer is a trusted third-party responsible for attesting and authorizing platforms to join a network. A verifier is any other system entity or trusted third-party that can verify a platforms' credentials in a privacy-preserving manner using DAA algorithms, i.e., without the need of knowing the platform's identity. The TCG has standardized the ECC-based DAA scheme in the TPM 2.0 Specification. This specification has also been published as the international standard ISO/IEC 11889:2015 [54] and comprises five algorithms: SETUP, JOIN, SIGN, VERIFY and LINK.

In a nutshell, DAA is essentially a two-step process where, firstly, the registration of a TPM is executed once, and during this phase the TPM chooses a secret key (SETUP). This secret key is stored in secure storage so that the host cannot have access to it. Next, the TPM talks to the issuer so that it can provide the necessary guarantees for its validity (JOIN). The issuer then places a signature on the public key, producing an AIK <*cre*>. The second step is to use this <*cre*> for anonymous attestations on the platform (SIGN), using Zero-Knowledge Proofs [104]. These proofs convince a verifier that a message is signed by some key that was certified by the issuer, without knowledge of the TPM's DAA key or AIK <*cre*> (VERIFY). Of course, the verifier has to trust that the issuer only issues <*cres*> to valid TPMs. More details on the underpinnings of each one of these phases and various proposed DAA schemes will be given in the context of WP2.

Based on the security and privacy requirements that have been specified for the three envisioned Reference Scenarios [42], the *anonymity*, *pseudonymity* and *unobservability* properties are built into DAA's algorithms, JOIN and SIGN / VERIFY by using anonymous digital signatures. Therefore, third-parties cannot identify and link subsequent service requests originating from the same user/system. This is also true in the presence of colluding third-parties. The JOIN protocol is intentionally not privacy-preserving as the Issuer needs to

be aware of the user/system to be authenticated. However, successful completion of the protocol results in the user/system solely owning a DAA credential.

Unlinkability (and/or different levels of *user linkability*) is controlled by the user through the DAA SIGN / VERIFY phases. A user/system has control over its credential, and can decide whether or not to "blind" it, thus, producing pseudonyms (and revocation) that are linkable. In addition, DAA also provides non-frameability and correctness properties which are security attributes that are vital to the envisioned scenarios. DAA ensures that only valid and trustworthy TPMs are able to join a network by ensuring that the endorsed TPM keys have not been previously compromised. This ensures that TPMs only produce valid signatures and can only be linked when specified by a particular authorized service.

On the data security side, there exist different kinds of attestation, particularly **static attestation** and **dynamic attestation** [105]. Static attestation allows the attestation of static properties and configuration of a remote platform. The most prominent example is the attestation of the integrity of binaries [106]. As the name implies, dynamic attestation deals with dynamic properties of the runtime. For instance, it is concerned about the execution and data flow of programs, and not the static integrity of binaries. Naturally, attesting dynamic properties is significantly more challenging than attestation of static (already known) properties. Hence, the majority of research has focused on static attestation including industry effort in the Trusted Computing Base introducing secure and authenticated boot loading mechanisms of operating systems. However, given the continuous attacks on dynamic properties such as zero-day exploits which corrupt program's control flows, static attestation alone cannot be considered a viable security solution in the long-run, and needs to enhanced with advanced dynamic attestation mechanisms.

There does not yet exist a comprehensive design nor an effective as well as efficient implementation to enabling dynamic attestation. The most prominent approach in this context is **Control Flow Attestation** [107]. Control Flow Attestation is one of the most important dynamic properties at the software layer since it captures diverse instantiations of software exploits that hijack a program's control flow. In DataVaults, we will leverage automated and scalable behavioural-based attestation techniques focusing on the attestation of properties of software and hardware for cyber-physical systems. For this, we plan to adopt and extend static and dynamic attestation techniques so that both static and run-time properties of a remote platform can be attested.

Control-flow attestation is one of the most important dynamic properties at the software layer since it captures diverse instantiations of software exploits that hijack a program's control flow. Such attacks tamper with state information in the program's data memory area, e.g., the stack and the heap. Software bugs allow an attacker to arbitrarily alter state information and hijack the program flow of applications to induce malicious operations. While traditional attacks require the attacker to inject malicious code, state-of-the-art attacks such as return-oriented programming leverage code that is already present in the vulnerable application thereby bypassing modern mitigation strategies. In other words, the attacker resembles malicious codes through a combination of already existing benign code pieces. In contrast to traditional PC platforms and mobile phones, software exploits against Internet of Things (IoT) devices can have severe safety consequences. Consider a modern

network which features a vast amount of heterogeneous hardware and software components with hundreds of millions of lines of code. A common theme of such composable infrastructures is that all of them are pushing the envelope with respect to how many application instances can be packed efficiently onto a certain physical infrastructure footprint. This co-existence of multiple micro-services, multiple applications, or even multiple tenants, enables a variety of Advanced Persistent Threats (APTs) to be exploited by adversaries.

5.1.4.2 User Personas

Besides such *cryptographic* privacy-preserving approaches (non-cryptographic approaches featuring a policy-based authorization infrastructure are described in Section 5.2), focusing mainly on user privacy, DataVaults will also investigate the integration of another line of privacy enhancement services, focusing on data anonymization, namely **Digital Twins**; **towards detecting privacy concerns and minimizing breaches and associated risks to which users can be exposed when sharing primary personal data** (Section 6.1).

Digital Twins, in their generic form, are virtual replicas (digital representation) of the "states" of physical devices (i.e., user devices) ranging from the software tasks running in the device to the data generation and collection processes. Especially for the latter, and as more complex "things" become connected with the ability to produce and share data, the concept of such digital twins enables the provision of strong data anonymization services towards the generation of aggregated "User Personas" (through the **blending of data**) which can be considered as fictional representative users. Personas and associated analytics can then be used to reveal/extract knowledge of interest (i.e., user activities, records, etc.) without compromising the users' real identities or exact activities performed, thus, ensuring total anonymity and privacy. In this way, digital twins can benefit data sharing businesses and economies by being able to extract patterns, insights and knowledge from aggregated data, collected from various sources, without compromising the privacy of each of these data sources and being completely aligned with the strict requirements enforced by GDPR.

In this context, the main challenge is this of publishing/sharing such microdata (for further data processing and knowledge extraction) without revealing sensitive information: Consider, for instance, a DataVaults-enabled data holder that wants to share a version of its private data with other economic operators and/or third-party data collectors. *How can the data holder release a version of its private data with strong guarantees that the individuals who are subjects of the data cannot be re-identified while the data remain practically useful?* It has been demonstrated [108] how *quasi-identifiers*, set of specific attributes (e.g., gender, date of birth, etc.), can be joined with information obtained from diverse sources in order to reveal the identity of individual records. This challenge has led to the privacy preserving paradigms of **k-anonymity and l-diversity and a set of accompanying policies for deployment**. k-anonymity protects against the identification of an individual's record. l-diversity, in addition, safeguards against the association of an individual with specific sensitive information.

In a nutshell, a data release provides k-anonymity [108, 109] protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release. For every record in a released table there

should be at least *k*−1 other records identical to it along a set of quasi-identifying attributes. Records with identical quasi-identifier values constitute an *equivalence class*. *k*-anonymity is commonly achieved either by **generalization** (e.g., show only the area code instead of the exact phone number) or **suppression** (i.e., hide some values of the quasi-identifier), both of which inevitably lead to information loss. Still, the data should remain as accurate as possible, in order to be useful in practice. Hence a tradeoff between privacy and information loss emerges.

In parallel, the concept of *l*-diversity [110] was introduced to address the limitations of *k*-anonymity. The latter may disclose sensitive information when there are many identical *sensitive attribute* (*SA*) values within an equivalence class (e.g., all persons suffer from the same disease). *l*-diversity prevents uniformity and background knowledge attacks by ensuring that at least *l SA* values are *well-represented* in each equivalence class (e.g., the probability to associate a tuple with an *SA* value is bounded by 1/*l*). [110] suggests that any *k*-anonymization algorithm can be adapted to achieve *l*-diversity.

DataVaults will investigate the integration of such data anonymization techniques for efficiently solving the **privacy-constrained and accuracy-constrained anonymization problems that exist in such data sharing environments**. The framework will investigate the provision of **context-aware user-centric adaptation services** for either enabling the users themselves to apply such generalization and suppression to the data before sharing (the level of anonymization, of course, will be defined in adequate security policies that will be deployed via smart contracts) or allowing the platform to perform such k-anonymity and l-diversity techniques to the collected data before storage and further sharing with third-party data requestors. More specifically, DataVaults will incorporate a **set of algorithms that can process the output of user context information and determine the level of security level and anonymization that needs to be applied based on some predefined "user persona" profiles**.

Operational Challenges: Integration is a key challenge. In many data sharing environments, there will be multiple systems interacting, each of which has numerous assets within it. In turn, these assets comprise many more components. If each of these systems, assets and components has a digital twin, this presents significant additional complexity and requires that issues are addressed at both a high level and a granular level of detail. Operating at such a level of complexity with such a potentially large number of data sets, **data quality** and **data integrity** are key issues. It will be necessary to be confident that the data comprising the digital twin is accurate and reliable.

Legal & Ethical Issues: To address this issue (as was also described in Section 2), allocation of **responsibility for data accuracy** between data creators/controllers and data processors will be a key contractual concern. These risks can be mitigated by creating roles to manage digital twin resources and compliance within an organization. Many cloud providers also offer security and identity tools to manage these types of risks. Ownership of digital twins and the **intellectual property** associated with them might also be a key consideration. There is room for digital twin based "product as a service" business models which could also include the selling of physical object-related performance data or physical object pricing based on performance data.

5.2 SECURITY ENFORCEMENT

Security policy enforcement is executed both during the **design- and run-time phases of product development and can target both users and software that is running in the system** (Figure 5). The design phase enforcement will enforce policies that are expected to reach a specific security level. If some criteria fail and the desired level is not met, the run-time enforcement sets new policies that are calculated with a security level goal system, in order to improve the system's security posture against newly identified attacks. The run-time phase can be iterative until all the criteria for the required security level are met.

When it comes to security policies enforced for **user behaviour**, security enforcement processes make sure that every user is in compliance with the security standards specific within an organization or system, and is usually a manual process performed by a person or group of people (security analysts). When it comes to this security enforcement, a trustable, unbiased and transparent party must be used to make sure that the whole process is done in a correct way. Therefore, it seems sensible to design a high-resilient, anti-tampering component to perform the enforcement of security and privacy policies, because if this component was compromised, the attacker would be in control of the security policies, enabling the execution of untrusted scenarios. **Blockchain and its distributed nature may provide a very decent solution when it comes to designing a component like this one.**

First of all, its peer-to-peer nature allows the Blockchain network to be highly resilient, as this network can work with just one node or one thousand nodes [18]. In addition, **Smart Contracts** (Section 6.1.2) **are the perfect tool to do the security enforcement process**, as it is a transparent, auditable code, shared between all the nodes, which can be reviewed by anyone but changed by no-one, and controls automatically the logic of the enforcement process without any human intervention. As a consequence, moving to the attacker´s perspective, it is infinitely more complicated for him to alter the security enforcement process if it has been coded into a Smart Contract, because it implies changing more than the half of the nodes where it is running, being this number bigger than in private approaches [19].

In this context, **the DataVaults platform will be able to act as a secure "oracle" (with trusted hardware - TPM) to convert security and privacy-preserving policies into smart contract logic** and further deploy the resulting smart contract to the underlying ledger. The deployment of the smart contract will be supported by the provided DataVaults ledger interfaces (Sections 6.1.4 & 6.4). Hence, the platform can be seen as a secure and trust anchor for security policy "conversion", "deployment" and "execution" via smart contract. It will keep "everything" recorded and traceable for attestation management and maintenance. Parties within the same system and across collaborative systems will be further allowed to search the smart contracts (via contract sequential number) over an off-chain structure (providing high efficiency search ability) to identify the location of the contracts on the (online) ledger. Finally, in order for these policies to be enforced, DataVaults will leverage **hardware-based roots of trust in each user's device** that will be responsible for measuring and attesting to the integrity and correct functionality of each device.

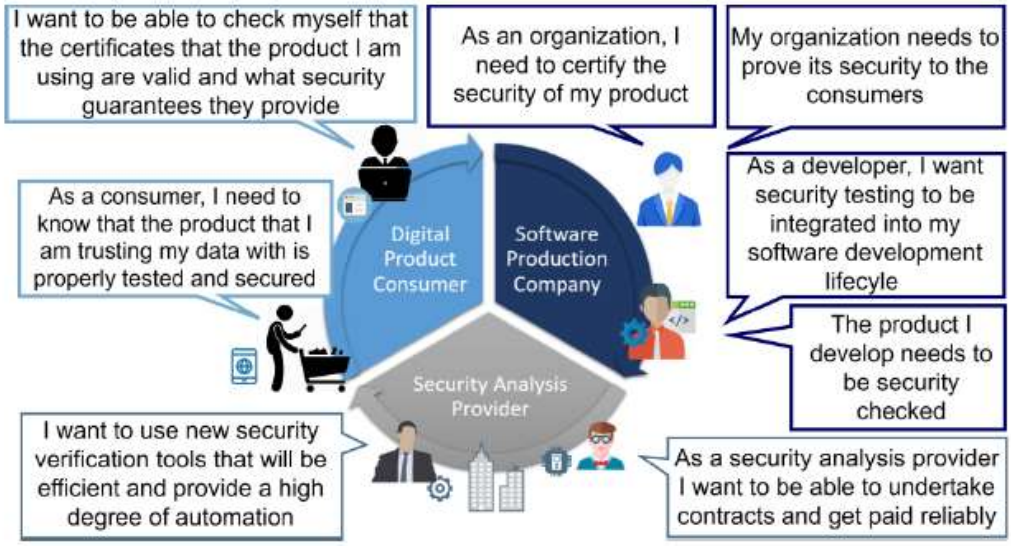

Figure 6 - Security policy enforcement during design- and run-time phases of product development.

Finally, from the user perspective, private Blockchains allow to establish an access control to participate (join) to the Blockchain, so it is quite straightforward to control who has access to these policies stored in the Blockchain and managed by the Smart Contracts. However, in this way, there could be a lack of transparency as not everyone can access to the content stored in the Blockchain. The choice between a public or a private ledger approach will fully depend on the requirements of the envisioned use cases as is also described in Section 6.1.

When it comes to security policies enforced for the **software processes** running in a system/network, there is a large body of related work on information flow security enforcement mechanisms. There are two major approaches to information flow security enforcement: **static techniques and dynamic techniques** [95]. In particular, during the design phase of the policy enforcement, the main approaches to enforce a particular information flow security policy, is called non-interference. Non-interference for programs means that a variation of confidential (high) input does not cause a variation of public (low) outputs. Static analysis techniques have one major drawback: they accept the program only if all its executions ensure non-interference. A common mechanism for ensuring that software behaves securely is to monitor programs at run-time and check that they dynamically adhere to constraints specified by a security policy. Policies are enforced by the run-time composition, configuration, and regulation of security services. The principle of separating security policy and dynamically enforcing security on applications is not new. Several authors have proposed security policy enforcement mechanisms using code modification as a technique for enforcing security policies such as resource limits, access controls, and network information flows. However, these approaches are typically ad hoc and are implemented without a high level abstract framework for code modification [96]. Another approach is by using reflection as a mechanism for implementing code modifications within an abstract framework based on the semantics of the underlying programming language. A recent survey [97], presents novel methods that employ machine learning and artificial intelligence in the pursuit for security vulnerability mitigation.

In the context of DataVaults, such a policy enforcement might also allow **Policy-Based Access Control (PBAC)** which is an access control model based on policy-based security management, which controls the access to resources by defining the rules and policy. There are many tools and frameworks for PBAC. Several PBAC frameworks are based on the IETF Framework for Policy-based Admission Control [98].

DataVaults shall enable the composition of large scale data sharing environments to be controlled via layered and cross domain authorization decisions based upon attestation. Such decisions shall be made at each layer (and across layers) to determine whether subsystems/systems conform to policies based upon the properties to which they can attest.

The Security Policy Enforcement platform will contain a Policy Admission Point, the logical component responsible for creating policies and policy sets and makes them available to the PDP. While policy creation will be managed by the PAP, a Policy information point (PIP) shall act as the source of the attribute values referred to within policies. Attributes can be the properties attested to by a component and the verification result of the attestation. The PIP shall be responsible for requesting and receiving such information from the attestation services. The PDP will be the logical component that shall evaluate the attested to properties/attributes against applicable policies and makes the final authorization decision. While the successful verification of the attestation provides evidence that the information supplied is correct, the PDP decides whether the collated information supplied sufficiently demonstrates conformance to policy.

6 TRUST ENHANCING DLT AND SMART CONTRACTS FOR FAIR AND SECURE PERSONAL DATA SHARING AND MANAGEMENT OF TRANSACTIONS: INITIAL INSIGHTS

The goal of DataVaults is the provision of a secure, trusted and auditable data sharing environment based on the use of **distributed ledger technologies and signed smart contracts** to capture data sharing (while complying with the prevailing GDPR legislation as highlighted in Section 2), collection, and compensation and trading preferences among the DataVaults parties for **guaranteeing the trusted consent management** among users. Advanced crypto primitives for enhanced **security and user-controlled privacy** (Section 5), aiming to put the users in control of their own privacy and that of their generated data, will be coupled with the provision of a set of components for the **secure and efficient computation, management and audit** of all data sharing transactions. This set of services, to be integrated into the envisioned distributed ledger infrastructure, will enhance the framework's overall security and reliability by guaranteeing **ledger management and maintenance**.

To this end, DataVaults will be built on a **hyper-ledger model** (Section 6.1) leveraging two general types of ledger infrastructure, namely a **private ledger** for capturing fine-grained details of data sharing actions and provided metadata and indexes to the stored data (Section 6.1) and a **public ledger** for recording security, privacy and sharing preferences (through the usage of smart contracts – Section 6.1.2) and documenting aggregated metadata towards efficient data search (Section 6.2).

Reflecting on DataVault's work and data flow and how provided data security, privacy, sharing and management services are to be engrained in a **policy-compliant Blockchain structure**, in what follows we will capture an initial blueprint of the distributed ledger technologies and smart contract protocols to be investigated within DataVaults towards offering the aforementioned set of, highly interconnected, functionalities and services:

a) **DataVaults Secure Data Trading** (Section 6.1.2), comprising all services related to the secure management and conveyance of data as well as secure data trading services;
b) **DataVaults Blockchain Operation Services** (Sections 6.1.4 & 6.3), comprising services related to the Blockchain computation, verification and auditing;
c) **DataVaults Trusted Blockchain Services** (Sections 6.4 & 6.5), that form the basis for enhanced security, privacy and reliability guarantees for ledger management and maintenance through the use of trusted computing technologies (i.e., TPMs).

This initial analysis, comprising technology identification and state-of-the-art documentation and investigation, will provide the roadmap for WPs 2, 3, and 4 towards the further design and implementation of such a Blockchain-based secure data-sharing environment.

6.1 LEDGER-BASED DECENTRALIZED DATA MANAGEMENT & ACCESS CONTROL

In DataVaults, we will leverage Distributed Ledger Techniques (DLTs) to maintain integrity, traceability and immutability throughout the entire lifecycle of all involved data; from data

collection and storage to flexible data search and sharing. In this context and to fully capture all behaviours involved in DataVaults, that has to be accommodated by the underlying distributed ledger infrastructure, as has also been defined in D1.1 [42], the envisioned data value chain can be seen as a structure of a trusted cycle which involves:

- **Primary Personal Data Providers (Individuals)**, who are **registered and authenticated users** willing to provide their own data, collected from various services, for the purposes of data outsourcing, storage and trading;
- (1st-tier) **Economic Operators (Data Brokers)**, that accommodate their business intelligence based on such primary personal data. In this tier, data brokers (organizations of any type) are enabled to **collect, manage** (e.g., data processing towards the production of relevant derivatives such as insights, reports, etc.) **and store** provided data while at the same time participate in **data trading activities** with other potential data collectors acting as 2nd-tier economic operators;
- (2nd-tier) **Economic Operators (Data Collectors)**, that are interested in **searching and extracting (primary) data of interest** towards providing services based analytics or data that is shared and generated by the data brokers.

Therefore, the DataVaults envisaged conceptual architecture will investigate (as described previously) a **hybrid-ledger model** leveraging two general types of ledger infrastructure:

- A **private ledger** for recording the actions of data sharing between the data providers and the platform, and
- A **public ledger** for documenting the transactions that shall be performed between external parties (e.g., data brokers/collectors) and the DataVaults platform, where also the smart data usage/sharing contracts will be stored to ensure the existence of a reference point regarding clauses relevant to the proper usage of the assets (including IPs, retention periods, ways of usages, etc.) that can be shared over the platform.

Essentially, the DataVaults platform will provide a **private ledger** (per data provider or a set of authenticated data providers) such that the **definition and recording of data sharing actions, read/write policies, etc.** can be applied while maintaining a **secure and auditable birds-eye-view of data flow access controls**. If a collaborative partnership with external stakeholders (i.e., data brokers/collectors) needs to be established (through the definition and instantiation of smart contracts – Section 6.1.2), DataVaults will provide a **"public shared" – permissioned – ledger** which will be maintained by all registered partners in an authenticated distributed and decentralized network.

Data sharing agreements and transactions will be reflected on the permissioned ledger. By permissioned, we mean that all the information sharing flow can only be accessed by the authenticated entities (via a membership access control layer). This use of the ledger is for ensuring the **data and event traceability**, during a data trading activity, and to be able to provide enhanced data operation security and confidentiality, user privacy and ledger management services. Such an approach will allow to not reveal the identity of individuals to data seekers (i.e., data brokers/collectors), while it places the core DataVaults platform as

a proxy service in the middle of transactions, enabling the set up of a sustainable business model for the platform itself.

6.1.1 DataVaults Distributed Ledger Infrastructure

Blocks will be the basic unit of the DataVaults hyper-ledger, each one of which comprises a block head, a block number and block data fields. We define the first block (*genesis block*) to capture all security- and privacy-related policies: it will embed **data sharing policies, GDPR-based policies and relevant data protection policies into the data field** (note that their hash values will be integrated to achieve policy integrity). The data field of subsequent blocks (*normal blocks*) will include the **information metadata, encrypted pointers, hash values of encrypted pointers, and transaction details, smart contract interfaces** (which are required for embedding smart contracts into coding level data), and other **information interfaces**. Hash values are used to guarantee the integrity of information in the platform. Blocks will be built and put sequentially onto a ledger.

In what follows, the most relevant candidate solutions for both the internal (private) and the external (public) ledger are described.

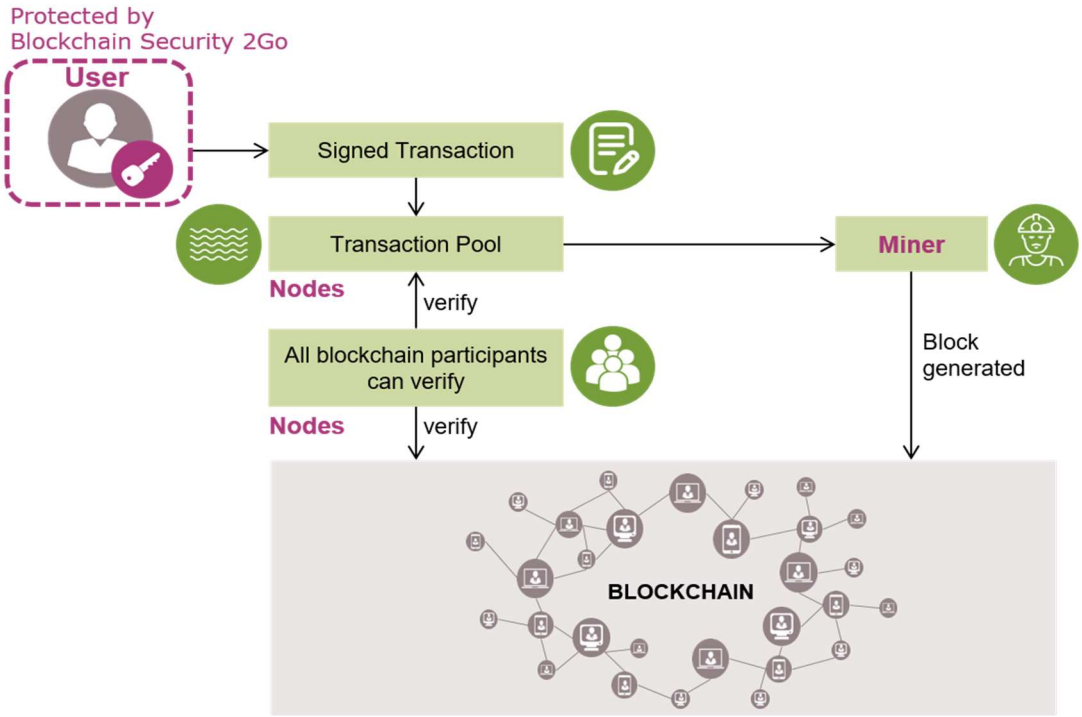

Figure 7 - The Blockchain Security 2Go starter kit offers protection for the user keys. As an alternative also a TPM could be used (if it supports the crypto primitives required by the applied Blockchain).

6.1.1.1 Consideration for the Internal Ledger

Blockchain Security 2Go Starter Kit: For the establishment of the internal (private) ledger, DataVaults will investigate the integration of Infineon's Blockchain Security 2Go Starter Kit which provides a fast and easy way to build best-in-class security into a Blockchain system

design (Figure 6). It is based on the use of NFC capabilities for empowering the user to be in control of his/her data and be aware of any use of the data created. Towards this direction, the starter-kit actively involves the user when an entity requests access to his/her data. For instance, he/she confirms a data request by tapping an NFC-enabled smartcard to his/her smartphone. Then he/she could be directly compensated for providing the data with e.g. cryptocurrency tokens that are sent to an Blockchain account secured by the smartcard. The high level flow of such an NFC triggered data decryption process looks like the following:

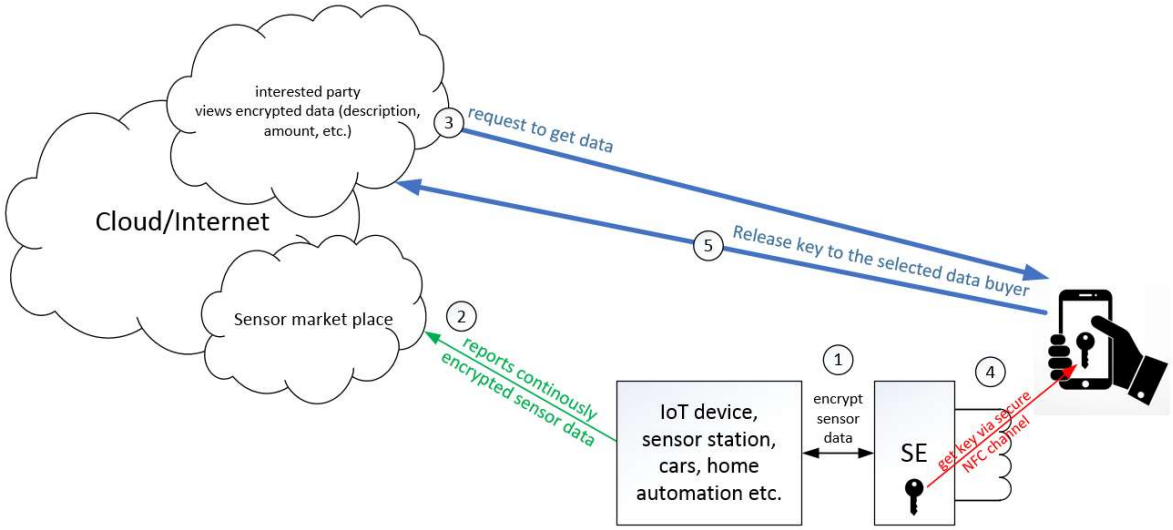

Figure 8 - High level flow of an NFC triggered data decryption process.

1. The data which is generated by the device is used in the particular system, but is also encrypted with a connected secure element with NFC capability. Note that the device can be everything: One example is home automation but also cars collecting information about the status of the road would be possible.
2. The collected, encrypted data is continuously reported to a cloud service. For instance, this could be also a sensor market place.
3. Interested parties get information about the data and can observe the offers. If a party wants to have access to the data, a request is made to the owner of the data provider. Note that this can be also the device itself.
4. There is only one way to get access to the key in order to decrypt the data → via NFC. By reading the particular key via NFC (of course there can be many keys or the key is renewed after reading it), only the device owner can forward the key to the particular request source. Without NFC connectivity there is no possibility to get access to the data.
5. By releasing the key to the selected request, the 3rd party can use it in order to decrypt the data and get access.

Another possible approach is that the data selling process is not secured by a smartcard that belongs to the user, but by a TPM that is directly embedded in the data generating device.

In the scope of the DataVaults project it will be investigated if privacy-friendly encryption methods (pairings, etc.) using secure controllers (e.g. Blockchain Security 2Go starterkit, TPMs) from edge to cloud may enhance the privacy of such an approach. If so, it will require investigation on how to realize such cryptographic primitives on highly resource-limited devices such a security controller.

Furthermore, this solution could be enhanced by a smart contract that manages the data providing and compensation process automatically. How to exactly define such a smart contract (Section 6.1.2) that also leverages the security provided by a security controller is an open research question.

6.1.1.2 Candidate Solutions for the Public Ledger

Ethereum[18]: Ethereum is a global, open source platform for decentralised applications. It was proposed in 2013 by Vitalik Buterin and launched in 2015 and is among the world's leading programmable blockchains. It can be classified as a public Blockchain technology, i.e, it is a public network, open for anyone to contribute and interact anonymously. It is not controlled by a centralised organisation or company. On the contrary, a community of diverse contributors around the world, work constantly on maintaining and improving it. The Ethereum platform generates its own cryptocurrency, Ether, as a reward to mining nodes for computations they perform, which is the only currency accepted in the payment of transaction fees and is required for the execution of smart contracts. Ethereum is programmable, meaning that developers can use it and build new applications exploiting the power of cryptocurrencies and Blockchain. These applications are trustworthy (always run as expected) and decentralised (no single entity controls them). Among others, these applications, which run on Ethereum as decentralised apps, include: wallets that enable payments using Ether, decentralised markets that allow trading of digital assets, financial applications that enable borrowing/lending and investments of digital assets, and more. The Blockchain community behind Ethereum is one the largest and most active in the world with members varying from ordinary users and app developers to mining organisations, protocol developers and many more.

Hyperledger Fabric[19]: Hyperledger Fabric (Fabric) is an open source, enterprise-grade distributed ledger platform, available for the development of solutions and applications in enterprise contexts. Fabric is a project originally contributed by IBM and Digital Asset. It is part of the Hyperledger Greenhouse, which was established under the Linux Foundation[20] . Fabric is governed by a group of maintainers from multiple organisations. Fabric is among the best performing platforms in transaction and smart contracts implementation due to a number of features. The modular and configurable architecture behind Fabric enables versatility and optimisation in a various range of applications, including finance, healthcare, supply chain, and more. It also supports the authoring of smart contracts in general-purpose programming languages (e.g. Java, Go, Node.js), rather than domain-specific languages, that require additional training from the developers[21]. The consensus protocols of Fabric do not need a native cryptocurrency for their execution. The Fabric community comprises nearly 200 developers, leveraging the features of Fabric for innovative use cases entailing the identifiability of transaction participants, permissioned networks and other requirements.

[18] https://ethereum.org/
[19] https://www.hyperledger.org/projects/fabric
[20] https://www.linuxfoundation.org/
[21] https://hyperledger-fabric.readthedocs.io/en/release-2.0/whatis.html

Cardano[22]: Cardano is an open source, decentralised public blockchain and cryptocurrency project, aiming to develop an advanced smart contract platform. It is led by three organisations: IOHK[23], Emurgo[24] and the Cardano Foundation[25].The development team behind Cardano comprises expert engineers and researchers, following a scientific-driven approach. Cardano has its own cryptocurrency, Ada, for the execution of digital funds transactions. Cardano is a platform capable of running decentralised financial-oriented apps, to be used by individuals, organisations and governments at a global scale. It has been implemented with Haskell, following a functional programming approach. The Cardano protocol has been built on peer reviewed academic research, to construct a robust system integrating distributed systems, cryptography and mechanism design. The multi-layered structure of this protocol makes it flexible and easily maintained and upgradeable. It has a large and active user community. The integration of new features, such as smart contracts, and the optimisation and scalability of the provided solutions are foreseen until the end of 2020, when Cardano will be a self-sustaining system[26].

Hyperledger Sawtooth[27]: Hyperledger Sawtooth (Sawtooth) is an open source enterprise solution for the implementation of distributed ledgers, with potential applications in fields varying from IoT to financial. Originally proposed by Intel and released in 2018, Sawtooth is part of the Hyperledger Greenhouse, under the Linux Foundation. The separation of the core system from the application domain, enabling the writing of smart contracts and the specification of business rules, without the need to know the details of the underlying core system. Additionally, it provides a highly modular architecture, that allows applications to support their unique business needs and policy decisions by choosing their consensus, permissioning and transaction policies. Sawtooth supports both permissioned and permissionless infrastructure. It interoperates with Ethereum via the Seth integration project, where smart Ethereum contracts are deployed to Sawtooth. Other features of Sawtooth include: dynamic consensus, parallel transaction execution, an event system and more. It supports multiple popular languages for the creation of contracts, including Python, Go, Java, C++ and more.

Corda[28]: Corda is an open source, enterprise distributed ledger platform for the writing and execution of applications, the CorDapps. Corda was developed by R3, an enterprise blockchain software firm, back in 2015. Today, Corda is maintained by a vibrant community of developers continuously adding new features, functionality and more. Unlike traditional blockchain, Carda could be better described as a shared ledger platform, meaning that transaction information is only shared with parties involved in the specific transaction in order to achieve consensus, rather than being broadcasted to all nodes in the ledger. This semi-private network design can address the scalability and privacy issues raised in original blockchains. The coding language of choice for Corda is Java-based Kotlin, making it

[22] https://www.cardano.org/en/home/
[23] https://iohk.io/en/about
[24] https://emurgo.io/en
[25] https://cardanofoundation.org/en/
[26] https://cardanowiki.info/wiki/Home
[27] https://www.hyperledger.org/projects/sawtooth
[28] https://www.r3.com/corda-platform/

compatible with other popular languages among fintech (such as Scala and Groovy). This makes it highly interoperable with existing systems. Initially oriented at financial applications, CorDapps could also be developed for other fields such as trade, healthcare and more, leading to one of the largest blockchain ecosystems in the world, comprising hundreds of industry participants.

Comparison & Evaluation: Following is a comparative presentation of the five distributed ledger technologies. Seven characteristics have been recognised as the main differentiating points among DLTs. These are: **the short description of the platform** (i.e. modular or generic, and being a DLT or blockchain implementation), **the industry focus** (i.e. the main field of the supported decentralised applications), **Governance** (i.e. who is in charge of the platform), **mode of operation** (i.e. permissionless or permissioned, public or private), **Consensus** (i.e. how is consensus achieved), **Smart Contract Language** (i.e., the language used to write the smart contracts), **Linked Cryptocurrency** (i.e. the cryptocurrency -if any- used by the platform for its operations)

Characteristic	Ethereum	Hyperledger Fabric	Cardano	Hyperledger Sawtooth	R3 Corda
Description of platform	Generic Blockchain platform	Modular DLT platform	Blockchain platform	Modular DLT platform	DLT platform
Industry Focus	Cross-Industry	Cross-Industry	Finance	Cross-Industry	Finance
Governance	Ethereum developers	Linux Foundation	Cardano Foundation, IOHK, Emurgo	Linux Foundation	R3 Consortium
Mode of operation	Permissionless, Public	Permissioned, Private	Permissionless, Public	Both Permissionless and Permissioned, Private	Permissioned, Private
Consensus	Proof of Work	Pluggable framework	Proof of Stake	Proof of Elapsed Time	Pluggable framework
Smart Contract Language	Domain Specific Language: Solidity	(Chaincodes) General Use Languages: Go, Javascript, Java)	N/A yet	General Use Languages:. C++, Go, Java, JavaScript, Python, Rust and Solidity via Seth)	General Use Languages: Kotlin
Linked Cryptocurrency	Native Ether Cryptocurrency	N/A	Native Ada Cryptocurrency	N/A	N/A

Table 4. Comparison of five distributed ledger technologies.

6.1.2 Smart Contracts

A smart contract (SC) [7, 43] is an enhanced digital version of a normal contract. Instead of stating consequences that are attached to certain actions or events, those consequences can be directly implemented into program logic, so e.g., if a data collector does not pay in time, a smart contract automatically prevents from accessing the data. Smart contracts are independently and autonomously executed by the Blockchain and as an actual part of the Blockchain are **immutable and transparently stored on the ledger**. In this way the execution of a contractually specified action cannot be prevented or manipulated.

Specifically, smart contracts are a set of programming codes that digitally **facilitate, verify and enforce** the contents of a conventional contract. Smart contracts are deterministically executed and are able to access the data stored on the ledger [20]. Thereby the repeated execution of a smart contract given an equal ledger state will always lead to equal computation results. The exact definition, extracted from [21], can be:

"A smart contract is a computerized transaction protocol that executes the terms of a contract. The general objectives are to satisfy common contractual conditions (such as payment terms, liens, confidentiality, and even enforcement), minimize exceptions both malicious and accidental, and minimize the need for trusted intermediaries. Related economic goals include lowering fraud loss, arbitrations and enforcement costs, and other transaction costs."

They are based on traditional legal contracts, were a non-interested third party does a transaction of a common good between two interested parties. Its objective is to enforce the agreement between the interested parties, in other words, to guarantee the transactions. The Smart Contracts are an extension of this concept, where all the interested parties share a common logic with a common verification mechanism, protected with cryptographic protocols [22].

The most famous cryptocurrencies [23] have a Smart Contract who defines the mining behaviour, transaction feeds, withdrawal limit... in this cryptocurrency. These are some of the uses related with the coin, but an Smart Contract is broader than that, including scenarios such as: financial contracts [24], gambling [25]... within the monetary platform. Almost everything can be coded in a Smart Contract in the same way it can be done in a normal software. However, some risks must be taken into consideration when using Blockchain-based systems and also some advantages as stated in [26].

In the context of DataVaults, **smart contracts are used to enable standardized and long-term comprehensible interactions between the participating parties** (data provider, broker and collector). Moreover, smart contracts enhance the flexibility and scalability of contract management and execution. Two parties (of a contract) are required to follow the DataVaults methodology to form and sign a digital contract in order to allow **comprehensible data sharing**. Note that cryptographic digital signatures replace user handwritten signature on the contract (Section 6.1.3). Such contracts are used to maintain **reliable system execution**, **guarantee trusted user consent** and **guide user's behaviour** in order to minimize potential (malicious) misbehaviours. Furthermore, smart contracts are used to capture user preferences/policies towards supporting flexible and customized levels

of data sharing and collection. DataVaults leverages smart contracts as event trigger to accelerate event reaction and shorten delay over real-time events (e.g., a smart contract triggers an event to proceed the compensation for the data provider).

More details on the modeling of the use of smart contracts, in the context of DataVaults, for fair and secure personal data management will be provided in D2.2 – "*Personal Data Market Design, Contracts and Rules*", however, envisioned types of contracts that may be considered (as an optional functionality):

Smart Contract between Data Provider and Broker. This SC provides the necessary guarantees that the data provider will follow the data sharing and collection policies of the DataVaults platform during its data preparation while protecting its rights from being violated from the broker. Data provider is required to sign a smart contract with the broker at the beginning of registration. User information, personal data sharing, payment details, data deletion requirements, trusted consent (which is in line with broker's data policy and related GDPR-based policies) and security deposit (which is the behaviour guarantee fee of both entities used as a deterrent to misbehaviours) will be merged with payment details (in case of data trading) into the smart contract.

Smart Contract between Data Collector and Broker. After data trading between the data collector and broker reaches an agreement, this needs to be translated into a smart contract. **This SC comprises of user information, data collection preference, payment details, trusted consent and security deposit.** To enhance the security level of collected data (after data collection), the trusted consent information allows the data broker and collector to negotiate a data sharing revoked time (before the commencement of the service), so that data access will be rejected after a future time slot. The security deposit in this contract is also bilateral - preventing collector from rejecting payment but also avoiding broker from denial of service after receiving payment – in order to guarantee the benefits of both entities in data trading.

Figure 9 - DataVaults Entities Data Trading.

6.1.2.1 Secure Data Trading

As an advanced data sharing and privacy management framework, DataVaults platform and Blockchain-backed protocols allow various forms of corporate and user data to be **monetized and exchanged between different parties**. Users can earn **rewards for sharing data** on a secure, decentralized network; rewards can either be monetization tokens, crypto currencies or other traditional financial services.

These **trading services** are enabled through the use of **smart contracts and trusted blockchain wallets** (Section 6.5) for securely storing and accessing all necessary user credentials for such trading deals. DataVaults leverages TPMs (Section 5.2.1.1) for delivering

trust and payment services. To eliminate impersonation and minimize transaction fraud, a TPM will be embedded into a user's account wallet for enabling payer/payee authentication. A user will receive payment if a data trading is successfully concealed and a payment event is triggered (note this is ensured through smart contracts). Payment transactions will be further recorded onto the ledger for validation (Figure 7).

6.1.3 Advanced Security, Privacy and Trust Layers

DataVaults will make use of **advanced encryption techniques** to protect user's data from being compromised and tampered with. The integration of encryption technologies will also guarantee data access rights to only authenticated and authorized system users. As described in Section 5, **data security** includes the integrity and confidentiality of data. **User privacy** will be partially adhered to data security as potential security breaches of data can severely harm user privacy. On top of that, DataVaults will also consider user privacy through authentication mechanisms, privacy-preserving signatures (DAA – Section 5.2.4) and the use of smart contracts to ensure user ledger access rights, data copyright, and contract rights. **Ledger security** mainly revolves around the correct control and operation of the blockchain structure.

The DataVaults platform will offer advanced protocol interfaces towards:

- **Integrity and verification of block data** for guaranteeing that stored data has not been tampered with;
- **Mining validation** for ensuring that a block mined by a user is valid;
- **Consensus agreement** for allowing a majority or all network users to reach an agreement on block or ledger validation;
- **Membership authentication** for providing access control mechanisms (read & write privileges) to authenticated users of the ledgers;
- **Undeniable actions commitment** for guaranteeing indisputable user operations over the ledgers, and
- **Customized block data security** for enabling users to put various levels of encrypted metadata onto the ledgers.

Towards guaranteeing the aforementioned properties, there is a plethora of security, privacy and operational assurance algorithms and techniques that DataVaults can investigate as core building blocks in the context of secure data sharing (see also Section 6.1.3.1):

- **Target Collision-Resistance Cryptographic Hash Functions** [44];
- **Merkle trees** [45] where data pieces are grouped in pairs and the hash of each of these pieces is stored in the parent node. In context, a data piece is captured as one transaction record and Merkle trees are used for efficient data storage and scalability;
- **Proxy Re-encryption** [46, 47]**, Searchable Encryption** [48, 49, 50] **and Hierarchical Identity-based Encryption** [51] have been proposed towards enhanced security in data storage, sharing and searchability. Such models if properly designed and implemented, can enable data querying even when the data is encrypted but in a resource-efficient manner (something that has been identified as a main limitation in existing Blokchain structures);

- **Digital Signatures** [52] **with various levels of anonymity** (e.g., linkable group signatures) can also be considered for achieving public verifiability and unforgeability;
- **Distributed Consensus Protocols** [53] for allowing the execution of user actions even if some of the users are faulty or malicious: a consensus mechanism is the way in which a majority, (or, in some mechanisms, all) of network users agree on the value of a piece of information.

On top of that, DataVaults will investigate the provision of advanced blockchain control services through the specification of novel **TPM-based security and privacy-preserving** protocols (through the leveraged Blockchain Security 2Go Kit) for advancing the state-of-the-art in scalability and computational efficiency for securing different levels of user privacy. As described in Section 5.2.1.1, a TPM [54] is a general-purpose and tamper-resistant security component. It is designed to be used as a root of trust as well as a cryptographic engine for the system that it is embedded into. The specifications of a TPM, developed by the Trusted Computing Group (TCG), are adopted as an international standard (ISO/IEC 11889). The core TPM services that will be investigated within DataVaults, are **platform authentication, system attestation and integrity, and secure storage** (Section 5.2.4)**.** From the nature of tamper-resistance, a TPM provides a platform cryptographic identifier, which allows the platform to be authenticated by a remote communication partner; moreover, a TPM can measure and report the platform configuration and software status by using the technique of Platform Configuration Register (PCR), which is the basis of platform attestation and integrity. From a powerful key hierarchy technique, a user can store an arbitrary amount and multiple types of secret values with a TPM. These three core services can provide enhanced solutions for DataVaults **user authentication and Blockchain attestation in a flexible and efficient manner**.

6.1.3.1 DataVaults Secure Components – State of the Art

Current Ledger Techniques: The first underpinning ledger component is cryptographic digital signature [52], protecting the ownership of transaction in decentralized network. DataVaults will use lightweight digital signature scheme e.g., Schnorr signature [55]. Consensus algorithm is another crucial component that guarantees the validity of block mining and the correctness of block validation. Practical byzantine fault tolerance [56], has been considered an efficient algorithm for DLT applications, but with a limitation that consequences of mining misbehaviour are ignored so that there is no guarantee for user to behave. Proof of "Something" is invented in such a way that miner is requested to give a proof of mining - a commitment to honest behavior. As a typical case of proof of something, proof of work (PoW) [57], the underlying consensus algorithm of Bitcoin, requires miner to invest computational resources on mining blocks. That is inefficient and not cost-effective due to resource consumed feature. To get rid of the heavy computational cost, proof of stake (PoS) [58] is designed to relate mining to stakeholder's fortune proportion. Compared to PoW, PoS is more efficient and environment friendly (without consuming huge amount of computer resources), but bias is easily taken into PoS in the sense that stakeholders with more proportion of system fortune will become richer (from block mining rewards) and, only stakeholders are qualified to be miners. Other proof of something, e.g., proof of space [59], proof of burn [60], have been designed in the literature as well. *But none of them can fully*

achieve efficiency, cost-effectiveness, fairness, and reliability simultaneously in permissioned ledger context and meanwhile, trusted hardware has not ever been employed to ledger verification. Proof of elapsed time [62] so far is the consensus algorithm supporting TEE and SGX [61]. *However, its efficiency and scalability may be a concern while merging with encryption and smart contract techniques.* Some current DLT platforms, e.g. Ethereum, leverage smart contracts to support efficient and customized contract execution (without a trusted authority) in particular for the execution of payment. *But smart contract should be further developed to record and trigger system events and support other trust-related functionalities (e.g., attestation) to provide trust, transparency and traceability for application users*.

***Lightweight Data Encryption*:** Hybrid encryption is one of the most prevalent encryption modes to protect data confidentiality without infringing data access efficiency. It requires a combination of symmetric encryption (e.g., AES, RC5/RC6) and asymmetric encryption. The former provides data encryption/decryption efficiency while the latter defines fine-grained level of data access control. Attribute-based encryption (ABE) [63, 64, 65], as a general extension of Public Key Encryption (PKE), is a classic type of advanced encryption, allowing sticky policies in data access control. It encrypts data under a description, so that only user(s) with the secret key matching the description can reveal the data from the encryption, in which a description could be a set of attributes or data access policies. ABE guarantees the confidentiality of data but also provide data owner policy-based data access control so that the owner can decide who can access its data via specified sticky policies. Proxy re-encryption (PRE) [46, 47] has been introduced for secure data sharing. Consider an encrypted data is uploaded to cloud server, PRE allows a semi-trusted proxy (i.e. the server) to convert the encrypted data intended for a data owner (called delegator) to another encryption of the same data intended for another user (called delegatee) by using a re-encryption key generated by delegator. In this encryption conversion, the proxy learns no information about the secret key of delegator as well as the underlying data, such that the privacy of the delegator and the data confidentiality are both secured. PRE has been explored into various contexts of encryption, e.g., attribute-based PRE [66] and functional-based PRE [67]. To securely search encrypted data (which is remotely stored in cloud server), searchable encryption (SE) has been introduced in such a way that a server (managing an encrypted database) can return the correct search results (which are encrypted files) to data searcher without knowing "what the search query is" and "what are the underlying plaintexts of the encrypted files". In general, SE is categorized into two classic types – one is public key based SE [68], which cares more on integrity check of outsourced data and strong security (but with less efficiency – due to heavy computational cost in matching); the other is symmetric SE [69], guaranteeing high efficiency in data search (but with less security – due to deterministic feature in search token). *It is still unknown that how one may integrate all the above advanced encryption technologies seamlessly as a whole system.*

6.1.4 Blockchain Computation & Verification Functionalities

6.1.4.1 Secure Computation

As described previously, in a Blockchain-based economy, there are various participating entities such as miner, nodes and the end users (**Error! Reference source not found.**).

Basically, a Blockchain is a decentralized digital ledger that manages a continuously growing list of data points (**chain of blocks**). Every block in the chain is cryptographically linked to the previous block. Consequently, to change one block and retain validity, an attacker would have to change the entire chain. The ledger records all transactions that have been sent to or from different accounts. This transaction history allows users to determine the current asset value that belongs to an account. Usually a transaction includes information such as the receiver's public key, the amount of assets that should be transferred or arbitrary data for a smart contract [27].

Towards providing certified information and actions over such digital ledgers, DataVaults provides an advanced set of Blockchain operation services where: (i) **all transactions are confirmed by the network** as entries forming blocks of transactions, and (ii) **the whole network monitors the legitimacy of each transaction**, guaranteeing a distributed control system. More specifically:

- ***Block Data Field Hash*:** To ensure the integrity of a block data field, DataVaults will use **cryptographic hash functions to hash the current block's data field and further record the value into the next block's header**. This will allow for the integrity of block data to be always guaranteed by checking the follow-up block header. Furthermore, DataVaults will provide a **fine-grained integrity check over all data stored in a block**. All data should have corresponding hash values; for instance, hash values of metadata, encrypted pointers and transactions. All of the hash values will be regarded as leaves of a Merkle Tree, in which a pair of hash values (out of all values) is bonded by a parent node. After building the Merkle tree, only a Merkle tree's root value will be stored in the block's data field. This significantly reduces the **storage and cost requirements.**
- ***Signature on User Behaviour*:** To prevent **malicious users from denying the actions they have performed**, DataVaults will make use of **cryptographic digital signature algorithms** in the sense that a user must digitally sign his/her operations over the ledger (note these operations must yield some type of change on the ledger status, e.g., transaction update, block mining, block verification, data auditing). Users will be equipped with a pair of keys **(signing and verifying keys)** to fulfil a valid signature but also to publicly verify the signature (Section 6.3). They will be allowed to generate their own signing keys while verifying keys must be embedded into the membership credential issued by a trusted authority (who takes charge of asserting membership credentials). This guarantees non-forgeability: verifying keys are not maliciously generated by attackers for the purpose of forgery but a real and valid key (matching with signing key) was generated and verified by the trusted authority.

6.1.4.2 Verification & Distributed Consensus

There are different approaches of how the "membership" for such Blockchain-based systems is handled. Bitcoin, the most used Blockchain and cryptocurrency today is a famous example of a "permissionless" Blockchain. Since it is "permissionless" and also "trustless", anyone (i.e., end-users and mining nodes) are able to join or leave the network at any point in time. With this approach, the "membership" of a permissionless Blockchain is dynamic. However, this leads to a number of challenges:

- *Unknown participation leading to concentration of power*: In a proof-of-work Blockchain, some entities could use a massive concentration of CPU power to manipulate the network.
- *Trustless model achieves limited trust*: As there is an only limited trust, service agreements or contracts cannot be established for a Blockchain system.
- *The value of transactions is limited by a lack of business trust*: The lack of service agreement or contract means that business cannot count on the availability or service of a given Blockchain system.

Thus, taking the above into consideration, there are three general approaches that DataVaults will investigate towards the verification of node membership:

- **Pre-identified participation**: Before participating, nodes most be identified and authenticated. Examples are implementations of Hyperledger Fabric [28]
- **Anonymous participation**: Nodes are not identified or authenticated. Thus, any entity with computing resources can operate as a node and can come and go whenever he likes. Bitcoin applies this approach.
- **Anonymous-verifiable participation**: Here, a node is able to cryptographically prove it is a member of a Blockchain without revealing its full identity (for example [27]).

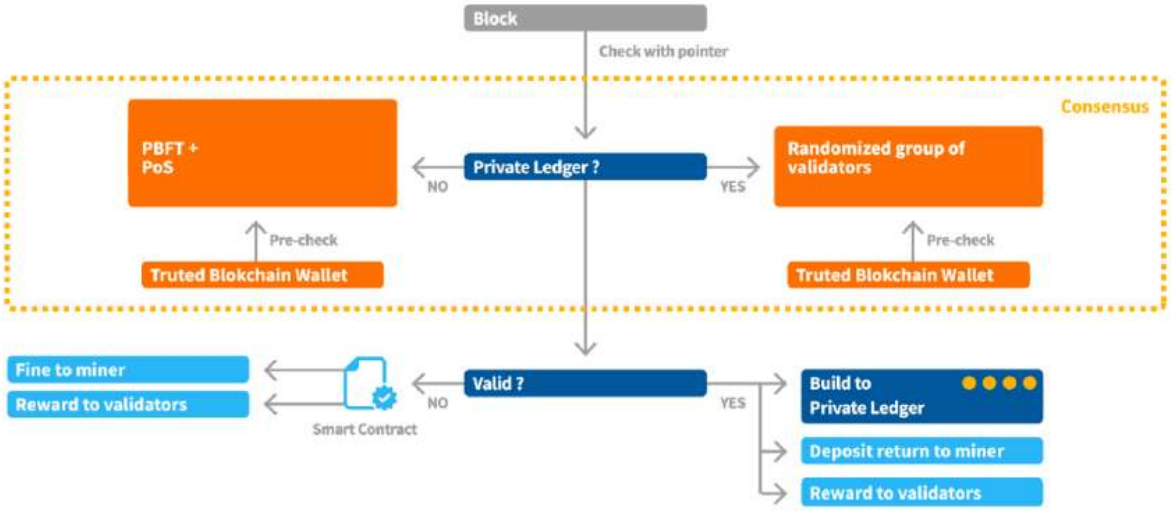

Figure 10 - Distributed Consensus on Mining.

In the context of DataVaults, we will investigate how security controllers (like the Blockchain Security 2Go starterkit or TPMs) can support **pre-identified or anonymous-verifiable participation**. For example, only platforms that are preregistered and show their authenticity, using the build-in TPM, will be allowed to participate in the network. Furthermore, it will be investigated how the **DAA feature of TPMs could allow anonymous-verifiable participation**. However, this poses various challenges for a practical realization, such as the exact definition of the trust protocol or the initial setup of verified nodes and participants. One possible technical solution to tackle this challenge of initializing trusted partners is to provide every trusted party a physical device, such as a smart card (i.e. the Blockchain Security 2 Go starterkit) or a USB stick that holds the credentials. Then, a certain

device can only participate in the network when the USB stick is plugged into the device. Such a solution could be realized by extending the interface of the currently exiting Blockchain Security 2Go starterkit with USB capabilities.

Furthermore, based on the current Blockchain intrinsic features, DataVaults will enables users to **mine the blocks to be captured on the ledgers**. Aiming at designing a new generation of Blockchain-based data sharing system and to guarantee the **validity of each mined block**, DataVaults will **output a novel design and implementation of consensus algorithms** (Figure 8): it will combine and extend PBFT and Proof-of-Stake (Section 6.1.3.1) to yield the following benefits: (i) **reduce response time** of voting/validation (i.e. shortening the waiting time of merging a block into ledger); (ii) **eliminate bias in choosing block miner** and, randomness is added into the election of miner (so that malicious users have the least chance to predict who is the next miner); (iii) **remove resources/computational power consumption limitations**. The novel twist of this merging will capture **efficiency**, **reliability** and **cost-effectiveness**. That will offer a potential solution to industry-level consensus applications. Furthermore, DataVaults will inspire a brand new **ledger verification direction** – based on the use of TPMs - to enhance the computational efficiency of the verification process.

6.2 Data-Driven Protection Engine – Integrity, Confidentiality & Advanced Data Sharing Tools

In DataVaults, data and its derivatives will also be protected by **encryption schemes**, while access to data will be performed by an **access policy engine** that should be driven by the information that will be part of the smart contracts that will be constructed upon sharing the data. The following sub-section describe some relevant aspects of these operations.

6.2.1 Decentralized and Scalable Data Storage, Search & Further Sharing

As described in the preliminary architecture, primary data in DataVaults will be stored either as **original data**, as **encrypted data**, or using a **hybrid approach** which will allow for more secure and trusted data seeking operations, while at the same time will enable the platform to perform certain analytics tasks using the original data to maximise performance, analytics accuracy and create snippets that can be shared with selected stakeholders at will. In what follows, we will give some insights in to how data can be stored in this hybrid state, and leveraging ledger-based storage, as well as information on how secure data searching can be provided in the DataVaults Core platform (more information will then be provided in the context of WP4).

6.2.1.1 Data Storage Approaches

Hybrid Data Storage: To provide efficiency and scalability for multi-layer trust chain data storage and access, DataVaults will investigate the usage of a hybrid model – **cloud-based storage system and ledger storage** – two-level storage infrastructure. In such a setup, symmetric keys are shared among users/devices, such that the transmitted data can be encrypted and sent to the platform. For further data sharing to other parties, DataVaults can use ABE (Section 6.1.3) to encrypt the symmetric key under some access policy which allows specified parties to gain access to the key for the data reading. Original copy of encrypted

data (including symmetric encrypted data and ABE encrypted symmetric key) will be stored in a cloud-based storage system, so that the system will return a pointer that indicates where the tuple is stored in the cloud. After receiving the pointer, the platform will encrypt it under the access policy, and will further write the encrypted pointer to the ledger. For the communication data flow among other levels within the network, a data sender will choose a one-time symmetric key, to maximize data security, use the key to encrypt a data flow, and further encrypt (using ABE) the key under an access policy. The data sender will then upload the encryptions to the cloud and merge the pointer to the consortium ledger. In such a hybrid data storage approach, the original full copy of data is stored on cloud server while its "mirror" (encrypted pointer) is on ledger. This design can maximize the **efficiency, flexibility and transparency of data traceability, immutability and sharing in decentralized networks**.

Ledger-Based Storage (LBS): LBS is one of the most recent cloud-based applications, enabling decentralised and trustworthy data storage services. Beyond the features provided by the conventional cloud-based storage systems, LBS injects traceability over data operation in its life circle due to a fact that all operations can be recorded in a distributed ledger with tamper immutability. To guarantee search efficiency, LBSs combine distributed ledger with a data-centric storage backend, in which the ledger is used to record data operation, storage location and data audit, while the backend is for centric data storage. Two LBSs, MyHealthMyData [70] and BigchainDB [71], have been proposed to date. *However, they cannot support customized fine-grained level of encryption for data sharing and search but also have not considered the use of hardware root of trust to strengthen the trust level.* Another LBS, proposed by FAREDGE [72], combines cloud system and Hyperledger Fabric [73] to maintain decentralized data storage across inter manufacturing structure, in which the ledger is deployed in between cloud tier and edger tire for data configuration, publishing and synchronization. In the ledger, FAREDGE provides membership and policy-based access control, end-point data encryption, the use of smart contract for transaction and majority consensus algorithm for ledger verification.

Figure 11 - DataVaults Secure Data Search & Collection.

6.2.1.2 Secure Data Search

Searching over secure data (e.g. encrypted data) in DataVaults will be performed by invoking **Searchable Encryption** which is a novel searching scheme that allows users to search over encrypted datasets, and get responses only on the queries they have performed (more information will then be given in the context of WP4). In general, as practise has shown, many security issues arise when a party wants to store data on untrusted entities (e.g. remote file servers and shared cloud infrastructures). Storing data in an encrypted format can be an effective workaround in terms of privacy. However, encryption imposes limitations

regarding actions that can be performed on these data. The problem of sacrificing functionality for security and how to overcome it, has been broadly investigated.

One of the most desired functionalities is the ability to search an encrypted database for particular entries using keywords. Various approaches have been proposed, that include performing functions on oblivious RAMs [74], and multi-party computations [75]. Performance issues due to the complexity of the former [76] and the overhead imposed by the need for multiple servers of the latter [77], have led to the emergence of a third approach: designing search-enabling cryptographic schemes. A popular scheme, which will be investigated in DataVaults, is the Searchable Symmetric Encryption (SSE) [78] scheme: **a symmetric cryptographic scheme designed to provide the functionality of performing direct searches over encrypted datasets.** SSE has been subject of active research for several years now and various security definitions and constructions have derived from this work. There is constant effort to optimise the solution in terms of performance and security, and construct a robust cryptographic scheme, however though at the moment many limitations exist which are being researched in order to improve the efficiency of this scheme.

In DataVaults, SSE will be investigated as the main candidate for implementing a secure search method that is able to search over encrypted datasets of individuals, which will return specific data snippets, in order for data seekers to identify if the data they seek is available on the platform and then engage with data owners to establish data contracts for consuming these datasets.

An off-chain searchable encrypted index structure will be generated and managed in DataVaults (Figure 9). The system will convert the most frequent metadata/search keyword/mode into the “secret information” which can be embedded into a 0/1 binary tree structure, and store the location of the pointer on the private ledger on the leaves of the structure. By this secret twist, the index structure builds up a strong link with the private ledger in such a way that the data searcher may only need to execute a privacy-preserving search over the structure to locate the leaf and then it will obtain the location information of the encrypted pointer on the ledger. To do so, the searcher must be given a search token by the platform, which can be seen as an approval of the permission of searchability, helping the searcher to find a correct path from the root to a specified leaf on the tree structure. The search token, however, will not allow the data searcher to know anything except a location on the private ledger. To synchronize the real-time ledger expansion, we will design a new type of dynamic SE mechanism that allows the system to build up encrypted index structure growing with the ledger. By using SE, DataVaults will aim not only to provide privacy but also high efficiency (O(logN)) in search over massive amount of ICT data flow. DataVaults will also offer fine-grained and expressive search services in the forms of single keyword, multiple keywords, formula, range search, and even regular language.

6.3 LEDGER-BASED SECURE DATA ACCESS CONTROL – KEY MANAGEMENT

The distributed nature of such Blockchain-based technologies may be inherently secure, but there are challenges when it comes to securely interacting with the backend platform (e.g., DataVaults). For example, generating transactions is an extremely sensitive process, creating vulnerabilities in spaces where it interacts with humans or, in the case of IoT, with devices.

Thus, **all transactions that are sent to a public ledger need to be protected by a digital signature**. This makes it extremely difficult to change or alter them without being detected. As described previously, to create such a digital signature, a secret private key that corresponds to the public key (address) of an account is needed. Other participants (i.e. nodes) on the Blockchain use the public key of the sender to verify that the transaction is authentic before adding the transaction to a new block in the chain. Typically, there is no third party and no possibility to alter the history of a Blockchain, there is no way to revoke such a transaction. Therefore, **keys (i.e. Blockchain credentials) require strong protection level in terms of security**. A successful Blockchain system needs highly reliable methods of interfacing with the strong key protection.

- **Level 1**: Storing the Blockchain user credentials on a personal device, such as a desktop, laptop or a mobile phone. While this practice may be convenient, it exposes the user to widely used software attacks.
- **Level 2**: A slightly better security level is achieved by applying a TEE (Trusted Execution Environment) on the device microcontroller, which allows the separation of security software from other, less secure software stacks and therefore provides higher protection against attacks.
- **Level 3**: The highest possible security level protects the Blockchain from micro-architectural as well as physical attacks. This level of security can only be achieved when a dedicated security microcontroller is in place for the operations and credential storage.

A security controller has dedicated countermeasures integrated that protect against these attacks and keep the credentials secret. Examples of security controllers that could be used is the Infineon Blockchain Security 2Go Starter Kit or a TPM.

To link the Blockchain Security 2Go smart cards to a Blockchain, you need an interface device that handles the communication with the Blockchain (see Figure 10). This could either be

- an NFC-enabled smartphone, or
- a host device (e.g. PC, RaspberryPi) connected to a contactless reader (e.g. via a PC/SC interface).

Figure 12 - An interface device that communicates via NFC to the Blockchain Security 2Go card and via a netfork (e.g. internet) to a Blockchain creates the link between the cards and the Blockchain network.

The card form factor with contactless interface is a good way to link the Blockchain credentials (i.e. private/public keypair) to a user. The card storing the keys can be carried by a user and does not depend on a specific device that is connected to the internet. Thus, it can be used with different devices such as the user's smartphone or PC, or even a 3rd party point-of-sales terminal.

A possibility to link the Blockchain credentials to a specific device and not to a user, would be to directly embed a security chip with the features similar to the Blockchain Security 2Go Starter Kit in a device. For example, a smartwatch or an IoT device itself could then store the credentials and is linked to one Blockchain account. This could be realized by adding a contact-based interface to the starter-kit (e.g. I2C). As such an interface is not available yet, it will be evaluated within the scope of the DataVaults project if it can be beneficial for the DataVaults approach. If so, the starter-kit will be extended accordingly.

To ensure authenticity of a transaction that is recorded in a public ledger digital signatures are used. Before the transaction is signed, the transaction is hashed (e.g. SHA-256 [29]). Then, the signature of this hashed data is calculated on the card with the senders' private key To calculate a signature with a Blockchain Security 2Go card or a TPM, it has to be hashed off-card. The Blockchain Security 2Go starter kit supports all hashes that lead to 32 bytes output data.

The vast majority of currently existing Blockchains use Elliptic-Curve Cryptography (ECC) as an asymmetric cryptography method to create signatures for transactions. Most of them use the elliptic curve secp256k1. Some selected examples existing Blockchains that use this ECC curve are Bitcoin, Ethereum and all ERC-20 tokens and many more.

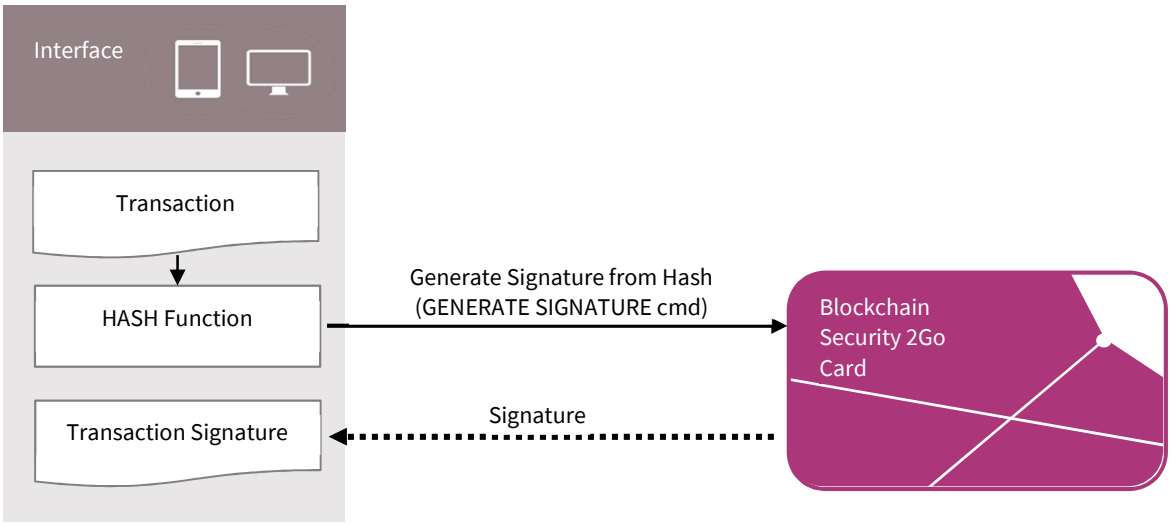

Figure 13 - To generate a signature of a transaction, first the transaction message is hashed, then the Blockchain Security 2Go card calculates a signature of this hashed message.

Another typical usage scenario of asymmetric cryptography is to realize a **key agreement protocol**. Public-key systems are very powerful but require much more computational effort than symmetric cryptography. Thus, these methods are only recommended for encrypting small amount of data. In established key agreement protocols, asymmetric cryptography is used to create a shared secret between two parties. This shared secret can then be used to exchange messages that are encrypted with symmetric cryptography as these cryptographic schemes offer a much better performance.

A very widely used key agreement protocol is the **Elliptic-curve Diffie-Hellman (ECDH) protocol** that allows two parties to establish a shared secret over an insecure channel when each party is related to an elliptic-curve public-private keypair. For example, the Transport Layer Security (TLS) protocol – a widely used encryption protocol for secure communication

in the internet (e.g. to secure HTTP, IMAP, SMTP, FTP etc.) – applies ECDH. The TLS specification supports the elliptic curves secp256r1, secp384r1, and secp521r1 [30].

It can be seen that popular Blockchain systems and popular internet protocols use different ECC curves. All curves mentioned above are defined in the SEC 2 standard [31]. The "k" in secp2564k1 stands for Koblitz and the "r" in secp256r1 for random. A Koblitz elliptic curve has some special properties that allow to implement the cryptographic calculations more efficiently. It is believed that there is a small security trade-off, meaning that completely random selected parameters are more secure. However, it offers some room for suspicion that the random coefficients may be selected by agencies in such a way they provide a back door. The NSA recommends the random curve secp256r1 – also known as NIST P-256 – for government use. To sum it up, apart from blockchain systems, traditional communication techniques typically use the NIST defined ECC curves.

In the scope of DataVaults, two types of hardware trust anchors can be used to create signatures: a TPM or the Infineon Blockchain Security 2Go starter kit. Currently exiting TPM solutions support the NIST curves. Thus, an out-of-the box TPM cannot be used to create signatures for transactions in currently widespread Blockchains. However, it offers support for traditional internet protocols. In contrast, the Blockchain Security 2Go starter kit supports all Blockchains based on ECC using the secp256k1 curve.

The specific ECC curve that is required by different Blockchains and offered by hardware-based solutions is an important aspect that will be considered when designing the underlying Blockchain infrastructure (as part of WP4) and deciding on which technologies should be used and/or how the hardware-based solutions should be extended.

6.4 DATAVAULTS TRUSTED LEDGER-BASED OPERATIONS

Trusted Platform Modules (TPMs) are a central building block of DataVaults and form the basis for enhanced **security**, **privacy** and **reliability** guarantees for ledger management and maintenance. The smart integration of the TPM technology will allow DataVaults to develop new Blockchain verification methods and significantly advance the state-of-the-art of Blockchain operation services: The DataVaults framework will not only use TPMs for user authentication and access authorization or to build secure Blockchain wallets, but also continuously attest and assess the security of involved devices in a privacy-preserving way and use TPM features to build efficient alternatives to rather inefficient or biased mining procedures. This hardware-based root of trust will be injected at the application points and endpoints to protect the integrity, validity, and usability of all data and process controls across the network.

***Trusted Authentication*:** To secure communication and prevent impersonation and man-in-the-middle attacks, peer authentication in such data sharing environments is of extreme significance. DataVaults will offer **multi-tier secure authentication based on hardware root of trust**: (i) **trusted identity authentication** between peers, (ii) **trusted membership authentication** for read and write on ledger, (iii) **trusted access authentication** for cloud-cased storage system, and (iv) **trusted actioner authentication** for data search and sharing. Using the hardware roof of trust anchor, DataVaults guarantees that a device or a party

claims what it is that is exactly what it is, which means that trust can be delivered throughout the whole data lifecycle.

Trust over Cryptographic Operations: The physical roof of trust anchor will guarantee that the operations based on cryptographic tools are really executed in a secure way. Recall that symmetric encryption, ABE, HMAC, SE, PRE, and digital signatures will be used to guarantee the data confidentiality, integrity, traceability and secure data search and sharing. Since these secure tools will be deployed in different scenarios and physical locations, it may open an opportunity for malicious network attackers to infiltrate into the adjunct software interfaces of the tools. DataVaults will explore research on the merge and extension of the current trusted hardware devices to support advanced cryptographic algorithms in an efficient and cost-effective way so as to definitely reduce probability of the case where a secure encryption tool is maliciously controlled and executed by network attackers. DataVaults's root of trust enhancement guarantee the reliability and trustworthiness of the execution of cryptographic operations over data sharing environments.

***Trusted Ledger Operations*:** The operation to the ledger is twofold: read and write. For the former, it is not difficult to achieve a **secure read** on ledger via trusted membership authentication, i.e., if a party with a trusted hardware device, who sends a read request, passes the membership authentication, then it will be allowed to read information stored on ledger. For the latter, **secure write operation**, DataVaults will focus on secure information mining with hardware roof of trust. All parties (vertical and horizontal) within the network may be allowed to manage the platform's private ledger, which means that these parties can be mined to verify the validity of data flow and further mine the data on the ledger. This can be achieved by reusing trusted authentication approach to make other parties believe that the current miner is an authentic party within the network.

Trusted Ledger Payment: DataVaults will leverage hardware roof of trust to deliver reliability in user compensation when sharing their data. To eliminate impersonation and minimize transaction fraud, a lightweight trusted hardware will be embedded into a party's account wallet for enabling payer/payee authentication. A party will receive payment from another if a data trading is successful concealed and a payment event is triggered (note this is ensured by using smart contract). The payment will be transferred directly from the payer's account wallet to that of payee by following the payment details instructed in corresponding smart contract. The transactions of the payment will be further validated and recorded onto the ledger. The trusted wallet will also efficiently help trusted membership authentication on ledger. More specifically, if a trusted hardware is embedded into a Blockchain wallet, inside of a user's platform, this authentication will directly be connected to secure access control over a ledger. Anytime when a user accesses to the ledger, the user is authenticated through the trusted hardware in the user's wallet.

6.5 DATAVAULTS TRUSTED BLOCKCHAIN CONTROL SERVICES

The DataVaults project aims to achieve **high security** and **privacy** guarantees by using these TPM services (security establishment, secure execution and attestation of a hardware TPM) as central building blocks. In doing so, the TPMs will be used as a **primary, hardware-based root of trust** and in multiple ways to provide security and privacy services. We will employ

TPMs to build **trusted blockchain wallets** and to **protect and continuously attest the platform integrity of all involved user devices**. To this end, every participating device will be equipped with a TPM to guarantee the platform integrity of involved devices and, in doing so, to ensure the authenticity of provided data by (and the trustful behaviour of) participating entities based on specified policies, may it be Data Providers, Data Brokers or Data Collectors. Furthermore, it will be possible to transfer blockchain wallets from one user device to the other, enabling each user to participate with several devices, thus, achieving **trusted user flexibility** (more information will be provided in the context of WP2).

Trusted Blockchain Wallet: In the DataVaults framework, TPMs are the basis for trusted Blockchain wallets. They will be used to:

- provide strong (i.e. two-factor) user authentication and to securely store the user credentials based on the TPM's secure key storage;
- control and authorize access to *private* or *public* ledger channels based on the user authentication process (e.g., to authorize access to or operations on different ledgers), and
- securely and efficiently verify Blockchain updates.

In this way, DataVaults will significantly advance the state-of-the-art of Blockchain verification methods: Unlike current mechanisms that often rely on computationally costly and wasteful proofs of work or biased proofs of stake**, DataVaults will use TPMs as central building block to build a very resource-efficient and trustful two-staged blockchain verification mechanism, which will be even suitable for resource-constrained devices** (such as smart devices - equipped with a TPM). Towards this direction, the platform will use the TPM's PCRs to securely store the current blockchain state hash and further use the TPM's hashing accelerator to speed up hash computations as required during blockchain operations (e.g. to compute the Merkle hash tree as described above). Where applicable, DataVaults will use other TPM hardware acceleration mechanisms as well (e.g., the TPM's asymmetric crypto coprocessors to accelerate the issuing and verification of digital signatures).

From a TPM perspective, the continuous verification procedure of Blockchain edits can be outlined as follows, where we will assume that all participating entities hold the current Blockchain state hash inside their TPMs: In Stage 1, the data broker will perform a pending Blockchain update, and will then determine the updated Blockchain state hash based on the ledger updates and the current state hash. Then, in Stage 2, the chosen verifiers (and any other DataVaults users) are able to verify the update. This involves checking the validity of the updated blockchain state based on the block update and the current blockchain state hash. On success, the users will then replace the current state hash inside their TPMs with the updated one.

Trusted Blockchain Attestation: In order to guarantee that only **trusted and uncompromised devices can participate in DataVaults**, all involved devices will use the TPM secure boot mechanism and their trust level will be continuously attested and assessed. To this end, all signatures on DataVaults data (e.g., transactions, smart contracts) will include the respective platform's integrity state (which is the hash value held by the device's PCRs at the end of the secure boot process), which will allow any other party to check whether the

data stems or was acknowledged by a trusted DataVaults user. **Depending on the selected privacy level, a conventional or a privacy-preserving signature scheme may be employed**. In the former case, a plain digital signature scheme supported by the TPM (e.g. ECDSA) will be selected, whereas in the latter case the TPM-provided DAA scheme can be used as strong privacy-preserving signature scheme. DAA (Section 5.2.4) can provide anonymous authentication, attestation and date integrity services. Several DAA schemes and their applications are specified in ISO/IEC 20008 and ISO/IEC 20009, respectively.

7 CONCLUSIONS

Towards the definition of a holistic DataVaults Data Security and Privacy Framework, this deliverable elicits the legal, ethical, security, privacy and trust requirement for DataVaults cloud-based platform and Personal App, lingering both on an ethical and legal perspective and on a technical viewpoint. This elicitation relies on several sources, ranging from the legal frame conditions of the European data protection framework (mainly GDPR and the current status of the upcoming ePrivacy Regulation) and other key regulatory and ethical sources, to the SoTA analysis on the data anonymization and pseudonymization techniques, encryption and authentication methods and data protection algorithms.

For the latter, this deliverable presented and assessed the most suitable and robust encryption technologies needed to secure different types of information, while still allowing advanced knowledge discovery through the provision of enhanced data search services, and advanced security and privacy-preserving primitives for authentication, authorization, attestation and verification through the use of trusted computing technologies. Such an analysis will serve as the basis and provide valuable insights on the identification of the most appropriate security technologies to be further investigated and enhanced in WP2 for the establishment of decentralized security and privacy-preserving environments with the inherent support of the ability reach a shared truth or trust that everyone agrees on without intermediaries, resulting in efficient allocation of resources and lower construction costs.

The survey is also enriched by initial insights on DLT and smart contracts that constitute core building blocks of DataVaults towards capturing data sharing (while complying with the prevailing GDPR legislation), collection, and compensation and trading preferences among the DataVaults parties for guaranteeing the trusted consent management among users. Advanced crypto primitives for enhanced security and user-controlled privacy, aiming to put the users in control of their own privacy and that of their generated data, were presented alongside a set of components for the secure and efficient computation, management and audit of all data sharing transactions. This set of services, to be integrated into the envisioned distributed ledger infrastructure, will enhance the framework's overall security and reliability by guaranteeing ledger management and maintenance.

The common ground of these SoTA surveys and requirements setting can be retrieved in 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. This has been envisaged and recommended by BDVA for unlocking the social value of personal data and fostering individual human empowerment and flourishing, in conjunction with their business and economic value and the same is at the core of DataVaults vision.

For achieving this, the future work of this WP will be consistent with the overarching project's progress, in order to embed GDPR compliance into the whole system and its tools and services, thereby allowing data subjects and data owners to remain in control of their data and its subsequent use, and, at the same time, to properly face with the management of privacy / utility trade-offs and to preserve utility for data analysis. This demands for technological-empowered balancing operations (for instance exploiting machine-learning capabilities) and a multi-layer approach in data sharing, within the boundaries of the rule of

law. Both of them will be further explored and fostered in the next stages of development of DataVaults reference framework, platform and application.

8 REFERENCES

[1] DATAVAULTS Consortium, «DataVaults proposal,» 2019.

[2] E. Union, «GDPR.eu,» 2020. [Online]. Available: https://gdpr.eu/article-4-definitions/. [Consultato il giorno February 2020].

[3] C. Gentry, A FULLY HOMOMORPHIC ENCRYPTION SCHEME, DEPARTMENT OF COMPUTER SCIENCE OF STANFORD UNIVERSITY, 2009.

[4] B. W. Amit Sahai, «Fuzzy Identity Based Encryption,» *IACR Cryptology ePrint Archive,* p. 86, 2004.

[5] S. N. ,. W.-O. ,. W. Jean-Philippe Aumasson, «BLAKE2: simpler, smaller, fast as MD5,» 29 29 2013. [Online]. Available: https://blake2.net/blake2.pdf. [Consultato il giorno 25 02 2020].

[6] M. J. Dworkin, «SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functior 04 August A2015. [Online]. Available: https://www.nist.gov/publications/sha-3-standard-permutation-based-hash-and-extendable-output-functions. [Consultato il giorno 25 02 202

[7] M. K. G. L. D. T. K. V. I. V. Andrey Bogdanov, «SPONGENT: The Design Space of Lightweight Cryptographic Hashing.,» *IACR Cryptology ePrint Archive,* p. 67, 2011.

[8] H. K. T. O. H. Y. Kota Ideguchi, «An AES Based 256-bit Hash Function for Lightweight Applications: Lesamnta-LW,» *IEICE Transactions,* n. 1, pp. 89-99, 2012.

[9] «ADVANCED ENCRYPTION STANDARD (AES),» 26 November 2001 . [Online]. Available: https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.197.pdf. [Consultato il giorno 25 02 2020

[10] L. R. K. C. P. A. P. M. J. B. R. Y. S. C. V. Andrey Bogdanov, «PRESENT: An Ultra-Lightweight Bl Cipher,» *International Workshop on Cryptographic Hardware and Embedded Systems - CHE* pp. 450-466, 2007.

[11] R. L. Rivest, 20 02 1997. [Online]. Available: https://people.csail.mit.edu/rivest/Rivest-rc5rev.pdf. [Consultato il giorno 25 02 2020].

[12] K. S. V. Anuj Kumar Singh, «A Lightweight Signcryption Scheme based on Elliptic Curve Cryptography,» 02 2014. [Online]. Available: https://www.researchgate.net/publication/260488960_A_Lightweight_Signcryption_Schei ased_on_Elliptic_Curve_Cryptography. [Consultato il giorno 25 02 2020].

[13] K. S. T. A. S. M. a. T. I. Taizo Shirai, «The 128-bit Blockcipher CLEFIA,» [Online]. Available: https://www.iacr.org/archive/fse2007/45930182/45930182.pdf.

[14] A. Will e C. D, A Practical Guide to TPM 2.0, Apress, 2015.

[15] S. Kinney, Trusted Platform Module Basics: Using TPM in Embedded Systems, 2006.

[16] Infineon Technologies, Blockchain Security 2Go User Manual, 2019.

[17] «NIST SP 800-35B Guide to Information Technology Security Services,» 2003. [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-35/final.

[18] S. Nakamoto, «Bitcoin: A Peer-to-Peer Electronic Cash System.,» 2009. [Online]. Available: https://metzdowd.com..

[19] M. &. S. J. &. N. L. &. K. C. &. S. S. &. N. D. &. M. A. Saad, in *Exploring the Attack Surface of Blockchain: A Systematic Overview.*, 2019.

[20] P. T. S. &. M. D. Seijas, in *Scripting smart contracts for distributed ledger technology. IACR Cryptology ePrint Archive, 2016, 1156.*, 2016.

[21] N. Szabo, in *Formalizing and securing relationships on public networks. First Monday, 2(9)*, 19997.

[22] S. S. P. a. D. J. B. K. Mohanta, «"An Overview of Smart Contract and Use Cases in Blockchai Technology",» in *9th International Conference on ComputingCommunication and Networki Technologies (ICCCNT), pp. 1-4.*, Bangalore, 2018.

[23] [Online]. Available: https://coinmarketcap.com/exchanges/yunbi/.

[24] J. Guggiari., «BlockChain: La tecnología que descentraliza al mundo.,» 2015.

[25] vDice, «vDice Ether Betting Game,» [Online]. Available: https://www.vdice.io/.

[26] M. C. S. F. S. P. A. R. P. T. A. B. W. I. X. X. Z. J. Staples, in *Risks and opportunities for systems using blockchain and smart contracts. Data61 (CSIRO)*, Sydney, 2017.

[27] T. Hardjoni e N. Smith, «Decentralized Trusted Computing Base for Blockchain Infrastructu Security,» 2019.

[28] A. B. V. B. C. C. K. C. A. D. C. D. E. C. F. G. L. Y. M. S. M. C. M. B. N. M. S. G. S. K. S. A. S. C. S. S. W. C. E. Androulaki, «yperledger Fabric: A Distributed Operating System for Permissione Blockchains,» 2018.

[29] NIST, «FIPS 180-4 Secure Hash Standard (SHS),» [Online]. Available: https://csrc.nist.gov/publications/detail/fips/180/4/final.

[30] Internet Engineering Task Force (IETF), «RFC 3279 Algorithms and Identifiers for the Intern X.509 Public Key Infrastructure,» 2002.

[31] Certicom Research, «SEC 2: Recommended Elliptic Curve Domain Parameters,» 2010.

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

[33] C. Batini, A. Rula, M. Scannapieco e G. Viscusi, «From data quality to big data quality,» in *B*

Data: Concepts, Methodologies, Tools, and Applications, IGI Global, 2016, pp. 1934--1956.

[34] E. Curry, «The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches,» in *New Horizons for a Data-Driven Economy*, Springer, 2016, pp. 29--37.

[35] "Data Protection in the era of Artificial Intelligence. Trends, existing solutions and recommendations for privacy-preserving technologies" (BDVA[29], October 2019).

[36] European Data Protection Supervisor, Opinion 7/2015 "Meeting the challenge of Big Data. A call for transparency, user control, data protection by design and accountability"

[37] Article 29 Working Party "Opinion 15/2011 on the definition of consent", adopted on 13 July 2011

[38] Article 29 Working Party "Opinion 03/2013 on purpose limitation", adopted 2nd of April 2013

[39] Article 29 Working Party "Opinion 01/2017 on the Proposed Regulation for the ePrivacy Regulation", adopted on 4 April 2017

[40] Article 29 Working Party "Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679", adopted on 4 April 2017

[41] Article 29 Working Party "Guidelines on consent under Regulation 2016/679", adopted on 10 April 2018

[42] The DataVaults Consortium, "D1.1 - DataVaults Data Value Chain Definition", 2020

[43] Buterin, Vitalik. "Ethereum Whitepaper". github. Retrieved 1 February 2020.

[44] Mihir Bellare, Phillip Rogaway: Collision-Resistant Hashing: Towards Making UOWHFs Practical. CRYPTO 1997: 470-484.

[45] Merkle, R. C. (1988). "A Digital Signature Based on a Conventional Encryption Function". Advances in Cryptology — CRYPTO '87. Lecture Notes in Computer Science. 293. p. 369.

[46] Giuseppe Ateniese, Kevin Fu, Matthew Green, Susan Hohenberger: Improved proxy re-encryption schemes with applications to secure distributed storage. ACM Trans. Inf. Syst. Secur. 9(1): 1-30 (2006).

[47] Matt Blaze, Gerrit Bleumer, Martin Strauss: Divertible Protocols and Atomic Proxy

[29] Big Data Value Association

Cryptography. EUROCRYPT 1998: 127-144.

[48] Michel Abdalla, Mihir Bellare, Dario Catalano, Eike Kiltz, Tadayoshi Kohno, Tanja Lange, John Malone-Lee, Gregory Neven, Pascal Paillier, Haixia Shi: Searchable Encryption Revisited: Consistency Properties, Relation to Anonymous IBE, and Extensions. J. Cryptology 21(3): 350-391 (2008).

[49] Mihir Bellare, Alexandra Boldyreva, Adam O'Neill: Deterministic and Efficiently Searchable Encryption. CRYPTO 2007: 535-552.

[50] David Cash, Stanislaw Jarecki, Charanjit S. Jutla, Hugo Krawczyk, Marcel-Catalin Rosu, Michael Steiner: Highly-Scalable Searchable Symmetric Encryption with Support for Boolean Queries. CRYPTO (1) 2013: 353-373.

[51] Ronald L. Rivest, Adi Shamir, Leonard M. Adleman: A Method for Obtaining Digital Signatures and Public-Key Cryptosystems (Reprint). Commun. ACM 26(1): 96-99 (1983).

[52] Shafi Goldwasser, Silvio Micali, Ronald L. Rivest: A Digital Signature Scheme Secure Against Adaptive Chosen-Message Attacks. SIAM J. Comput. 17(2): 281-308 (1988).

[53] Shehar Bano, Alberto Sonnino, Mustafa Al-Bassam, Sarah Azouvi, Patrick McCorry, Sarah Meiklejohn, George Danezis: Consensus in the Age of Blockchains. CoRR abs/1711.03936 (2017).

[54] "ISO/IEC 11889-1:2009 - Information technology -- Trusted Platform Module -- Part 1: Overview". ISO.org. International Organization for Standardization. May 2009.

[55] Schnorr C.P. (1990) Efficient Identification and Signatures for Smart Cards. CRYPTO' 89 Proceedings. CRYPTO 1989. Lecture Notes in Computer Science, vol. 435. Springer, New York, NY.

[56] The Swirlds hashgraph consensus algorithm: fair, fast, Byzantine fault tolerance. Technical Report SWIRLDS-TR-2016-01, Swirlds, Inc.,2016.

[57] S. Nakamoto. Bitcoin: A peer-to-peer electronic cashsystem. https://bitcoin.org/bitcoin.pdf, Dec 2008. Accessed: 2019-03-06.

[58] P. Daian, R. Pass, and E. Shi. Snow white: Provably secure proofs of stake. Cryptology ePrint Archive, Report 2016/919, 2016. http://eprint.iacr.org/2016/919.

[59] T. Hønsi. SpaceMint: A Cryptocurrency Based on Proofs of Space. IACR Cryptology ePrint Archive, 2017.

[60] P4Titan. Slimcoin: A Peer-to-Peer Crypto-Currency with Proof-of-Burn. http://www.slimcoin.club/whitepaper.pdf, 2014.

[61] Intel Software Guard Extensions - https://software.intel.com/en-us/sgx/details.

[62] Hyperledger Sawtooth. https://intelledger.github.io/introduction.html. Accessed: 2019-03-17.

[63] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proceedings of the 13th ACM conference on Computer and Communications Security, pp. 89–98, 2006.

[64] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in Proceedings of IEEE Symposium on Security and Privacy, pp. 321-334, 2007.

[65] R. Ostrovsky, A. Sahai, and B. Waters, "Attribute-based encryption with non-monotonic access structures," in Proceedings of the 14th ACM conference on Computer and Communications Security, pp. 195–203, 2007.

[66] Kaitai Liang, Man Ho Au, Joseph K. Liu, Willy Susilo, Duncan S. Wong, Guomin Yang, Yong Yu, Anjia Yang: A secure and efficient Ciphertext-Policy Attribute-Based Proxy Re-Encryption for cloud data sharing. Future Generation Comp. Syst. 52: 95-108 (2015).

[67] Kaitai Liang, Man Ho Au, Joseph K. Liu, Willy Susilo, Duncan S. Wong, Guomin Yang, Tran Viet Xuan Phuong, Qi Xie: A DFA-Based Functional Proxy Re-Encryption Scheme for Secure Public Cloud Data Sharing. IEEE Trans. Information Forensics and Security 9(10): 1667-1680 (2014).

[68] Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky, Giuseppe Persiano: Public Key Encryption with Keyword Search. EUROCRYPT 2004: 506-522.

[69] Reza Curtmola, Juan A. Garay, Seny Kamara, Rafail Ostrovsky: Searchable symmetric encryption: Improved definitions and efficient constructions. Journal of Computer Security 19(5): 895-934 (2011).

[70] My Health My Data. http://www.myhealthmydata.eu/. Accessed 2019-03-11.

[71] McConaghy, T., Marques, R., Müller, A., et al. "BigchainDB: A Scalable Blockchain Database". https://www.bigchaindb.com/whitepaper/.Accessed 2019-03-11.

[72] FAREDGE. http://www.faredge.eu/. Accessed 2019-03-17.

[73] Hyperledger Fabric. https://www.hyperledger.org/projects/fabric. Accessed 2019-03-11.

[74] Goldreich, O., & Ostrovsky, R. (1996). Software protection and simulation on oblivious RAMs. Journal of the ACM (JACM), 43(3), 431-473.

[75] Canetti, R. (1996). Studies in secure multiparty computation and applications. pp73-79, March.

[76] Chang, Y. C., & Mitzenmacher, M. (2005, June). Privacy preserving keyword searches on remote encrypted data. In International Conference on Applied Cryptography and Network Security (pp. 442-455). Springer, Berlin, Heidelberg.

[77] Song, D. X., Wagner, D., & Perrig, A. (2000, May). Practical techniques for searches on encrypted data. In Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000 (pp. 44-55). IEEE.

[78] Curtmola, R., Garay, J., Kamara, S., & Ostrovsky, R. (2011). Searchable symmetric encryption: improved definitions and efficient constructions. Journal of Computer Security, 19(5), 895-934.

[79] European Cyber Security Organization (ECS). European Cybersecurity Strategic Research and Innovation Agenda (SRIA) for a contractual Public-Private Partnership (cPPP). June 2016. [Available Online]: https://ecs-org.eu/documents/ecs-cppp-sria.pdf

[80] IEEE 1609 WG. Family of Standards for WAVE. 2009.

[81] M. Zhao, J. Walker, and C.-C. Wang. "Security Challenges for the Intelligent Transportation System". In: Security of Internet of Things. SecurIT '12. 2012.

[82] M. Feiri, J. Petit, and F. Kargl. "Formal model of certificate omission schemes in VANET". In: IEEE VNC. 2014.

[83] M. Gerlach and F. Guttler. "Privacy in VANETs using Changing Pseudonyms - Ideal and Real". In: IEEE Vehicular Technology Conference. 2007.

[84] S. Gisdakis, M. Lagana, T. Giannetsos, and P. Papadimitratos. "SEROSA: SERvice Oriented Security Architecture for Vehicular Communications". In: IEEE Vehicular Networking Conference, 2013.

[85] J. Petit, F. Schaub, M. Feiri, and F. Kargl. "Pseudonym Schemes in Vehicular Networks: A Survey". In: IEEE Communications Surveys and Tutorials (2015).

[86] S. Gisdakis, T. Giannetsos, and P. Papadimitratos. "SPPEAR: Security & Privacy-preserving Architecture for Participatory-sensing Applications". In: ACM. WiSec '14.

[87] ISO/IEC 11889:2009 (all parts) Information technology – Trusted platform module.

[88] Trusted Computing Group (TCG), Trusted Platform Module Library - Part 1: Architecture (Family 2.0, Revision 01.38), 2016.

[89] Trusted Computing Group (TCG), Trusted Platform Module Library - Part 2: Structures (Family 2.0, Revision 01.38), 2016.

[90] Trusted Computing Group (TCG), Trusted Platform Module Library - Part 3: Commands (Family 2.0, Revision 01.38), 2016.

[91] Trusted Computing Group (TCG), Trusted Platform Module Library - Part 4: Supporting Routines (Family 2.0, Revision 01.38), 2016.

[92] W. Arthur, D. Challener and K. Goldman, A Practical Guide to TPM 2.0 - Using the Trusted Platform Module in the New Age of Security, Apress Media, 2015.

[93] G. Proudler, C. Liqun and C. Dalton, Trusted Computing Platforms - TPM 2.0 in Context, Springer, 2014.

[94] A. Segall, Trusted Platform Modules - Why, when and how to use them, The Institution of Engineering and Technology, 2017.

[95] H. Shahriar and M. Zulkernine, "Mitigating Program Security Vulnerabilities: Approaches and Challenges," ACM Comput. Surv., vol. 44, no. 3, pp. 11:1-11:46, 2012.

[96] I. Welch and R. J. Stroud, "Using Reflection as a Mechanism for Enforcing Security Policies in Mobile Code," in European Symposium on Research in Computer Security (ESORICS), Toulouse, France, 2000.

[97] S. M. Ghaffarian and H. R. Shahriari, "Software Vulnerability Analysis and Discovery Using Machine-Learning and Data-Mining Techniques: A Survey," ACM Comput. Surv., vol. 50, no. 4, pp. 56:1-56:36, 2017.

[98] R. Yavatkar, D. Pendarakis and R. Guerin, "RFC 2753 - A Framework for Policy-based Admission Control," 2000.

[99] E. F. Brickell, J. Camenisch and L. Chen, "Direct anonymous attestation," in ACM Conference on Computer and Communications Security (CCS), 2004.

[100] 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.

[101] E. Brickell, L. Chen and J. Li, "Simplified security notions of direct anonymous attestation and a concrete scheme from pairings," *International Journal of Information Security,* 2009.

[102] J. Camenisch, L. Chen, M. Drijvers, A. Lehmann, D. Novick and R. Urian, "One TPM to Bind Them All: Fixing TPM 2.0 for Provably Secure Anonymous Attestation," in IEEE Symposium on Security and Privacy (S&P), 2017.

[103] J. Camenisch, M. Drijvers and A. Lehmann, ""Anonymous Attestation with Subverted TPMs," in Advances in Cryptology - CRYPTO, 2017.

[104] S. Goldwasser, S. Micali and C. Rackoff, "The knowledge complexity of interactive proof systems," in SIAM Journal on computing, 1989.

[105] K. E. Defrawy, G. Holland and G. Tsudik, "Remote Attestation of Heterogeneous Cyber-Physical Systems: The Automotive Use Case," in ESCAR, 2015.

[106] R. Sailer, X. Zhang, T. Jaeger and L. v. Doorn, "Design and implementation of a TCG-based Integrity Measurement Architecture," in 13th USENIX Symposium, 2004.

[107] N. Asokan, F. Brasser, A. Ibrahim, A. Sadeghi, M. Schunter, G. Tsudik and C. Waschmann, "SEDA: Scalable Embedded Device Attestation," in Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015.

[108] Sweeney, L. 2002. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10, 5, 557–570.

[109] Samarati, P. 2001. Protecting Respondents' Identities in Microdata Release. IEEE Transactions on Knowledge and Data Engineering (TKDE) 13, 6, 1010–1027.

[110] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. 2006. l-Diversity: Privacy Beyond k-Anonymity. In Proc. of Intl. Conf. on Data Engineering (ICDE). SP800-108

[111] GDPR, "General Regulation on data protection 2016/679, of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data". It can be retrieved, for instance, at the following link: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[112] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). It can be retrieved, for instance, at the following link: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:l24120&from=EN

[113] Charter of Fundamental Rights of the European Union, 2016/C 202/02. It can be retrieved at the following link: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12016P/TXT&from=EN

[114] The European Convention on Human Rights, adopted in 1950 and entered into force in 1953. The Convention and its Protocols can be retrieved at the following link:https://www.coe.int/en/web/conventions/search-on-treaties/-/conventions/treaty/results/subject/3

[115] EC's Communications "AI for Europe" (25 April 2018) and "Building Trust in Human-Centric AI" (8 April 2019)

[116] "Data Protection in the era of Artificial Intelligence. Trends, existing solutions and recommendations for privacy-preserving technologies" (BDVA, October 2019)

[117] "Meeting the challenge of Big Data. A call for transparency, user control, data protection by design and accountability" (Opinion 7/2015, European Data Protection Supervisor, 2015)

[118] Greek Law 4624/2019 on the protection of natural persons with regard to the processing of personal data

[119] Belgian Law of 30 July 2018 on the protection of natural persons with regard to the processing of personal data ("Law of 30 July 2018")

[120] Organic Law 3/2018 of 5 December on the Protection of Personal Data and the Guarantee of Digital Rights

[121] Italian Legislative Decree n° 101 of 10th August 2018

[122] DataVaults D1.1 “DataVaults Data Value Chain Definition”.

[123] ISO/IEC 9797-2

[124] ISO/IEC 9797-2

[125] ISO/IEC 10116:2006

[126] SP800-108

9 ANNEX 1. REGULATORY FRAMEWORK IN THE SELECTED JURISDICTIONS

9.1.2 Demonstrator #1 – Sports and Activity Personal Data and Demonstrator #2 – Strengthening Entrepreneurship and Mobility

National regulatory landscape

The Greek Law 4624/2019 on the protection of natural persons with regard to the processing of personal data, fully covers the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. The Greek Law only deviates from the General Data Protection Regulation (hereafter GDPR) to a limited extent and will therefore have a minor impact on the processing of personal data by private companies and organizations.

Greek Law 4624/2019

Structure

Utilizing the "opening clauses" of the GDPR provided to member states, the new law supplements the GDPR in three significant ways. First, it supplements the GDPR on general issues that are left to the discretion of member states. Second, it regulates special cases of processing, which are considered important for the national legislator. Third, the law imposes restrictions on the rights of data subjects when necessary and proportionate for purposes of public interest. Specifically, Section A of the law stipulates its objective and scope, the definitions of public and private entities, and the role of the data protection officer in public bodies. Section B includes provisions regarding the organization and operation of the Hellenic Data Protection Authority. In Section C, supplemental measures for the application of the GDPR are implemented, whereas Section D incorporates the LED Directive into Greek law.

Main provisions and highlights relating to processing of personal data

The main provisions of the law concern the following regulatory issues:

Age of consent: If a minor is 15 years old and provides consent, their data in relation to information society services can be lawfully processed. If a minor is under the age of 15, the consent of parent or guardian is required.

Special categories of data: Apart from the legal basis of Article 9 of the GDPR, processing special categories of data by public and private entities is permitted without the consent of the data subject, when it is mandatory for health care, social care, Social Security and to assess an individual's ability to work. Furthermore this should be done with proviso that there are appropriate measures in place to safeguard data subjects. Processing special categories of data by public entities for further purposes is permitted, in cases of public interest, the necessity of preventing a significant threat for public safety and the necessity to take humanitarian measures. Nevertheless, processing genetic data for health and life insurance is expressly prohibited.

Processing for further purposes: The processing of personal data by public entities for purposes other than those for which they have been collected is permitted in cases in which it is necessary for the prosecution of criminal offenses, public safety reasons and prevention of harm of another person. Similarly, the processing by private entities is permitted in cases in which they are subject to national security issues or for the foundation, exercise or support of their legal claims. Such processing by private entities is permitted in order to prevent threats against national security or public health after a public entity's request for either the prosecution of criminal offenses or the establishment, exercise or defence of legal claims, unless the interest of the data subject to his/her data not to be processed is outweighed.

Limitations on the rights of data subjects

The law provides for exceptions from the obligation to inform data subjects when such information would jeopardize the proper performance of the controller's duties, public security or the establishment or exercise or defence of legal claims. The exercise of the right of access is also restricted when there is not any obligation to inform the data subject or when their data has been recorded and cannot be deleted due to regulatory provisions about their obligation to retain or control them, such as when their information is stored on tax documents, fingerprints, passports, etc. The right to erasure of personal data does not apply in cases of non-automated processing, when erasure is impossible due to the special nature of their storage or requires a disproportionate effort and where it is contrary to conventional or legal retention periods. In certain cases of automated processing, the right to erasure may also be lawfully replaced by restrictions to processing of the relevant data. Finally, the right to object before public entities may not be applicable if processing is required for the public interest, when the latter prevails over the interests of the data subject.

Special cases of processing

The law stipulates specific provisions about the cases of processing related to the freedom of expression and information, the context of employment, and archiving purposes in the public interest, as well as scientific purposes and purposes of historical research or purposes related to the collection or retention of statistics. In the particular issue of employment, the law delimits the lawful purposes of processing to only those which are necessary for the recruitment, the performance and execution of the employment contract. If the processing is based on the legal grounds of the employee's consent, the validity of consent is evaluated according to the circumstances of the specific employment contract and the conditions of consent pursuant to Article 7 of the GDPR. The processing of personal data is also permitted on the basis of collective labor agreements. Finally, the surveillance through CCTV systems in the workplace is only permitted when it is necessary for the protection of persons and goods and when written or electronic notice is provided to employees.

Variations of GDPR on right of information to be provided

When personal data is collected from the data subject, the controller is exempt from the obligation to inform data subjects of further processing of personal data pursuant to Article 13(3) of the GDPR in the following cases (Article 31 of the Data Protection Law):

- the processing purpose of the further processing is compatible with the initial purpose, the communication with the data subject is not conducted via digital means and data subject's interest to be informed is not particularly high; or
- when, in case of a public entity, such information would jeopardise:
    - the proper performance of the controller's duties;
    - the national or public security and the controller's interests not to provide the information override the data subject's interests;
    - the establishment, exercise or defence of legal claims and the controller's interests not to provide the information override the data subject's interests; or
    - the confidential transfer of personal data to public entities.

The controller must take appropriate measures for the protection of data subjects' legitimate interests, including the provision of information outlined in Article 13(1) and (2) of the GDPR in an accurate, transparent, intelligible and easily accessible manner, in a clear and plain language.

In addition, broader exceptions apply for public entities when personal data have not been obtained from the data subject, under Article 32 of the Data Protection Law.

Variations of GDPR on right to erasure

Under Article 34 of the Data Protection Law, the right to erasure does not apply, in cases of non-automated processing, when due to the special nature of storage, erasure is impossible or is possible only following a disproportionate effort and data subject's interest for the erasure is not considered important. In such cases, erasure is substituted by restriction of processing. The same exception applies where erasure would be contrary to conventional or legal retention periods. The above does not apply in case of unlawful processing.

Variations of GDPR on right to object

Under Article 35 of the Data Protection Law, the right to object may not be applicable before a public entity, if processing is required for the public interest, when the latter prevails over data subjects' interests or processing is obligatory under a legal provision.

Variations of GDPR on right of access

Under Article 33 of the Data Protection Law, the right of access is restricted when:

- there is no obligation to inform data subjects; or
- when data subjects' data:
    - were recorded only because they could not have been deleted due to regulatory provisions of obligatory retention; or
    - serve exclusively purposes of protection or control of data,

- and the provision of information would require a disproportionate effort and the necessary technical and organisational measures to make processing impossible for other purposes.

The reasons for refusing to provide access to the data subject must be documented.

National data protection authority

The Hellenic Data Protection Authority (HDPA) is responsible for monitoring the implementation of the GDPR provisions, the Data Protection Law and other provisions related to the protection of persons against the processing of personal data in the Greek territory.

Besides its powers under Article 58 of the GDPR, the HDPA has been provided with the following investigative and corrective powers under Article 15 of the Data Protection Law:

- to carry out, *ex officio* or following a complaint, investigations and audits over compliance with the provisions of the Data Protection Law;
- to address warnings to the controller or processor that intended processing operations are likely to infringe provisions of the Data Protection Law;
- to order the controller or processor to bring processing operations into compliance with the provisions of the Data Protection Law, in a specified manner and within a specified period, particularly by means of an order for the rectification or erasure of personal data;
- to order and impose a temporary or definitive limitation and/or ban on the processing of personal data;
- to order and impose the delivery to the authority of documents, filing systems, equipment or processing means of personal data and their content;
- to seize any documents, information, filing systems of any equipment and means of a personal data breach, including their content, that comes to its attention when exercising its investigatory powers and be declared as a sequestrator until issuance of a decision by competent judicial authorities;
- to order the controller or processor to interrupt the processing of personal data, to return or 'freeze' the relevant data, or to destroy the filing system or relevant data;
- to impose administrative sanctions under Article 83 of the GDPR and Article 39 of the Data Protection Law;
- to impose administrative sanctions under Article 82 of the GDPR;
- to issue a provisional order; and
- to issue administrative regulatory acts in order to regulate specific, technical and detailed matters.

Contact information of the Greek Data Protection Agency:

Data Protection Authority Offices: Kifissias 1-3, 115 23 Athens, Greece

Call Centre: +30-210 6475600

Fax: +30-210 6475628

E-mail: contact@dpa.gr

9.1.3 Demonstrator #3 – Healthcare Data Retention and Sharing

National regulatory landscape

The Belgian Law of 30 July 2018 on the protection of natural persons with regard to the processing of personal data ("Law of 30 July 2018") entered into force on 5 September 2018.

The Law only deviates and/or complements the General Data Protection Regulation (hereafter GDPR) to a limited extent and will therefore have a minor impact on the processing of personal data by private companies and organizations.

Hereafter are some of the most noteworthy provisions of the Law:

Material and Territorial scope

The Law explicitly states in article 6 that the GDPR remains fully applicable in relation to the processing of personal data in the private sector except in those cases where the Law supplements the GDPR. In principle, this article 6 is unnecessary but Belgian legislation has introduced it for reasons of clarity.

Moreover, it has been stated that the Law will apply to companies and organizations that process personal data:

- in relation to the activities of an establishment which is situated on Belgian territory, irrespective of where the processing takes place or
- in relation to data subjects residing on Belgian territory, even if the company is not established there, and it offers goods and services to these subjects on Belgian territory or it monitors the behaviour of such data subjects, for as far as this behaviour takes place on Belgian territory or
- which are established in a place where Belgian law is applicable under public international law.

The Law will not apply to a processor established on Belgian territory, if the controller is established in another EU Member State and when the processing takes place on the territory on which the controller is established. In that case, the law of the other EU Member State will be applicable.

Child's consent

The GDPR allows EU Member States to provide for an age lower than 16 years regarding a child's consent, as long as it does not go below the age of 13 years old. Belgian Law has chosen to make full use of this possibility and lower the age to 13. If a child is below the age of 13 years, the processing of its personal data will only be lawful and valid if and to the extent that consent is given or authorized by the holder of parental responsibility over the

child. Therefore, each data controller should implement an adequate system that can verify the parental consent for children under 13 years old.

Special categories of personal data

The GDPR prohibits the processing of special categories of personal data (i.e. racial or ethnic origin, etc.). However, there are several exemptions in which case the processing of these special categories of personal data is allowed. One of the exemptions allows Member States to determine when the processing of these special categories of personal data is necessary for reasons of substantial public interest. Consequently, the Law lists three situations in which processing is deemed to be of substantial public interest. The most relevant of these three situations relates to processing by associations who have as their statutory goal the defence and improvement of human rights and fundamental freedoms. The two other situations apply to sex offenders and are irrelevant to private companies and organizations.

Data concerning health

The General Data Protection Regulation (GDPR) recognises data concerning health as a special category of data and provides a definition for health data for data protection purposes. Though the innovative principles introduced by the GDPR (privacy by design or the prohibition of discriminatory profiling) remain relevant and applicable to health data as well, specific safeguards for personal health data and for a definitive interpretation of the rules that allows an effective and comprehensive protection of such data have now been addressed by the GDPR. Processes that foster innovation and better quality healthcare, such as clinical trials or mobile health, need robust data protection safeguards in order to maintain the trust and confidence of individuals in the rules designed to protect their data.

The Belgian Law introduces three new obligations for the data controller or processor:

- to indicate which categories of persons, have access to the data and explain their relation to the processing of the personal data.
- to maintain a list of these categories of persons for the Belgian data protection authority.

to make sure that the designated persons are subject to a legal, statutory or equal contractual obligation to ensure the confidential character of the personal data.

National Data Protection Authority

The Data Protection Authority (DPA) is an independent body ensuring the protection of privacy when personal data is processed.

The DPA is the successor of the Commission for the protection of privacy as of 25/05/2018 and was established by the Belgian Federal House of Representatives with the Act of 3 December 2017 establishing the Data Protection Authority.

Contact

Autorité de la protection des données - Gegevensbeschermingsautoriteit (APD-GBA)
Rue de la Presse 35 – Drukpersstraat 35
1000 Bruxelles - Brussel

Tel. +32 2 274 48 00
Fax +32 2 274 48 35
email: contact@apd-gba.be
Website: https://www.autoriteprotectiondonnees.be/ - https://www.gegevensbeschermingsautoriteit.be/
Member: Mr David Stevens, President

9.1.4 Demonstrator #4 – Smarthome Personal Energy Data

National Regulation

The Spanish Data Protection Legislation has incorporated the General Data Protection Regulation. In order to adapt and develop certain matters contained in the European Regulation, the Spanish Parliament has approved the Organic Law 3/2018 of 5 December on the Protection of Personal Data and the Guarantee of Digital Rights, full text can be found (in Spanish) in the following link, https://www.boe.es/boe/dias/2018/12/06/pdfs/BOE-A-2018-16673.pdf .

Organic Law 3/2018 of 5 December

The Law facilitates the exercise of the specific rights conferred to the data subjects by requiring that the means for exercising such rights are easily accessible.

- **The right to information**

  In order to enforce the principle of transparency, the new Law regulates the way in which citizens are informed about the processing of their data and opts for a layered information system. In a first layer, Data subjects shall therefore be informed about the basic aspects of the processing Data collection: identity of the Data Controller, the purpose of processing and the rights the subjects possess among other basic information. They must be informed also about how to access to more detailed information -through a direct link- contained in a second layer, if they so require.

- **The right to access**

  The new Law recognizes the right of access and, where appropriate, the right to rectify or suppress the data of deceased persons to persons connected with them, unless the deceased had prohibited such access, rectification or deletion.

- **Minors consent**

  With regard to the processing of the personal data of minors, the Law sets the minimum age for autonomous consent at 14 years. Similarly, it regulates the right to be forgotten in relation to data provided by minors to social networks and other information society services. This right may be exercised by the minor her/himself or by third parties while she/he´s still a minor.

- **Whistleblowing systems**

The Law also contains a specific article relating to the processing of personal data within the framework of whistleblowing systems; It allows anonymous dilation from employees to communicate infractions in Data Protection.

- **Video surveillance, sound and digital devices**

The Law updates the guarantees applicable to citizens in relation to the use of video surveillance devices, geolocation, sound recording and other digital devices in the workplace. The employer can use these devices to control their employees' work nevertheless they must be previously informed, and the devices cannot be placed in changing rooms, rest or dining areas.

Video and audio surveillance devices can be used in public places but only for security or safety reasons. All non-criminal related Data must be deleted in less than a month.

- **Processing Personal Data**

Controllers must retain data only for as long as is deemed necessary for the purpose of the collection and process thereof (3 months unless specific cases).

Controllers must keep it accurate and up to date.

Controllers must block Personal Data and avoid any possibility of Data processing and display, once it has been required to be rectified or deleted.

- **Credit Information System**

The new Law requires a minimum amount of 50 Euros for the inclusion of a person in a credit information system (delinquent file) and reduces from 6 to 5 years the maximum period of inclusion of debts in such files.

- **Digital Rights**

The Law specifically includes new "digital rights" such as universal access to internet, secure and appropriate use of Personal Data, and the right for users to rectify and object that applies in digital media and social networks.

National Data Protection Authority

The Spanish Data Protection Agency is the independent Public Authority in charge of the Data Protection and was established with the Organic Law 5/1992 (LORTAD).

Contact information of the Spanish Data Protection Agency:

Agencia Española de Protección de Datos (AEPD)
C/Jorge Juan, 6
28001 Madrid
Tel. +34 91399 6200
Fax +34 91455 5699
email: internacional@aepd.es
Website: https://www.aepd.es/

In previous projects the process in order to get data from users was as follows:

1. An explanatory letter was sent to the AEPD explaining the projects, including objectives and data to be gathered from users. In this letter it has to assure the compliance of the Data Protection Regulation.
2. The AEPD sent us a response letter with comments about the project. If the Agency observe some issues it would claim for further explanations. In previous projects no issues were found but in all cases it's not a final approval, because at any time AEPD could require validations of the compliance of the regulation.

9.1.5 Demonstrator #5 – Personal Data for Municipal Services and the Tourism Industry

National Regulation[30]

On 19th September 2018 Legislative Decree n° 101 of 10th August 2018 came into force to adjust the Italian personal data protection code (Legislative Decree no. 196 of 30th June 2003) to the provisions of (EU) Regulation 2016/679.

The general part of the Italian Privacy Code is almost entirely replaced by the provisions of the Regulation, so that the previously valid rules on the principles, legal basis of the processing, information and consent are now repealed and replaced by those of the European Regulation.

Nevertheless, since the GDPR leaves to Member States the possibility of introducing further conditions, including limitations, with reference to the treatment of genetic, biometric or health data, legislation decree 101/2018 transfers such possibility to the National Data Protection Authority who will be able to issue specific regulations, that need to be taken into account.

With regard to the special part of the text, **the main novelties** are listed below:

Curriculum vitae

Legislative Decree 101/2018 states that the notice under art. 13 GDPR is to be given on the "first suitable occasion" after the sending of the curriculum vitae. Within the limits of the purposes described in article 6 par. 1) letter b) of GDPR, the consent of the applicant to the processing of personal data contained in the curriculum is not required.

Remote monitoring

The provisions of Article 4 of the Workers' Statute (as amended in 2015 by the Jobs Act) shall be expressly without prejudice and the penalty pursuant to Article 38 of Law 300/1970 is also confirmed for cases of violation of paragraph 1 of Article 4 of the Workers' Statute.

Simplification for SMEs

[30] Content extracted from https://www.laborproject.it/en/2018/09/20/legislative-decree-n-101-of-10th-august-2018-now-in-force/

The reform protecting SMEs included in the new art. 154-bis, par. 4 of the Privacy Code (and introduced by Legislative decree 101/2018) is especially important, as it states that, with regard to micro and small and medium enterprises, given the simplification requirements, the Supervisory Authority shall include simplified ways to comply with the obligations of the data controller in its guidelines.

Consent of minors

With regard to the direct offer of "services of the information company", consent can be given by minors upon reaching 14 years of age. Below this limit consent shall be given by the adult who exercises parental responsibility.

Codes of ethics and general authorisations

Lawmakers decided to guarantee continuity by accepting the provisions of the Supervisory Authority on a provisional basis, to be reviewed later on. The Supervisory Authority, with a general provision to be discussed and published within ninety days from the coming into force of the decree, shall identify the provisions contained in the general authorisations that are compatible with the provisions of the GDPR and of legislative decree 101/2018 and, if necessary, shall update them.

The general authorisations thus audited that are considered incompatible with the GDPR shall cease to be effective.

The Supervisory Authority is also required to promote the issue of codes of ethics dealing with the processing of personal data in some sectors (work, journalism, statistics and scientific research), involving the interested parties and making a public consultation.

Sanctions

The Italian lawmakers decided to introduce penalties, as allowed by the GDPR with regard to all Member Countries, for some violations of the privacy laws; such penalties are to be added to the severe administrative sanctions established in the Regulation (up to 20 million Euro or 4% of the gross annual world turnover). The penalties punish:

- the unlawful processing of personal data;
- the illegal acquisition of personal data subject to large-scale processing;
- the illegal communication and dissemination of personal data subject to large-scale processing;
- false statements made to the Supervisory Authority;
- non-compliance with the Supervisory Authority provisions;
- violation of any provisions on remote monitoring and workers opinion surveys.

In the presence of an especially wide-ranging and strict system of administrative sanctions, characterised by a strong deterrence, some early commentators highlighted a possible violation of the "ne bis in idem" prohibition, in respect of some behaviours.

Personal data of deceased persons

The rule dealing with personal data of deceased persons is worth mentioning here.

The rights under articles 15 to 22 of the GDPR concerning the personal data of deceased persons can be exercised by subjects holding a personal interest or acting to protect the interested party as an agent or for family reasons worth protecting.

The exercise of the abovementioned rights is not allowed when it is prohibited by law, or when – limited to the "direct offer of services of the information company" – the data subject expressly forbade it with a written and unequivocal statement.

This prohibition cannot however produce effects that penalise the exercise by third parties of the rights of property derived from the death of the data subject or the right to defend one's own interests in court.

National Data Protection Authority

The Italian Data Protection Authority (Garante per la protezione dei dati personali) is an independent administrative authority established by the so-called privacy law (Law No. 675 of 31 December 1996) and regulated subsequently by the Personal Data Protection Code (Legislative Decree No. 196 of 30 June 2003) as amended by Legislative Decree No. 101 of 10 August 2018, which also established that the Italian DPA is the supervisory authority responsible for monitoring application of the General Data Protection Regulation (pursuant to Article 51 of Regulation No. 2016/679).

Contact information:
Piazza Venezia 11 – 00187 Roma
Phone: +39-06-6967 71
Fax: +39-06-6967 73785
Phone: +39-06-6967 71 / +39-06-6967 72917
Certified mail: protocollo@pec.gpdp.it
Email: urp@gpdp.it