D3.2 Distributed Privacy-preserving Data Management and Storage Final Version

Document Identification			
Status	Final	**Due Date**	28/02/2025
Version	1.0	**Submission Date**	28/02/2025

Related WP	WP3	**Document Reference**	D3.2
Related Deliverable(s)	D3.1	**Dissemination Level**	PU
Lead Participant	DUT	**Lead Author**	Kaitai Liang (DUT) Dazhuang Liu (DUT)
Contributors	NOR, DUT, UTH, UMU, VTT, FUJ_GE, KUL, DBC, FHG	**Reviewers**	Athanasios Stratikopoulos (UOM)
			Sotirios Michagiannis (DBC)

Keywords:
Data sharing, trustworthiness, confidentiality, distributed data storage, privacy, ePrivacy, Energy Efficiency

Disclaimer for Deliverables with dissemination level PUBLIC
This document is issued within the frame and for the purpose of the TANGO project. This project has received funding from the European Union's Horizon Europe Framework Programme under Grant Agreement No. 101070052. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.
The dissemination of this document reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains. This deliverable is subject to final acceptance by the European Commission.
This document and its content are the property of the TANGO Consortium. The content of all or parts of this document can be used and distributed provided that the TANGO project and the document are properly referenced.
Each TANGO Partner may use this document in conformity with the TANGO Consortium Grant Agreement provisions.

Document Information

List of Contributors	
Name	Partner
Kaitai Liang	DUT
Dazhuang Liu	DUT
María Hernández	UMU
Jesús García	UMU
Mohamed Belkhechine	FSDE
Jürgen Neises	FSDE
Felix Hermsen	FHG
Ville Ollikainen	VTT
Alexandra Papageorgiou	KUL
Andrea Palumbo	KUL
Peggy Valcke	KUL
Ioannis Drivas	DBC
Christophoros Manolis	NOR
Vitalii Demianets	NOR
Natalia Borgoñós	UMU

Document History			
Version	**Date**	**Change editors**	**Changes**
0.1	20/12/2024	Kaitai Liang (DUT) Dazhuang Liu (DUT)	Table of Contents is created
0.2	24/12/2024	Tomas Pariente Lobo (EVIDEN)	Table of Contents to be approved
0.3	03/02/2025	Dazhuang Liu (DUT)	Merge inputs from each task
0.4	12/02/2025	Kaitai Liang (DUT)	Refined document based on reviewers' comments
0.5	24/02/2025	Dazhuang Liu (DUT)	Format issues corrected
1.0	25/02/2025	Kaitai Liang (DUT)	FINAL VERSION TO BE SUBMITTED

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	Kaitai Liang (DUT)	05/01/2025
Quality manager	Jürgen Neises (FSDE)	25/02/2025
Project Coordinator	Tomás Pariente (ATOS)	26/02/2025

Table of Contents

Document Information 2
Table of Contents 3
List of Tables 7
List of Figures 8
List of Acronyms 9
Executive Summary 10
1 Introduction 11
1.1 Purpose of the document 11
1.2 Relation to other project work 11
1.3 Structure of the document 11
2 Overview of WP3 12
2.1 Overview of WP3 main components 13
3 Description of components 15
3.1 Blockchain-based Data Storage and Sharing [T3.1] 15
3.1.1 Introduction 15
3.1.2 Description of the component 15
3.1.3 Internal Architecture 15
3.1.4 Features implemented 16
3.1.5 Advancement of components 16
3.1.6 Support for pilots 17
3.1.7 Future Investment 17
3.1.8 Conclusion 17
3.2 Trustworthy Data Sharing [T3.2] 17
3.2.1 Introduction 17
3.2.2 Description of the component 18
3.2.2.1 TSM 18
3.2.2.2 PRS 19
3.2.2.3 UPCV 20
3.2.3 Features implemented 21
3.2.3.1 TSM 21
3.2.3.2 PRS 22
3.2.3.3 UPCV 22
3.2.4 Advancement of components 22
3.2.4.1 TSM 22
3.2.4.2 PRS 22
3.2.4.3 UPCV 22
3.2.5 Support for pilots 22

3.2.5.1 TSM 22
3.2.5.2 PRS 22
3.2.5.3 UPCV 23
3.2.6 Future Investment 23
3.2.6.1 TSM 23
3.2.6.2 PRS 23
3.2.6.3 UPCV 23
3.2.7 Documentation 23
3.2.7.1 TSM 23
3.2.7.2 PRS 23
3.2.8 Conclusion 23
3.3 Confidentiality and Privacy by Design [T3.3] 23
3.3.1 Introduction 23
3.3.2 Description of the component 24
3.3.2.1 User Consent and Access Control 24
3.3.2.2 Standard API for Access Management 25
3.3.2.3 Defining and Registering New Policies in the PAP component 25
3.3.2.4 Integration with TSM and UCBA (context-aware policies) 26
3.3.2.5 Sticky Policies 26
3.3.2.6 Hybrid encryption with SEDSS 26
3.3.2.7 Interaction PEP & Sticky Policies 27
3.3.3 Internal Architecture 27
3.3.3.1 User Consent and Access Control 27
3.3.3.2 Sticky Policies 28
3.3.4 Features implemented 30
3.3.5 Advancement of components 31
3.3.6 Support for pilots 31
3.3.7 Future Investment 32
3.3.8 Conclusion 32
3.4 Self-encryption and Decryption Techniques with Multi-Factor Information Recovery Mechanisms [T3.4] 32
3.4.1 Introduction 32
3.4.2 Description of the component 33
3.4.3 Internal Architecture 33
3.4.4 Features implemented 34
3.4.5 Advancement of components 35
3.4.6 Support for pilots 35
3.4.6.1 Functional Testing Workflow and Toolkit for FMAKE 35
3.4.6.2 On the Encryption and Decryption Efficiency of SEDSS for Real world Data 37
3.4.6.3 Encryption and Decryption on various Data Format 37
3.4.7 Future Investment 39

3.4.8 Conclusion 39
4 Recommendations for secure and privacy-preserving data storage and sharing [T3.5] 40
4.1 Legal Considerations 40
4.1.1 Blockchain-Based Data Storage and Sharing [T3.1] 40
4.1.1.1 Identification of personal data stored in FIDES 40
4.1.1.1.1 Qualification of encrypted data as pseudonymous personal data 40
4.1.1.1.2 Qualification of individual data chunks separately stored in DRILL as personal data 41
4.1.1.1.3 Personal data in plaintext or encrypted (non-hashed) form 41
4.1.1.1.4 Reference data stored in the blockchain: stamps of approval, timestamps and logs 42
4.1.1.1.5 Reference data stored in the blockchain: hashes 43
4.1.1.2 Application of the GDPR to the processing of personal data in a decentralised environment 45
4.1.1.2.1 Preliminary considerations 45
4.1.1.2.2 Territorial and material applicability of the GDPR to personal data processing in a decentralised environment 45
4.1.1.2.3 Identification of the controller(s) and processor(s) for FIDES use in the banking pilot 46
4.1.1.2.4 Identification of the processor 48
4.1.1.2.5 Compliance with data protection principles of Article 5 GDPR in a decentralised environment 48
4.1.1.2.6 Compliance with data subjects' rights in a decentralised environment 51
4.1.1.2.7 Compliance with requirements on security of the processing 53
4.1.1.2.8 Conclusions 53
4.1.2 Elements that can constitute a proof of customer consent to process personal data 55
4.1.2.1 General considerations 55
4.1.2.2 Elements to be used to prove consent and general factors to take into account 55
4.1.3 Re-evaluation of the previous legal advice on FIDES 56
4.1.4 WP3 Protocol compliance and recommendations 56
4.2 Trustworthy Data Sharing [T3.2] 59
4.3 Confidentiality and Privacy by Design [T3.3] 60
4.3.1 Confidentiality and Privacy by Design mechanisms 60
4.3.1.1 Sticky policies (CPE-ABE) 60
4.3.1.2 User Consent (CPE & XAMCL) 60
4.3.2 WP3 Protocol compliance and recommendations 61
4.4 Self-Encryption and Decryption Techniques with Multi-factor Information Recovery Mechanism [T3.4] 63
4.4.1 Material and territorial applicability of the GDPR 63
4.4.1.1 Encrypted and hashed data as pseudonymous data 64
4.4.1.2 Identification of the personal data processed in the self-encryption process in the pilots 64
4.4.2 Identification of data controller(s) and data processor(s) in the self-encryption process 65
4.4.3 Compliance with data protection principles of Article 5 GDPR 66
4.4.3.1 Identification of a legal basis for the processing 66

4.4.3.2 Display of users' IDs and data minimisation 67
4.4.3.3 Continued processing of user IDs and data ownership records: compliance with storage limitation 68
4.4.3.4 Security requirements 69
4.4.3.5 Exercise of data subjects' rights 72
4.4.4 WP3 Protocol compliance and recommendations 72
4.5 On the ePrivacy Directive 74
4.5.1 General confidentiality obligation 74
4.5.1.1 Applicability of the obligation to personal and non-personal data 74
4.5.1.2 Applicability of the obligation to the activities of the platform 74
4.5.1.3 Applicability to the activities of the platform and interplay with the GDPR 76
4.6 Ethical considerations 76
4.6.1 General privacy aspects 76
4.6.2 Regarding the blockchain-based Data Storage and Sharing 78
5 Conclusions 80
6 References 81
Annexes 83
A.1 Encryption API of SEDSS 83
A.2 Decryption API of SEDSS 84
A.3 On the Encryption and Decryption Efficiency of SEDSS for Real world Data 84

List of Tables

Table 1: WP3 components _____ *13*
Table 2 Fides features implemented in D3.2. _____ *16*
Table 3 Features implemented in T3.3 _____ *30*
Table 4 A summary table to summarize the features of previous version in the last deliverable and the current status. _____ *34*
Table 5 The description of Encryption API design. _____ *83*

List of Figures

Figure 1 WP3 components relation ____ *12*
Figure 2 Internal architecture of Fides platform ____ *16*
Figure 3 - TSM Process and APIs ____ *19*
Figure 4 Overview PRS ____ *20*
Figure 5 UPCV web app, executed on mobile and desktop browsers ____ *21*
Figure 6 - TSM backend API ____ *22*
Figure 7 User Consent and Access Control components ____ *25*
Figure 8. CP-ABE submodules ____ *26*
Figure 9. Internal Workflow of the PEP/PDP Component for User Access Requests ____ *28*
Figure 10. CP-ABE workflow ____ *30*
Figure 11 The overall architecture of SEDSS. ____ *34*
Figure 12 The diagram of functional testing workflow between user's side and SEDSS service of TANGO server ____ *35*
Figure 13 The encryption step for functional testing. ____ *36*
Figure 14 The decryption step for functional testing. ____ *36*
Figure 15 The hash verification step for functional testing. ____ *36*
Figure 16 Comparison of encryption times for TXT, PDF, CSV and MP4 files. The measurements are presented in seconds for file sizes of 10 MB. ____ *37*
Figure 17 Comparison of encryption times for TXT, PDF, CSV and MP4 files. The measurements are presented in seconds for file sizes of 2 GB. ____ *38*
Figure 18 Comparison of decryption times for TXT, PDF, CSV and MP4 files. The measurements are presented in seconds for file sizes of 10 MB. ____ *38*
Figure 19 Comparison of decryption times for TXT, PDF, CSV and MP4 files. The measurements are presented in seconds for file sizes of 2 GB. ____ *39*
Figure 20 Encryption and decryption times for TXT files of varying sizes, ranging from 1 MB to 2 GB. ____ *85*
Figure 21 Encryption and decryption times for PDF files of varying sizes, ranging from 1 MB to 2 GB. ____ *85*
Figure 22 Encryption and decryption times for CSV files of varying sizes, ranging from 1 MB to 2 GB. ____ *85*
Figure 23 Encryption and decryption times for MP4 files of varying sizes, ranging from 1 MB to 2 GB. ____ *86*

List of Acronyms

Abbreviation / acronym	Description
ABE	Attribute-based Encryption
API	Application Programming Interface
CP-ABE	Ciphertext-Policy Attribute-Based Encryption
CJEU	Court of Justice of the European Union
DGA	Data Governance Act
DID	Decentralized Identifiers
DnAMS	Decentralized Access Management System
DoA	Description of Action
Dx.y	Deliverable number y belonging to WP x
EC	European Commission
EECC	European Electronic Communications Code
EDPS	European Data Protection Supervisor
EU	European Union
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
IIC	Industrial Internet Consortium
M2M	Machine to Machine
MFIR	Multi-Factor Information Recovery
OKVS	Oblivious Key-Value Storage
P2P	Peer-to-peer
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PIP	Policy Information Point
PRS	Privacy Risk Score
RRI	Robot Revolution Initiative
SaaS	Software as a Service
SED	Self-Encryption and Decryption
SSI	Self-Sovereign Identity
TSD	Trade Secrets Directive
TSM	Trustworthiness Scoring Module
UBEM	User Behaviour Exchange Module
UC	User Consent
UI	User Interface
UPCV	Ubiquitous Personal Context Vectors
VC	Verifiable Credentials
VP	Verifiable Presentations
VRN	Volatile Random Number
W3C	World Wide Web Consortium
WP	Work package
XACML	eXtensible Access Control Markup Language
ZKP	Zero-Knowledge Proof

Executive Summary

The deliverable D3.2 of the TANGO project presents the final outcomes of the software components developed within Work Package 3 (WP3), which focuses on distributed privacy-preserving data management and storage. The document provides an in-depth analysis of the implementation status of tools supporting AI-related activities, offering a comprehensive overview of the methodologies, architectural advancements, functionalities, and integrations developed since the previous deliverable, D3.1.

WP3 encompasses several key tasks designed to ensure secure, scalable, and privacy-compliant data management within the TANGO framework. Task T3.1 focuses on blockchain-based data storage and sharing, highlighting its readiness for pilot use cases in public organizations and financial institutions. Enhancements in functionality, scalability, and compliance have been integrated, with continued improvements planned to address evolving privacy and security challenges. Task T3.2 advances the development of three primary components: the Trustworthiness Scoring Module (TSM), the Privacy Risk Score (PRS), and Ubiquitous Personal Context Vectors (UPCVs). These components enhance data privacy evaluation and ensure secure, trustworthy data sharing. Significant progress has been made in making these components available for pilot use cases, with ongoing refinements to align with real-world applications. Task T3.3 emphasizes confidentiality and privacy by design, incorporating advanced cryptographic mechanisms to enhance data protection. This task integrates privacy-preserving techniques into the TANGO platform, ensuring regulatory compliance while enabling secure data processing and sharing. Task T3.4 successfully implements self-encryption and decryption mechanisms coupled with a multi-factor information recovery system. This innovation ensures confidentiality during data transmission and storage while providing a secure and scalable encryption framework. The developed APIs enable seamless integration into TANGO, offering encryption, decryption, and secure key management functionalities. Usability testing demonstrates the system's effectiveness across various data formats and sizes, with a functional testing toolkit enhancing accessibility for both technical and non-technical users.

Additionally, the deliverable outlines recommendations for secure and privacy-preserving data storage and sharing by T3.5, in section 4. It provides legal considerations regarding compliance with GDPR and the ePrivacy Directive, particularly addressing challenges related to blockchain-based data storage. The document emphasizes the need for a thorough Data Protection Impact Assessment (DPIA) when handling personal data, ensuring adherence to regulatory frameworks.

The conclusion of this deliverable underscores the achievements of WP3 in delivering robust, privacy-preserving data management solutions. Each task contributes to the overarching goal of providing secure, scalable, and efficient AI-driven data handling mechanisms. Collectively, these advancements position TANGO WP3 as a critical work package of secure and compliant data processing within diverse pilot applications, paving the way for continued enhancements and future research in privacy-preserving data management and processing solutions.

1 Introduction

1.1 Purpose of the document

The objective of this deliverable is to present the release of methodology and techniques that are used in WP3 of TANGO project. The document summarizes the work progress regarding functionality, connection to pilots. Within the document, each task uniformly presents its outcomes. In particular, the new contributions and techniques as well as updates compare to deliverable 3.1 are highlighted in this document.

1.2 Relation to other project work

This document is related to the outputs of T2.1 – Gap Analysis in Distributed Data Management, Processing & Storage, T2.2 – User Needs and Requirements for Data Management, Processing & Storage and T2.3 – Use Case Scenarios & KPIs Definition, where the GAP analysis in terms of technologies, the TANGO offerings, including those of WP5, and the mapping with the user requirements from pilots have been thoroughly described.

The document is also related to the work to be carried out in WP6 in terms of architecture, integration and testing in tasks T6.1 - Continuous Integration and Delivery and T6.2 Functional Testing and Monitoring. In the case of T6.1 the relation is bidirectional, as the integration will require the outputs of this deliverable in terms of software artifacts, as well as the architectural choices discussed in WP2 and WP6 have an impact in the way of implementing in the most effective way the results to facilitate the integration.

1.3 Structure of the document

The document is structured into the following key sections.

Chapter 1 outlines the purpose of this document as well as the relationship of each component in this deliverable to other project works, deliverables, documents and usecases.

Chapter 2 presents the methodological framework of the distributed privacy-preserving data management and storage system, and briefly summarize the progress of each technical component (task) produced after deliverable D3.1. The Recommendations for secure and privacy-preserving data storage and sharing, which is proposed to assist the design and implementation of technical tasks, is also summarized.

Chapter 3 details the outcomes of components. In particular, we divide this chapter into four subsections, each one correspondent to a technical component (from T3.1 to T3.4). Each subsection provides a detailed overview of the component, starting with an introduction and description, followed by its internal architecture and implemented features. It highlights advancements made since D3.1, with a summary table comparing previous and current features, and discusses how the component supports pilot use cases. The section concludes with future development plans, usage documentation, and a summary of the component's significance.

Chapter 4 presents the output of T3.5, which gives the recommendations for secure and privacy-preserving data storage and sharing.

Chapter 5 and 6 draw the conclusion of this deliverable and list the bibliography respectively.

2 Overview of WP3

In the scope of TANGO, WP3 Distributed Privacy-preserving Data Management and Storage is mainly related to provide confidential data storage and distribution to support use case application and other components in Tango framework. The work package has the following main objectives:

The objective of T3.1 is the development of a privacy-preserving, decentralized data-sharing framework within TANGO. T3.2 contains trustworthiness scoring module (TSM), the privacy risk score (PRS) and the ubiquitous personal context vectors (UPCVs), which provides privacy attribute values which make up a small part of the final trustworthiness score; The aim of T3.3 is to emphasizes confidentiality and privacy by design, incorporating advanced cryptographic mechanisms to enhance data protection; T3.4 aims to develop an flexible privacy-preserving encryption and decryption mechanism in order to protect confidential data during the transmission and in storage for TANGO project. The objective of T3.5 is to provide a holistic assessment of the technologies developed under WP3, by applying the protocol presented in D3.1, which is valid for both personal and non-personal data in order to guarantee confidentiality.

Distributed data management, storage and sharing solutions are at the core of the TANGO platform, interconnecting various tools that enhance untampered and secure data handling, sharing [T3.2] and re-use in various application areas such as public administration, smart hospitality, autonomous vehicles, smart manufacturing, banking and retail. The distributed data management and sharing will be based on energy-efficient blockchain technology [T3.1] ensuring the traceability, integrity and ownership of the data whilst maintaining GDPR compliance. A distributed data tokenisation solution combined with privacy and confidentiality by design [T3.3] will be developed, providing access control to the data. Distributed self-encryption/decryption combined with recovery solutions [T3.4] based on trust, will provide high security for data storing and sharing. Finally, the design and implementation of all technical components are supported by the policies given by T3.5.

Figure 1 shows graphically the main tasks and components implemented in WP3 in support of the distributed privacy-preserving data management system.

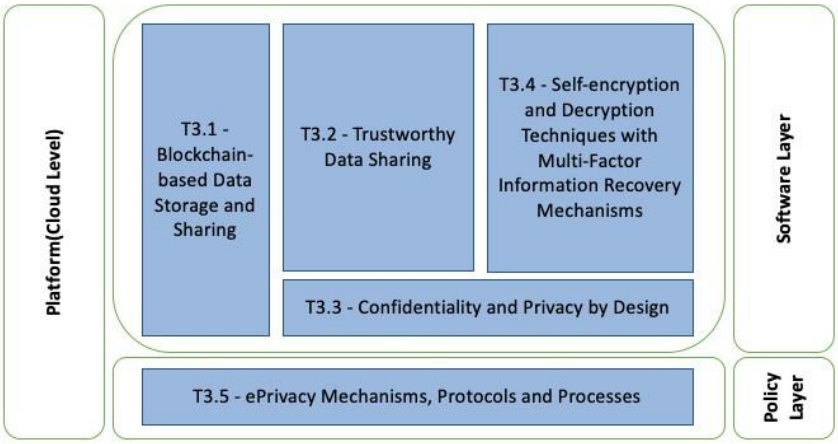

Figure 1 WP3 components relation

2.1 Overview of WP3 main components

The main components of WP3 and the main enhancements regarding the previous deliverable 3.1 are summarized below. In the next sections, the document provides a detailed view of the current status of the implementation of these components for the intermediate release.

Table 1: WP3 components

Task	Component name	Short description	What's new in D3.2
T3.1	Fides	This task will focus on the design and development of a novel energy efficient Blockchain mechanism that will act as the basis for distributed procedures for data management, sharing and storage. The platform will be based on Fides solution, a secure and trustworthy blockchain data sharing platform that currently focuses on exchanging Know-Your-Customer data among financial institutes to prevent fraud.	Improved integration with decentralized modules; Full development and deployment of DnAMS; Expanded pilot use cases, providing a more comprehensive framework.
T3.2	TSM	A sub-task is to design and develop the Trustworthiness Scoring Module (TSM) that is tasked with assessing the trustworthiness of both data-sharing partners (consumer and provider). This evaluation relies on the IIC trustworthiness model, which leverages system characteristics. These characteristics are appraised as a weighted sum of the degree to which attributes fulfill protection objectives or properties.	The Trustworthiness Scoring Module is available to the Automotive and the Manufacturing (RIA) use case as backend version providing information to the PEP or as user integrated version. The scoring is based on historic pilot data regarding a user specific configuration. The Privacy Risk Score may be leveraged as contributor to the scoring.
	UPCV	Another sub-task is associated with the technique of Ubiquitous Personal Context Vectors (UPCV). That is incorporated into a distributed architecture, where each user and vendor has their own dedicated instances of a User Behaviour Exchange Module (UBEM). This symmetric component implements an algorithm that remains uniform on both the user and vendor (service) sides. Moreover, UBEM incorporates an additional capability for producing recommendations, generated on users' UBEMs and then presented on their respective user interfaces (UI).	For the 'Smart hospitality' use case, UPCV is now implemented as a distributed service, each vendor having data of their items in their servers, and users having their data on mobile apps. Collaborative recommendations take place locally on user devices, without disclosing any details of the user. For the 'Retail' use case, a shopping basket dataset has been received and the analysis is ongoing.
	Privacy risk score	The last sub-task is to establish a technical privacy risk score for the data to be shared, based on simulated	The privacy risk score has been implemented for the autonomous driving use case. The module takes as

Task	Component name	Short description	What's new in D3.2
		attacks and metrics rooted in information theory. The score is a vector or a single scalar value indicating the likelihood that individuals will be negatively affected by the sharing or processing of the data. It can also be used to estimate the risk of intellectual property disclosure.	input a sequence of anonymized images and assesses how well the anonymization satisfies a pre-defined privacy policy.
T3.3	Sticky policies	A sub-task is to focus on policy-based encryption. Data is encrypted according to attribute-based policies (ABE), which allows giving the user the power to control their data during the whole lifecycle.	API design finished. Integration complete with ABE toolset.
	User consent	Another sub-task is to use policy for access control. Policy-based access control is used to address user consent. Attribute-based policies will be used to enable fine-granular checks based on identity attributes.	Successfully implemented a standard API for storage. Integrated with TSM. Possibility to add policies coming from PAT tool. Testing implemented. TANGO deployment finished.
T3.4	SEDSS	This component transforms plaintext into ciphertext and reversed ciphertext to its original form through decryption using a specific algorithm and key. The cryptographic technique used to distribute a piece of sensitive information among a group of participants in such a way that only a specific subset of them can reconstruct the original secret.	We present the enhanced system architecture and newly implemented API design, all aimed at integrating SEDSS into the TANGO framework. We tested the SEDSS with various file types. Results demonstrates that SEDSS is able to efficiently deal with various file types.
T3.5	ePrivacy Mechanisms, Protocols and Processes	This task aims to expand the notion of confidentiality of communication, included in art.5 ePrivacy Directive, broadening the scope of application of such a principle to all data exchanged and processed online.	Building upon the overview of all technologies developed in WP3 and the privacy protocol which were provided in D3.1, a holistic legal assessment of the technologies supported by a practical application of this protocol are included in D3.2. for the purposes of this task.

3 Description of components

3.1 Blockchain-based Data Storage and Sharing [T3.1]

3.1.1 Introduction

The Fides platform is a blockchain-based solution designed by norbloc to address the challenges of secure and compliant data sharing, specifically in regulated environments such as finance and public administration. Its primary function is to facilitate Know-Your-Customer (KYC) and Know-Your-Business (KYB) data exchanges, reducing redundancies and increasing operational efficiency among financial institutions.

Within the TANGO project, Fides serves as a key component in WP3, focusing on distributed data management and privacy preservation. By eliminating data silos and enabling real-time updates Fides fosters a "single version of truth" approach, critical for regulated industries aiming to streamline processes and enhance transparency.

3.1.2 Description of the component

The Fides platform addresses the complexities of data sharing in highly regulated industries. Designed with privacy, security and efficiency at its core, the platform ensures GDPR compliant exchanges of sensitive information, including KYC and KYB data. The following features underscore its functionality:

- **Decentralized Storage:** Fides uses blockchain-enabled infrastructure to store data securely, ensuring resilience and compliance. This eliminates reliance on centralized data storage systems, reducing risks associated with single points of failure.
- **Consent Management:** Through its Decentralized Access Management System (DnAMS), Fides enforces strict access control policies, ensuring that user consent is explicitly granted and traceable for all data-sharing activities. This guarantees alignment with data protection regulations.
- **Immutable Audit Logs:** Leveraging blockchain technology, Fides maintains tamper-proof records of data sharing events, creating a transparent and verifiable audit trail for both internal and external compliance checks.

3.1.3 Internal Architecture

The internal architecture of Fides platform, as shown in Figure 2 is designed to ensure robust, secure, and decentralized data sharing. Below is presented an overview of the Fides design and its internal components.

The Fides internal components that will be utilized in TANGO are:

- Proprietary norbloc Data Backend (DRILL)
- Proprietary norbloc Access Management System (DnAMS)
- Blockchain Layer

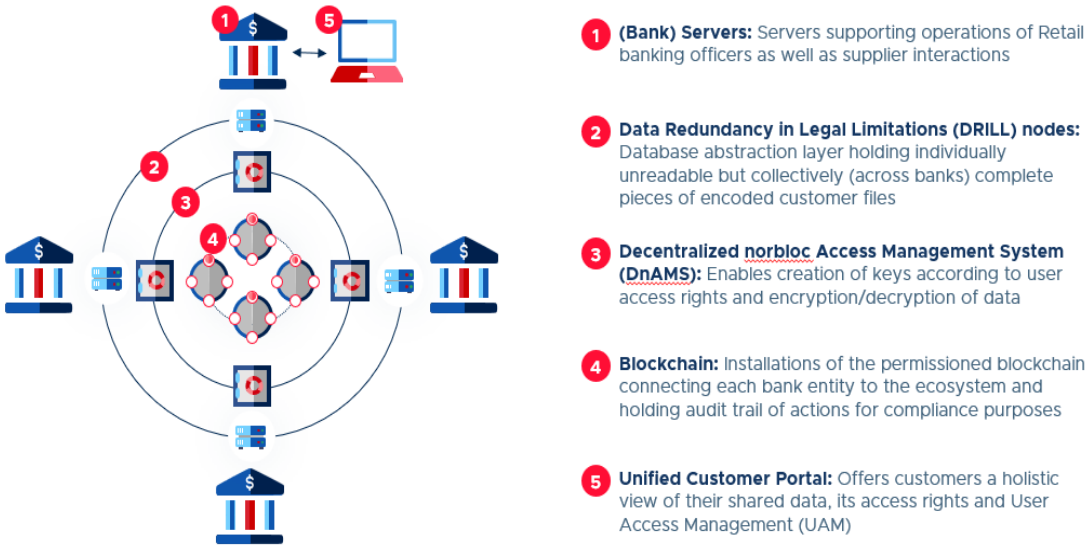

Figure 2 Internal architecture of Fides platform

3.1.4 Features implemented

Table 2 Fides features implemented in D3.2.

Feature	Status in D3.1	Status in D3.2	Updates & Explanations
Decentralized Data Storage (DRILL)	Fully Implemented	Fully Implemented	-
Decentralized Access Management (DnAMS)	Partially Implemented	Fully implemented	This is main development of D3.2 - DnAMS has been fully implemented
Blockchain Integration, including Immutable Audit Logs	Fully implemented	Fully implemented.	-
Real-time Synchronization	Fully Implemented	Fully Implemented.	-
Tango integration	Conceptual design	Partially implemented	Pilot T7.6 requirements were refined

3.1.5 Advancement of components

During the project duration a new Fides version was developed: Fides v3. Fides v3 introduces two main advancements: significant improvements in DnAMS and enhanced integration with the TANGO project.

1. In DnAMS, the key advancement is the achievement of full decentralization, which now allows for more robust and secure management of data access permissions and explicit

consent management. This decentralization is a core feature of Fides v3, reinforcing the platform's compliance with privacy regulations and eliminating the need for centralized control.

2. On the TANGO integration front, the platform underwent a detailed refinement process based on pilot feedback. The necessary integrations with TANGO were designed and developed to meet the specific requirements of the pilot. This ensures that Fides' functionalities are fully aligned with the objectives of the TANGO project, optimizing the platform for real-world use cases.

3.1.6 Support for pilots

Fides platform's robust architecture and modular design allow for flexible deployment across various pilots, demonstrating its applicability to diverse regulatory and operational contexts.

Fides' contribution concentrates on Public Organizations use case (T7.6). Fides is utilized to enable secure and transparent sharing of visa-related data between VISARIGHT and Immigration authorities. The platform's infrastructure ensures data integrity and transparency throughout the process.

Significant improvement for the pilot use case brought by Fides is the inclusion of the third party in the process, making the use case distributed between three independent parties instead of two. Alongside envisioned VISARIGHT and an Immigration authority, the third party is an Employer. The employer also participates in data exchange and produces some verification stamps that can be utilized by another two parties.

In such a setup Fides all essential Fides features (see 3.1.4) are fully exploited.

3.1.7 Future Investment

The development of Fides within the TANGO project revealed some integration challenges, so future work will focus on improving its integration with other TANGO components.

User Experience (UX) improvements will be a priority to simplify usage and onboarding. Enhancing the accessibility of Fides for different user roles, including financial institutions regulatory bodies, will be crucial to fostering adoption. Future iterations will also address usability refinements in the Unified Customer Portal, ensuring a more intuitive interface for managing data access and sharing permissions. Additionally, other key areas for improvement include integrating AI and Machine Learning for better data access management, security, and compliance automation.

Overall, these enhancements will ensure that Fides continues to maintain its competitive edge, and will remain a top-tier solution for secure and privacy-compliant data sharing across various sectors.

3.1.8 Conclusion

The deliverable highlights the platform's readiness through its integration into pilot use cases, including public organizations and financial institutions. The advancements achieved since D3.1 reflect a focused effort to enhance functionality, scalability and compliance.

As TANGO progresses, Fides will continue to evolve, incorporating insights from pilot activities and addressing emerging challenges in privacy and security.

3.2 Trustworthy Data Sharing [T3.2]

3.2.1 Introduction

As previously described the component consist of three independent modules and related sub-tasks. These are the Trustworthiness Scoring Module (TSM), the Privacy Risk Score (PRS) and the Ubiquitous Personal Context Vectors (UPCVs).

The TSM is a backend solution which has been designed for usage as contributor to the Policy Information Point (PIP). The PIP as a component of the PEP gathers information enabling the Policy Decision Point to take decisions which shall be enforced by the PEP granting or denying access (e.g. in the Manufacturing use case). Moreover, the TSM may be also used as a means of allowing a resource to be scheduled by a use case specific component (e.g. in the Automotive use case).

The TSM has been developed from a stand-alone version described in D3.1 to a backend version providing APIs and it has been prepared to leverage historic (timeseries) data to take a configurable history into account evaluating trustworthiness.

The PRS is a backend solution that supports the TSM in the automotive use case by providing privacy scores. This functionality can also operate independently of the TSM. Its core function is to conduct quality assurance tests on previously anonymized data and summarize this test in a time series of privacy scores.

These privacy scores are then used in the TSM as privacy attribute values which make up a small part of the final trustworthiness score.

In the automotive use case, images are collected, anonymized, and uploaded to the cloud for future use. To assess privacy protection, we implemented a module that allows the evaluation of the quality of image sequence anonymizations (time series) across different scenarios. The idea is that this aids the customers to verify that their vehicles effectively safeguard privacy.

The UPCV recommendation technology has two main components: server on vendor side, having information of offered items, and a web app for users – it is the web app that contains personal data and computes and presents recommendations. Since vendors have only item data, nothing about users, data can be shared between vendors which have similar items in their offerings. The UPCV algorithm may have properties that can be utilized for privacy-preserving business intelligence in a multi-vendor environment.

3.2.2 Description of the component

3.2.2.1 TSM

The TSM allows a scoring based on a configurable scoring scheme and continuously gathered attribute data. It is based on so-called system characteristics aligned with the IIC Trustworthiness schema. The component provides APIs for configuration, submitting attribute data and evaluating Trustworthiness scores.

Configurations and data are related to a registered user, which is identified by its DID. Attribute data are gathered in a timeseries data according to a configurable time window. Trustworthiness score then may be used by the PEP (primary case) or any local use case component for decisions on granting access to resources like machines or scheduling resources like autonomous vehicles.

The process flow includes

1. Configuration of the evaluation schema, i.e. selection of characteristics, weights of characteristics and related attributes, timeseries window
2. Submitting attribute data according to the configuration to the TSM's timeseries data base
3. Evaluation of Trustworthiness scores by the PEP or a use case specific module, e.g. CAVride

These steps and the API are illustrated in Figure 3.

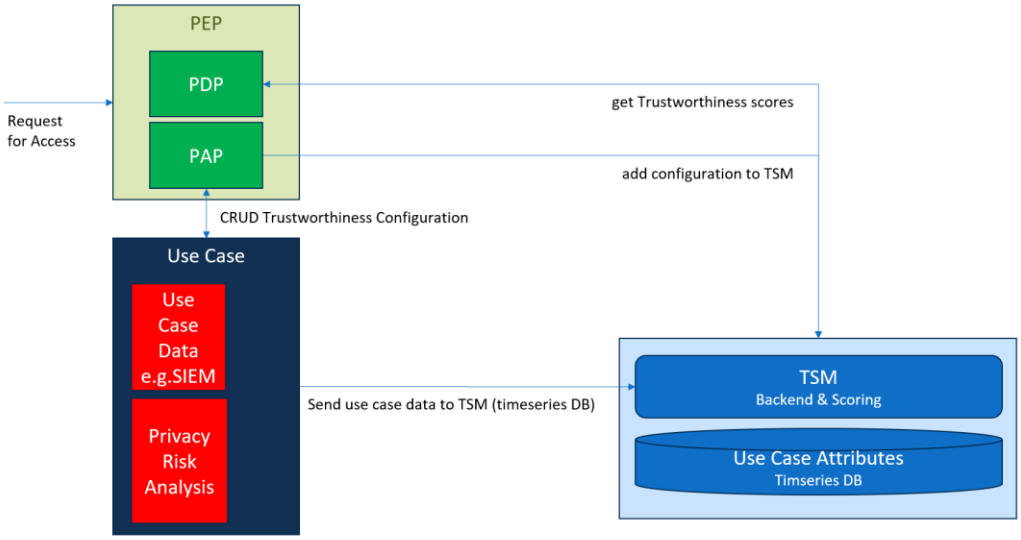

Figure 3 - TSM Process and APIs

3.2.2.2 PRS

The primary objective of the PRS is to rate anonymized images within a predefined time window, providing the TSM with a time series of privacy scores. However, the PRS can also serve as a platform for experimenting with different anonymization methods and evaluating their effectiveness under various privacy settings and locations.

For development, we primarily relied on open-source data (e.g., PP4AV). Although implementing an anonymization module is beyond the core functionality of the PRS, we opted to include it because the autonomous driving use case could not share its proprietary anonymization module due to confidentiality constraints and delays in providing anonymized images. As a result, we implemented a module with default anonymization approaches that, to the best of our knowledge, closely align with the use case provider's anonymization procedure based on the descriptions provided. Notably, this could also help the autonomous driving use case provider identify a potentially more suitable anonymization procedure in the future.

The high-level architecture of the PRS is depicted in Figure 4.

The process flow includes:

1. Collection of batches of images and location timestamps
2. An anonymization module which anonymizes the images using an anonymization policy. The anonymization policy specifies which hyperparameters and algorithms are used in the anonymization. For instance, the policy could specify blurring with gaussian blur and the specific hyperparameter that controls the bluing intensity.
3. A privacy module that uses the blurred images and the collected location timestamps to compute a privacy score. The specific method of anonymization is unknown to the privacy module. The privacy policy specifies a default privacy requirement. It can also specify a specific policy for a given location region. For instance, in the vicinity of a hospital or military installation a higher privacy level could be required than on a standard highway. Whenever no specific policy is associated with a specific location the default policy is active.

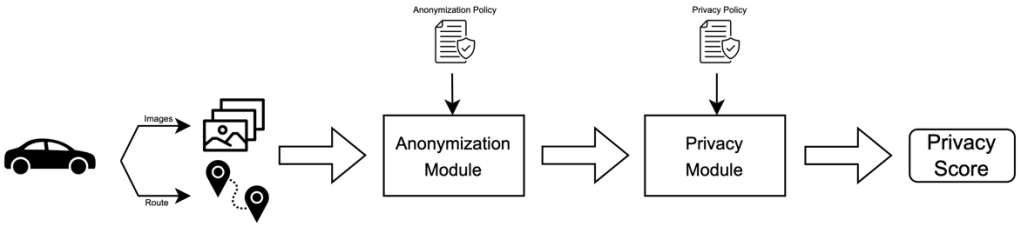

Figure 4 Overview PRS

3.2.2.3 UPCV

In the UPCV, the arrangement needs to anonymously keep track of user behaviour. The recommender is built around the concept of a privacy wallet: Users and items have their own privacy wallets, which contain a set of Volatile Random Numbers (VRN) that *cannot* be traced back to an individual user.

When user views an item and thus expresses interest, a single VRN is copied from the privacy wallet of the user to the privacy wallet of the item and vice versa. Eventually, similar users get same VRNs in their privacy wallets; so do also items that have been accessed by same users. Item recommendations are based on privacy wallet similarities:

When the user wants to see recommended items from the web service, they request the server to send all the items' privacy wallets to them, after which the user's web browser or similar client – the web app may be wrapped as a mobile app – calculates the recommendations and displays them.

Considering a process flow, it includes the following steps:

1) Vendors (e.g. tourist services) have created a listing of their offerings (items), with thumbnail images and descriptions. This listing resides on vendor servers, together with a privacy wallet for each item; a single VRN is created to the privacy wallet.
2) When accessing the recommendation system for the first time, the privacy wallet is automatically created for the user. It is stored in web app's local storage. Symmetrically, a single VRN is created also to the privacy wallet of the user.
   - Over time, also other users are likely to receive this VRN as well, and it also is subject to be deleted from the original privacy wallet.
   - *Hence, the VRN is not associated with any single user*.
3) In the web app, the user can click any item (e.g. tourist attraction) for details; This indicates user interest in the item and initiates VRN exchange between the user and the item over a server API.
4) The list of items on the web all is updated by requesting privacy wallets of individual items, computing on the web all similarities between the items' privacy wallets and the user's privacy wallet, and presenting the items from most to least similarity.
5) Figure 5 UPCV web app, executed on mobile and desktop browsers illustrates an example of web app screen on a desktop browser; the same web app[1] runs in mobile platforms as well.
6) UPCV technology has also been evaluated for analyzing shopping baskets in Retail use case. The ultimate goal is to find out, if the privacy-preserving recommender could be used for business intelligence in a distributed market potentially containing sensitive data.

[1] UPCV Recommender. Accessed in January 31st, 2025 at https://upcv.k8s-cluster.tango.rid-intrasoft.eu/

Figure 5 UPCV web app, executed on mobile and desktop browsers

3.2.3 Features implemented

3.2.3.1 TSM

The TSM features were adapted to its usage as a backend component. Major adaptations were

1. Introduction of user (DID) related scoring. This includes user related configurations and data structures
2. Implementing a timeseries database for attribute data. This enables future consideration of historical behaviour for the trustworthiness scoring. The time window is considered in the configuration.
3. Definition of the backend APIs.

The API is illustrated in Figure 6.

Figure 6 - TSM backend API

3.2.3.2 PRS

The PRS features are implemented in python and adapted to be compatible with the TSM such that the output of the PRS can be stored in a time series database.

3.2.3.3 UPCV

UPCV is implemented as a whole, to the extent that it can be used in Smart Hospitality piloting.

3.2.4 Advancement of components

3.2.4.1 TSM

The TSM is available as a backend component. Integration into the PEP is ongoing. Validation as a local use case component in the Automotive use case is planned. Deployment depends on data availability for the contributing PSR. Design of the application in the Manufacturing use case in ongoing. The TSM is available as dockerized components.

3.2.4.2 PRS

The PRS is in its final development stages to couple it with the TSM. Since, we are still awaiting real data from the autonomous driving partner we anticipate some fine-tuning tasks on the data compared to open-source data from the PP4AV dataset we are currently using. This may involve adjusting anonymization algorithms and adding certain functionalities to deal with detection issues of certain attributes to compute the privacy score.

3.2.4.3 UPCV

After D3.1, UPCV has got a distributed implementation, as promised in the DoA. Regarding ‘Retail’ use case, the acquired shopping basket data set data has been processed through a desktop UPCV implementation; the analysis will continue once certain details of the data set have been received.

3.2.5 Support for pilots

3.2.5.1 TSM

The TSM’s integration into the PEP is ongoing. With its availability the use cases can leverage the Trustworthiness scoring. Moreover, it is available for local integration and deployment by the use cases themselves for generating and utilizing Trustworthiness scores.

3.2.5.2 PRS

Since the primary objective of the PRS is to support the TSM, its functionality aligns with that of the TSM. However, as previously noted, the secondary use case can assist autonomous driving providers in

exploring innovative approaches to anonymization—going beyond the basic methods implemented thus far

3.2.5.3 UPCV

The distributed web app implementation of UPCV is ready for Smart Hospitality piloting. Because of distributed nature of computation, and lightweightness of the technology as a whole, neither response times nor scalability is limiting its use.

The study of applying UPCV recommendation technology into business intelligence (Retail use case) is ongoing.

3.2.6 Future Investment

3.2.6.1 TSM

The TSM has been prepared for utilizing historical data. However, its use has not been pursued yet. Ways to evaluate a historical behaviour extending the scoring may be assessed with the use cases.

3.2.6.2 PRS

The PRS has been designed in a modular fashion, enabling the seamless integration of new privacy requirements. A future research project will explore the assumption that existing privacy regulations may not adequately protect individuals' privacy in certain scenarios, addressing this concern through further investigation.

3.2.6.3 UPCV

At VTT, there have been some ideas and thoughts to further improve UPCV algorithm, to be validated and potentially reported in an academic publication. Also, the ‘Retail’ study will be concluded.

3.2.7 Documentation

3.2.7.1 TSM

The TSM has been deployed in the TANGO development and test environments. The API documentation is available there, e.g. in the development environment see: backend - Swagger UI.

3.2.7.2 PRS

The PRS has not yet been deployed in the TANGO development and test environments. The API documentation will be available there as soon the final integration stages are finished.

3.2.8 Conclusion

The three components of task T3.2 considerably progressed towards availability for the use cases in the recent period. Each of them provides its functionality and is ready for integration while further alignment with the use cases may be needed for final validation.

3.3 Confidentiality and Privacy by Design [T3.3]

3.3.1 Introduction

Ensuring **confidentiality and privacy by design** is fundamental in data access management, particularly in distributed and secure environments. These principles aim to protect sensitive information by embedding privacy and security measures directly into system architectures rather than treating them as add-ons. By leveraging fine-grained access control mechanisms and cryptographic techniques, data access can be regulated in a way that minimizes exposure and ensures compliance with privacy requirements.

In this context, the process of granting a consumer access to requested data proposed in the solution follows two parallel sub-processes designed to uphold these principles which are integrated in the data space connector:

First, **user consent and access control** are enforced using the **XACML framework**, which includes the Policy Enforcement Point (PEP), Policy Decision Point (PDP), and Policy Information Point (PIP). These components work in conjunction with Verifiable Credentials, including Verifiable Presentations and zero-knowledge proofs from the SSI module, to determine whether a consumer is authorized to access a specific asset.

Second, **data encryption** based on attribute-based policies (**CP-ABE**) allows producers to encrypt data according to predefined policies. Once access is granted by the Policy Enforcement Point, the consumer must decrypt the data using a key linked to their identity attributes. This ensures that decryption is only possible if the consumer's attributes match the **sticky policy**, maintaining strict access control and data confidentiality.

3.3.2 Description of the component

3.3.2.1 User Consent and Access Control

The **Policy Enforcement Point (PEP)** and **Policy Decision Point (PDP)** are critical components in enforcing access control policies.

What we refer to as usage control in this context refers to **applying access policies** that determine how data can be accessed (e.g., read-only or write-only) before it is shared. However, once data has been accessed or shared, these policies do not regulate its further use. The primary function of these policies is to govern access to data and services based on predefined rules, ensuring that only authorized entities can retrieve or modify data as permitted.

When a requester intends to perform a specific action on a resource or service, they must first send a **request to the PEP**, which then forwards it to the **PDP for evaluation**. This request contains details regarding the requested action, the target service, and relevant identity information of the requester.

Most of the requester's identity information is provided in the form of an **access token issued by the SSI Verifier**. The key details contained in the access token include:

- Expiration: Date and time when the access token becomes invalid.
- Issuer: Entity that issued the access token.
- Key ID: ID of the key needed for verifying access token's signature
- Subject: Represents the user or entity to whom the token has been issued.
- Verifiable Credential: This is a credential that prove claims about the identity of the subject.
    - Credential Subject: The entity to which the credential applies. The VC contains claims describing the subject, such as an ID, email, role, or name.

Using the information contained in the **access token**, the **Policy Decision Point (PDP)** verifies the requester's identity against the policies retrieved from the **Policy Administration Point (PAP).** These policies define the access rules for the requested resource or service. The PDP also uses additional information retrieved from the **Policy Information Point (PIP)** that helps it to make the decision. This additional information is a Trust Score (coming from a Trust Score Manager) and a Behavioural Score (coming from User Continuous Behavioural Authentication) that the requester receives. These components will be described later in this section. The policies have some fields that evaluate what is the minimum trust score, for example, that a requester needs to have whether he wants to access a resource.

If the request is **authorized** and complies with the proper policies, the PDP issues a **Capability Token**, which is subsequently returned to the requester. This token serves as **proof of authorization**, granting the requester the ability to perform the requested action within a limited timeframe.

To utilize the **Capability Token**, the requester submits a **second request to the PEP**, including the Capability Token, the action to be performed, the target resource, and any additional parameters required for execution. Upon receiving the request, the **PEP verifies the token's validity** by ensuring it is correctly signed, unexpired, and aligned with the requested action and resource.

Once the verification is successful, the **PEP establishes a connection with the relevant API** to grant the requester access to the desired service. Depending on the nature of the service and action, this interaction may involve various communication methods, such as query parameters or JSON payloads in a POST request.

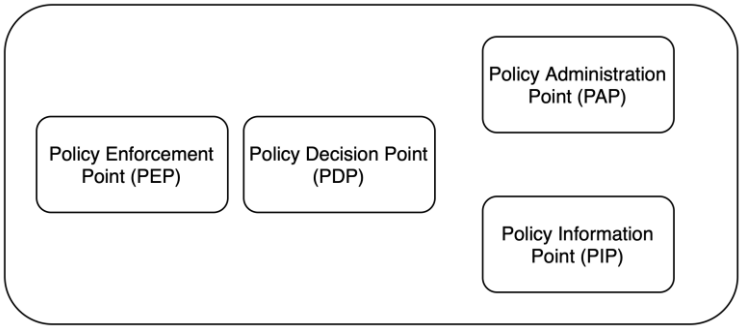

Figure 7 User Consent and Access Control components

3.3.2.2 Standard API for Access Management

In this context, a **standard API** has been established to manage access to various services. Pilots are required to adapt and **configure this API** according to their specific resources (database) and the interaction process necessary for their services.

To ensure seamless integration, services should be designed to accept requests from this API, which may include query parameters in a request or JSON payloads in a POST request.

By aligning their implementations with this standardized approach, pilots can ensure compatibility and efficient access management across different services.

3.3.2.3 Defining and Registering New Policies in the PAP component

The system enables requesters to define and **register new policies** within the PEP/PDP component. This process consists of two key steps:

1. **Requesting Access Using a Long-Lived JWT**

- Endpoint: <pep-component>/api/auth
- The requester must first obtain a short-lived JWT, which will be used to submit new policies.
- To generate this token, the requester sends a request containing:
  - Their identifier (used as the JWT subject)
  - The access scope (defining the level of granted permissions)
  - An authorization header containing a previously issued long-lived JWT

2. **Registering New Policies**

- Endpoint: <pep-component>/api/new-policy
- Once the short-lived JWT has been acquired, the requester can register new policies by sending a request that includes:
  - The policy definition
  - The resource to which the policy applies
  - The short-lived JWT obtained in the previous step for authentication

This mechanism ensures that **only authorized entities** can define and register new access policies while maintaining security and compliance with established authorization frameworks. As a proof of concept, this approach will be tested by integrating with the PAT (Privacy Assurance Tool, T5.4) component to incorporate policies defined in the user consent process. This integration aims to streamline policy management and enhance alignment between access control mechanisms and user-defined consent policies.

3.3.2.4 Integration with TSM and UCBA (context-aware policies)

In the integration with TSM (**Trustworthiness Scoring Module**, T3.2) and UCBA (**User Continuous Behavioural Authentication**, T4.3), the information provided by these components contributes to the decision-making process within the PDP (Policy Decision Point).

This process extends beyond simply evaluating predefined policies to determine whether a user can access a service. It also considers two key factors:

- **Trustworthiness Score** – Provided by the TSM, this score is calculated based on specific user-related factors. The PIP (Policy Information Point) manages the interaction with the TSM, ensuring that when the PDP evaluates a request, it retrieves the user's trustworthiness score to assist in policy decision-making.
- **Behavioural Score** – This score is also used by the PDP to enhance access control decisions. It is retrieved through the PIP, which handles communication with the UCBA component.

By incorporating these **dynamic scoring mechanisms**, the system enhances security and adaptability, making access control decisions more **context-aware** and **risk-sensitive**.

3.3.2.5 Sticky Policies

Sticky policies remain attached to the data as it moves across different systems, platforms and organizations, ensuring that predefined restrictions and permissions are enforced throughout the data lifecycle. The implementation of sticky policies is achieved through Ciphertext-Policy Attribute-Based Encryption (**CP-ABE**), encrypting data based on **attribute-based access polices** defined by the data owner. Only users whose attributes (provided by SSI **Verifiable Credentials**) satisfy the encryption policy can successfully decrypt and access the protected data.

The following figure illustrates the CP-ABE submodules that constitute the sticky policies components. A detailed flow description is provided in Section 3.3.3.

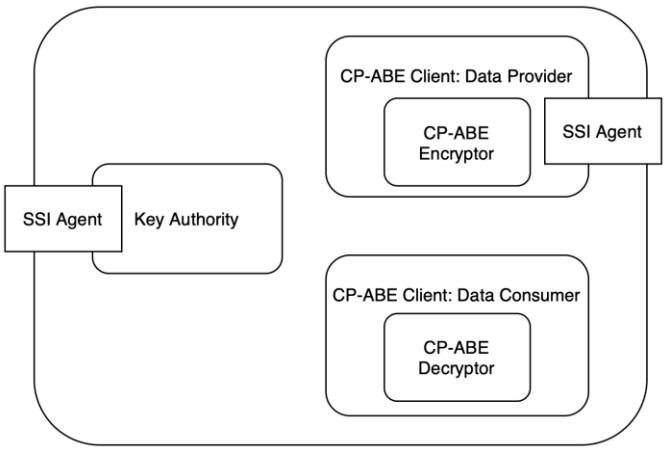

Figure 8. CP-ABE submodules

3.3.2.6 Hybrid encryption with SEDSS

In scenarios where both encryption mechanisms are deployed, **SEDSS** (Self-Encryption, T3.4) and **CP-ABE** (sticky policies) operate in a layered manner to ensure end-to-end data protection. The SEDSS

module encrypts data when it is published by the data provider, safeguarding its confidentiality. Simultaneously, CP-ABE is used to encrypt the SEDSS decryption key along with the encrypted data, ensuring that only authorized recipients possessing the required attributes can retrieve the key and decrypt the underlying information. This dual-layer encryption strategy enhances security by preventing unauthorized access to sensitive data, even in cases where storage or transmission environments are compromised.

Workflow was defined in D6.1., Annex A.5. If any final modification is considered, it will be specified in D6.2.

3.3.2.7 Interaction PEP & Sticky Policies

As outlined earlier, granting a consumer access to a requested data resource or service relies on the two described mechanisms: Policy Enforcement Point (PEP) authorization and Sticky Policies through Ciphertext-Policy Attribute-Based Encryption (CP-ABE). These mechanisms complement each other to ensure security, confidentiality, and privacy by design throughout the data-sharing process.

In the context of different pilot implementations, the interaction between these mechanisms follows a consistent process across two primary use cases:

- **File Upload** with Policy Enforcement and Encryption: before a user uploads a file, the Policy Enforcement Point (PEP) verifies their authorization. If authorized, the file is encrypted in accordance with predefined attribute-based policies, ensuring that only users with the appropriate attributes will be able to decrypt it in the future.
- **File Download** with Policy Enforcement and Decryption: when a user attempts to download a file, the PEP checks their authorization to access the resource. If access is granted, the file is decrypted based on the attribute policies set during encryption, ensuring that only users meeting the required criteria can successfully retrieve the content.

By integrating PEP authorization with CP-ABE encryption, this approach enforces strict access controls while embedding security directly into the data, regardless of where it is stored or shared.

3.3.3 Internal Architecture

3.3.3.1 User Consent and Access Control

The internal architecture of the PEP/PDP component consists of several subcomponents:

- **Policy Enforcement Point (PEP):** This is the component that intercepts the different requests and acts in consequence. Depending on the action that the user wants to make, the PEP is the main entity that receives the requests in the following situations:
    - When a user wants to get a Capability Token to get a resource. Here the PEP forwards the request to the PDP to evaluate it and when PDP issues the Capability Token if everything is correct, the PEP answers the user request with the capability token.
    - When a user wants to access a resource using a Capability Token. The PEP just verifies the Capability Token and let the user access the resource whether the information is ok.
- **Policy Decision Point (PDP):** The PDP receives the request from the PEP and evaluates it. The PDP executes the process to verify if the user can access the resource. It just executes the process, the different details that the PDP needs for making the decision come from other components. The policies that are associated with the resource and the user needs to comply to access it come from the PAP and the trustworthiness score and the behavioural score from the PIP.
- **Policy Administration Point (PAP):** This component is used:
    - When the PDP wants to retrieve the policies that regulate the access to a resource or services.

- When a user wants to register new policies. The user sends the proper requests to the PEP, and it oversees evaluating the policy if it has any syntax error and the store it into the policy storage associated with the resource or service that it regulates.

- **Policy Information Point (PIP):** This component contains additional information that the PDP requires for performing the decision-making process. This additional information is retrieved by the PIP from the corresponding platforms (the trustworthiness score from the TSM and the behavioural score from UCBA) and is sent to the PDP.

The next workflow represents the interaction between the different elements of the PEP/PDP Component's internal architecture in the main workflow of the component, a user requesting to access a resource.

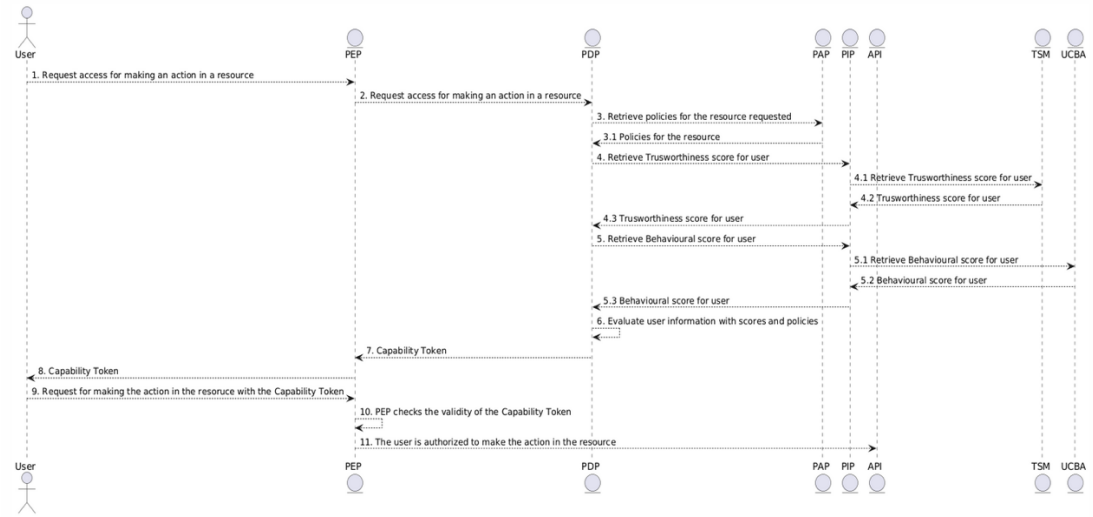

Figure 9. Internal Workflow of the PEP/PDP Component for User Access Requests

3.3.3.2 Sticky Policies

As defined in D3.1., the **ABE Toolset Prototype** enables secure and privacy-preserving data exchange using Ciphertext-Policy Attribute-Based Encryption (CP-ABE). It ensures that only users or devices with specific attributes can decrypt data, making it crucial for TANGO pilot scenarios requiring granular access control.

The toolset consists of three submodules:

- Key Generator (Issuer Authority): Generates master public and private keys for encryption and decryption, issuing CP-ABE decryption keys based on user attributes.
- Encryptor: Used by data providers to encrypt data according to an access policy, ensuring only authorized users can decrypt it.
- Decryptor: Used by consumers to decrypt data if their attributes satisfy the encryption policy, obtaining the necessary decryption key from the Issuer Authority.

Above this ABE Toolset, which provides the cryptographic tools, the sticky policies workflow takes place. Three main steps can be identified:

- **Initialization**:
  The initialization phase sets up the cryptographic infrastructure necessary for secure data encryption and decryption. It involves the **creation and distribution of encryption and decryption** keys based on user attributes.
  - The **Key Authority Server** Module (Issuer Authority) initializes by generating a master public key and a master secret key.

- The master public key is published in a Public Registry or made accessible through other key management mechanisms.
- When a consumer (user or device) requests access to encrypted data, it must first **obtain a decryption key** corresponding to its attributes.
- The consumer generates a Verifiable Presentation (VP) through its **SSI Agent**, proving its attributes.
- The Key Authority Server **verifies the VP** and, if valid, issues a CP-ABE decryption key containing attributes that match the consumer's identity.
- The **decryption key is securely stored** for future use, enabling the consumer to decrypt data later without needing to request keys again.
- This phase ensures that only authenticated and authorized entities receive decryption keys, maintaining strict access control.

- **Encryption**:
  The encryption phase is executed by the data provider (producer), ensuring that access control is enforced directly within the encrypted data.
  - Before sharing data, the producer **defines an access policy**, specifying which attributes are required to decrypt the data.
  - The **data** is then **encrypted using the CP-ABE Encryptor** with the master public key and the defined policy.
  - The encryption process ensures that only consumers whose attributes satisfy the policy will be able to decrypt the data.
  - The encrypted data is then uploaded to the storage, along with metadata that may include policy information, a reference to the encryption key, and a data identifier.

  Since the access control rules are embedded in the encryption process, the data remains protected regardless of where it is stored or transmitted. Even if an unauthorized party gains access to the encrypted data, they cannot decrypt it without possessing the required attributes and a valid decryption key.

- **Decryption**:
  The decryption phase ensures that only authorized consumers can access the plaintext data.

  - When a consumer requests encrypted data, it must first pass through the Policy Enforcement Point (PEP), which verifies access authorization based on predefined policies.
  - If authorization is granted, the encrypted data is provided to the consumer.
  - The consumer then uses the **CP-ABE Decryptor** to attempt decryption.
  - If the consumer possesses a **valid decryption key** that satisfies the encryption policy, the data is successfully decrypted.
  - If the necessary decryption keys are not available, the consumer must request them again from the **Key Authority**, following the verification process outlined in the initialization phase.

  This phase ensures that data access remains strictly controlled, preventing unauthorized decryption while allowing legitimate users to retrieve the information securely.

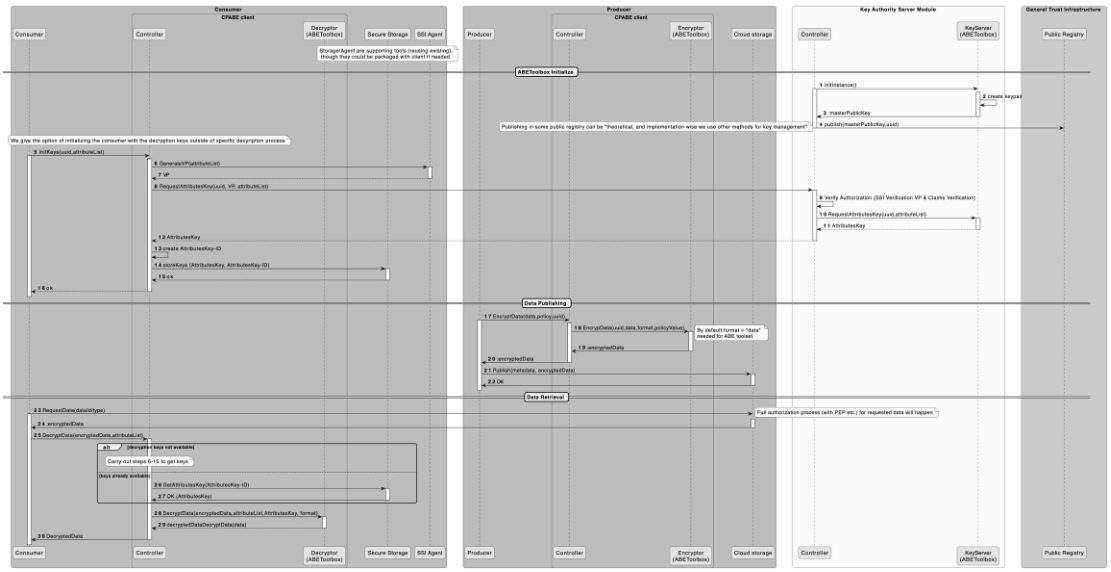

Figure 10. CP-ABE workflow

This workflow ensures a privacy-preserving, attribute-based access control mechanism, where encryption and authorization policies are enforced directly within the data itself. By embedding access control within the encryption process, the mechanism remains persistent and self-enforcing, independent of where the data is stored or shared, making it particularly suitable for distributed environments like TANGO.

3.3.4 Features implemented

Table 3 Features implemented in T3.3

Component	Feature	Status	Comments
PEP	Standard API for storage	Done	Standard API to be adapted per pilot to their own storage DB.
PEP	Integration with TSM	Done	PEP can incorporate policies based on trust scores by leveraging the functionality provided by the PIP.
PEP	Integration with PAT	Ongoing	PAT will have the capability to introduce new policies into PEP to regulate access to additional services. Additionally, access to the PAT service itself will be accessed through the PEP authorization workflow.
PEP	Integration with UCBA	Ongoing	Testing API to integrate scores via PIP component.
PEP	Testing	Done	Functional and integration testing implemented.
PEP	Deployment	Done	Deployed within the TANGO cluster and tested.
Sticky Policies	API design	Done	API workflow functionalities fully defined.

Component	Feature	Status	Comments
Sticky Policies	Vault secure storage	Ongoing	Integration with Vault for secure key management.
Sticky Policies	Integration with SEDSS	Ongoing	Integration with SEDSS encryption/decryption API for hybrid mode to be used in FMAKE pilot.
Sticky Policies	Testing	Ongoing	Functional testing implemented. Still some integration testing missing.
Sticky Policies	Deployment	Ongoing	Deployed demo version. Working in deployment in three different instances (infrastructure, provider, consumer)

3.3.5 Advancement of components

Significant progress has been made in the development and deployment of key components within the TANGO framework. The PEP (Policy Enforcement Point) has successfully implemented a standard API for storage, adaptable per pilot to their respective databases, and has been fully integrated with TSM, enabling trust-based policy enforcement via the PIP (Policy Information Point). Functional and integration testing have been completed, and PEP has been successfully deployed within the TANGO cluster. Ongoing efforts focus on integrating PAT, which will enable dynamic policy updates and regulated access to additional services, as well as UCBA, which will enhance policy-based decision-making. Still evaluating possible integration with DCBA (device behavior) for demonstration purposes.

In Sticky Policies, the API design has been finalized, providing a well-defined workflow for enforcement. However, key integrations remain in progress, particularly with SEDSS encryption/decryption APIs, crucial for hybrid encryption in the FMAKE pilot, and final adjustments with Vault secure storage, which will enhance secure key management. While functional testing has been carried out, additional integration testing is still needed, along with the full deployment across infrastructure, provider, and consumer environments. These advancements ensure that both PEP and Sticky Policies evolve toward a more robust and secure data-sharing ecosystem.

3.3.6 Support for pilots

As specified in D3.1., each feature added to the components is designed to be independent of any specific pilot, meaning no particular pilot is taken into account. Instead, features are integrated into the connector's generic architecture and workflows. Stated that:

PEP component will support pilots:

Smart Hospitality	Authorization to access services: PAT tool, EDAE.
Autonomous Vehicles	Authorization to request car service, and maps from the car. Authorization for B2B clients to access appropriate maps (together with CP-ABE)
Smart Manufacturing FMAKE	Authorization to upload printing designs (together with CP-ABE) Authorization to download printing designs according to the defined policies (together with CP-ABE)
Smart Manufacturing RIASTONE	Access to RIAS system

Retail	Authorization to access EDAE service

CP-ABE will support:

Autonomous Vehicles	Authorization for B2B clients to access appropriate maps (together with PEP)
Smart Manufacturing FMAKE	Authorization to upload printing designs (together with CP-ABE) Authorization to download printing designs according to the defined policies (together with CP-ABE) Hybrid mode CP+ABE & SEDSS

3.3.7 Future Investment

Future investment should prioritize advancing the ongoing functionalities of PEP and Sticky Policies to ensure full integration and seamless operation. In PEP, the focus remains on completing the integration with PAT, enabling dynamic policy updates and enforcing access control for additional services, as well as progressing with the UCBA integration to enhance policy-based decision-making. Meanwhile, Sticky Policies require further development in Vault secure storage integration to strengthen key management security and completing the integration with SEDSS to support hybrid encryption for the FMAKE pilot. Additionally, refining integration testing is essential to validate these functionalities across different deployment environments. By addressing these ongoing efforts, the system will achieve enhanced security, adaptability, and interoperability within the TANGO framework.

3.3.8 Conclusion

The integration of PEP authorization and CP-ABE encryption ensures a secure and privacy-preserving approach to data access management. By combining policy-based access control with attribute-based encryption, the solution enforces strict authorization while embedding security directly into the data itself. This dual-layered mechanism not only protects sensitive information from unauthorized access but also ensures that access control policies remain effective even after data is shared. By leveraging these techniques within the data space connector, the system upholds confidentiality, integrity, and privacy by design, making it well-suited for secure and distributed environments.

3.4 Self-encryption and Decryption Techniques with Multi-Factor Information Recovery Mechanisms [T3.4]

3.4.1 Introduction

This task implements self-encryption and decryption techniques with multi-factor information recovery mechanisms, a secured information encryption and recovery mechanism to meet the demands for confidential information storage and distribution in TANGO use cases. Compared to the previous deliverable (D3.1), this deliverable highlights the current progress on this task, especially the updating on architecture design, the API design for TANGO integration and. To make the deliverable self-inclusive, we also briefly include an overview of each module in this component. Specifically, we implement three components to fulfil these functionality requirements: self-encryption module, self-decryption module, multi-factor information recovery mechanism. Meanwhile, we implement the API regarding each functional component in order to integrate them into TANGO framework. We also tested our component regarding the computational costs, including the time cost and memory consumption, on various file types and various data size. Ultimately, our encryption and decryption software can not only execute encryption and decryption with information sharing mechanism to flexibly produce key(s) for multiple parties, but also scalable to different data formats and data sizes. The detailed contribution of task 3.4 is as follows:

- Encryption and Decryption: Encryption converts a document into an unreadable format (ciphertext) to prevent unauthorized access, using a key for the process. Decryption reverses this, allowing authorized users to recover the original document using the same key. Self-encryption is a specialized method where the document's contents are used as part of the encryption and decryption process, eliminating the need for an external key.
- Multi-Factor Information Recovery Mechanism: This mechanism enhances security by dividing the decryption key into multiple shares distributed among several individuals. To decrypt the document, a minimum number of key shares must be combined, reducing the risk of malicious access, as it requires collaboration among a group.
- API design for TANGO integration: The API provides secure encryption, decryption, and multi-factor key management. It includes endpoints for self-encryption of documents, decryption of ciphertext, splitting keys into shares for distribution, and recovering keys using a threshold of shares.

3.4.2 Description of the component

Task 3.4 implemented a secure encryption and recovery mechanism to ensure confidential information storage and distribution within TANGO use cases. Two key components are developed: the self-encryption and decryption module and the multi-factor information recovery mechanism: (1). Self-Encryption and Decryption, which encrypts documents without user-specified keys, instead deriving encryption keys from the document's own content. This process ensures that only those possessing the original file can decrypt it, enhancing security and eliminating the need for external key management; (2). Multi-Factor Information Recovery: To further strengthen security, the encryption key is divided into multiple shares and distributed among different entities. Decryption requires a predefined threshold of key shares, mitigating risks by preventing unauthorized access through a single compromised key holder.

3.4.3 Internal Architecture

Figure 11 briefly shows the system architecture of the self-encryption and decryption with multi-factor information recovery system. As shown in ‘main components’, the whole system consists of four sub-modules, encryption, decryption, multi-factor information sharing and multi-factor information recovery module (the former two sub-modules consist of encryption/decryption module, while the latter two sub-modules consist of multi-factor information recovery module). Based on the architecture design, the encryption and decryption module are both connected to the data connector of the TANGO framework. Then the multi-factor information sharing, and recovery modules are callable by the encryption and decryption module respectively. After the encryption process, the encrypted information and shared keys will be returned to the user-specified location (such as a remote file system, a remote database or data storage in TANGO framework). In the decryption process, a decryption request which contains the key (pieces) for data recovery and location to obtain encrypted files will be submitted to our framework. Then, multi-factor information recovery mechanism will be called to recover the original key for data encryption from the user-provided key pieces. Finally, the decryption process will be executed based on the recovered key and encrypted data chunks. Once the decryption is successfully done, the plain text will be returned to user via the data connector.

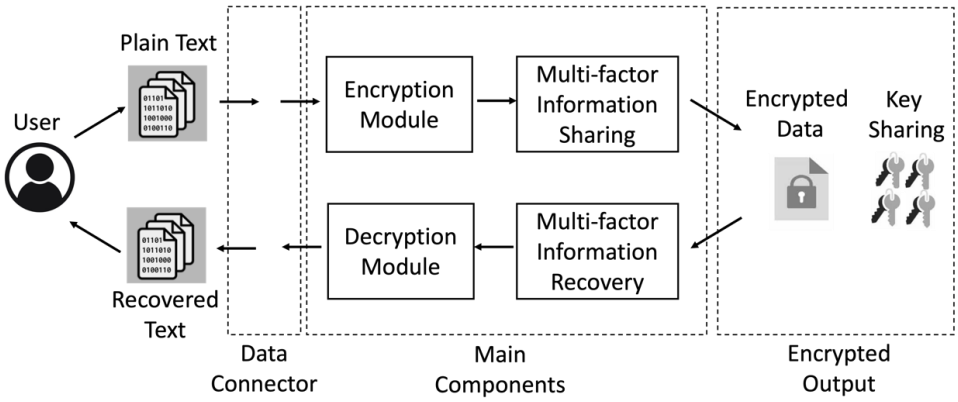

Figure 11 The overall architecture of SEDSS.

3.4.4 Features implemented

In SEDSS we have two three main functions, encryption, decryption, and multi-factor information sharing. We implemented two APIs, namely encryption and decryption. Besides, we integrate the information sharing and recovery functions into the encryption and decryption APIs respectively, so we don't explicitly mention the multi-factor information sharing/recovery mechanism for clarity. To highlight the progress we made after D3.1, we present a table in Table 4 to summary the features implemented in the previous version and the progress so far.

Table 4 A summary table to summarize the features of previous version in the last deliverable and the current status.

Item	Progress in D3.1	Status in D3.2
Encryption	Implemented.	Updated functions based on pilot demands. Encryption API is operational.
Decryption	Implemented.	Updated functions based on pilot demands. Encryption API is operational.
Multi-factor Information Sharing	Implemented, a standalone API is designed.	Updated functions based on pilot demands. Decryption API is operational.
APIs	Initially designed but not fully tested or deployed.	APIs are tested with toy data. APIs are deployed to TANGO platform and is ready for use.
Testing	Not implemented.	Functional testing includes data of various size and formats, as well as real data from usecase. An UI-based functional testing toolkit is developed for pilot users.
Deployment	Demo codebase deployed in TANGO server. APIs are not available.	An executable docker image containing all functions and features, including encryption and decryption APIs, is deployed in TANGO server. The APIs are ready for use.

3.4.5 Advancement of components

Compare to D3.1, SEDSS in this deliverable advance in methodology and API design. This section introduces the design and functionality of the newly designed Encryption API and Decryption API, which are integral to the secure handling of data in the SEDSS framework. These APIs enable users to encrypt and decrypt files efficiently while incorporating advanced features such as multi-chunk data encryption, multi-factor key generation, and secure key recovery mechanisms. We showcase the encryption and decryption API design as well as technical explanations in Section 0 and 0 respectively.

3.4.6 Support for pilots

This section presents efforts to support pilots in real-world scenarios. A functional testing toolkit is introduced for pilot users without prior experience in programming or API development. The efficiency of SEDSS is then evaluated in encryption and decryption, with results reported across various file sizes and types. To accommodate diverse data formats and encodings, SEDSS is customized with a unified encoding method. Details are provided below.

3.4.6.1 Functional Testing Workflow and Toolkit for FMAKE

A UI-based functional testing toolkit was developed to evaluate the encryption and decryption process within SEDSS, a self-encryption and decryption algorithm. This testing toolkit provides an intuitive graphical interface that allows users to interact with SEDSS. The primary motivation for designing this toolkit was to facilitate accessibility and usability for pilot users who may not have technical expertise in encryption algorithms or API-based interactions. Instead of relying on command-line operations or complex coding, users can perform encryption and decryption tasks with simple interactions, making it easier to assess the algorithm's efficiency and effectiveness in real-world applications. By integrating core encryption and decryption functionalities, it enables a user-friendly testing environment.

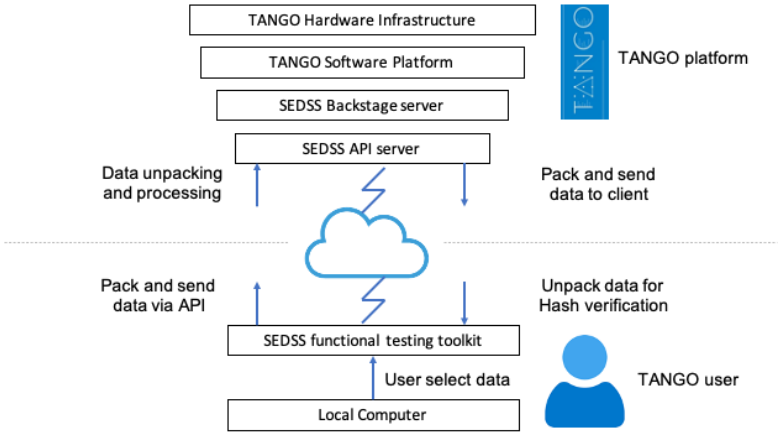

Figure 12 The diagram of functional testing workflow between user's side and SEDSS service of TANGO server

There are three steps in the SEDSS functional testing workflow, (1) encryption testing workflow, (2) decryption workflow, and (3) hash verification. We briefly showcase the relationship of these steps in Figure 12, and the design of functional testing user-interface in Figure 13, Figure 14 and Figure 15 respectively.

Figure 13 The encryption step for functional testing.

In the first step, the user selects a file from the local disk and initiates the encryption process using the SEDSS functional testing toolkit. The toolkit encrypts the file and generates an encryption key, both of which are then returned to the user. This step evaluates the efficiency and correctness of the encryption process across different file sizes and formats. The encryption API, as shown in Figure 13, is specified by the SEDSS service provider as a standalone service in TANGO platform.

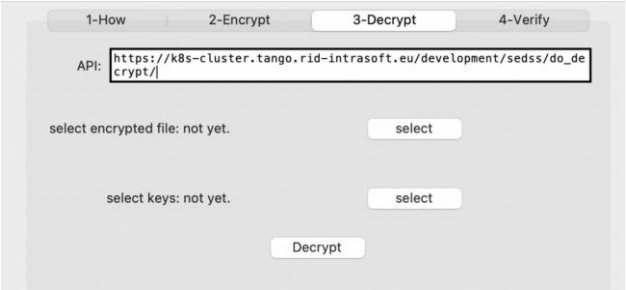

Figure 14 The decryption step for functional testing.

In the second step, the encrypted file and its corresponding encryption key are used in this step to test the decryption process. The user selects both inputs through the toolkit, which then executes the decryption algorithm to restore the original file. The success of this step confirms the accuracy and reliability of the decryption mechanism. The decryption API, as shown in Figure 14, is specified by the SEDSS service provider in TANGO platform as well.

Figure 15 The hash verification step for functional testing.

To ensure the integrity of the encryption-decryption process, a hash verification is executed in the last step. A hash is a fixed-length string generated from input data using a cryptographic hashing algorithm. It acts as a unique digital fingerprint, ensuring data integrity by producing the same output for the same input while making even the slightest change result in a completely different hash. Hash functions are widely used in data verification, encryption, and digital security to prevent unauthorized modifications and ensure authenticity. In the last step, the toolkit calculates and compares the hash values of the original file and the decrypted file. If the hash values are identical, it indicates that the decryption process was successful, and the original data has been perfectly restored. The hash verification will be done in users' local computer.

3.4.6.2 On the Encryption and Decryption Efficiency of SEDSS for Real world Data

We present the analysis of the encryption and decryption times for four typical types of files with varying sizes. The files tested include TXT, CSV, PDF, and MP4 files, and the measurements were taken using an encryption-decryption algorithm in SEDSS. The results comprehensively reflect that SEDSS is capable of dealing with large files with reasonable time cost. The results are presented in Annex. 3.

3.4.6.3 Encryption and Decryption on various Data Format

In this section, we investigate the compatibility of encryption and decryption of SEDSS on various data formats and discuss the time consumption on the same file size with four typical file formats to showcase that SEDSS can process various data formats, indicating that SEDSS is compatible with different usecases. Figure 16 and Figure 17 present the encryption times for 10 MB and 2 GB files across TXT, PDF, CSV, and MP4 formats. Similarly, Figure 18 and Figure 19 illustrate the decryption time for the same set of files.

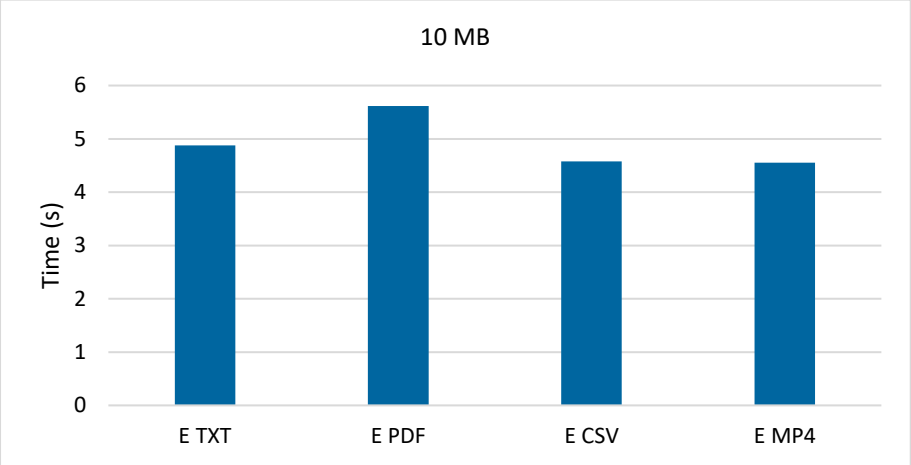

Figure 16 Comparison of encryption times for TXT, PDF, CSV and MP4 files. The measurements are presented in seconds for file sizes of 10 MB.

Figure 16 illustrates the time consumption of various encryption algorithms when applied to a 10 MB file. The results indicate that encryption times are relatively consistent across different algorithms, with only slight variations. This suggests that for smaller files, the choice of encryption algorithm may not significantly impact the overall performance. The efficiency observed in encrypting 10 MB files highlights the algorithms' capability to handle smaller datasets effectively, ensuring minimal processing delay and rapid data protection.

Figure 17 presents the encryption time for a 2 GB file, revealing a stark difference in performance among the algorithms. As file size increases, the time required for encryption escalates considerably, emphasizing the scalability challenges faced by certain algorithms. This figure underscores the importance of selecting an appropriate encryption algorithm when dealing with large datasets, as some algorithms demonstrate markedly better performance in managing substantial data volumes. The comparative analysis of these figures underscores the critical need for balancing security and efficiency, particularly in environments where large data encryption is routine.

We showcase the time cost of decryption process on 10 MB file size against four data formats in Figure 18. The result shows the time consumption of decrypting various data formats on a 10 MB file, with the times closely clustered between 4.2 and 4.9 seconds for different file types such as TXT, PDF, CSV, and MP4. Despite slight variations, the overall download times remain within a narrow range, indicating that the encryption algorithm is performing efficiently across different file types. This efficiency is particularly evident in the relatively quick processing times, suggesting that the algorithm effectively balances security and speed for smaller files, ensuring minimal delay and consistent performance.

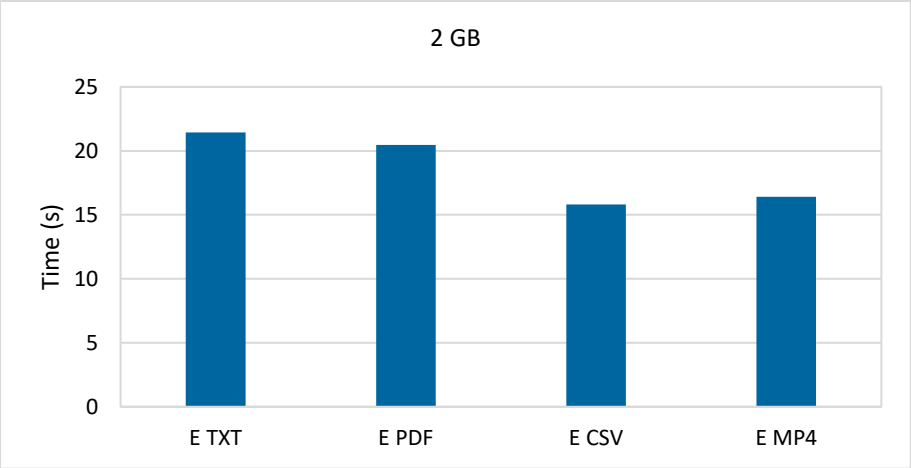

Figure 17 Comparison of encryption times for TXT, PDF, CSV and MP4 files. The measurements are presented in seconds for file sizes of 2 GB.

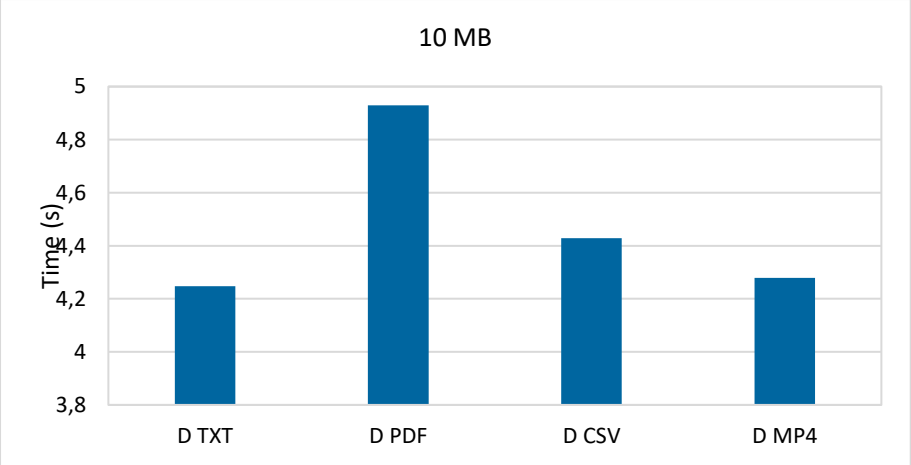

Figure 18 Comparison of decryption times for TXT, PDF, CSV and MP4 files. The measurements are presented in seconds for file sizes of 10 MB.

Figure 19 represents the time consumption of decrypting files with various formats when processing a 2 GB file. Our results show that decryption times for different file types of range between 17 to 22 seconds. Despite the larger file size, the encryption algorithm maintains efficiency across different formats, with only moderate increases in processing time. The consistent performance across diverse file types underscores the algorithm's capability to handle significant data volumes, ensuring reliable and swift encryption for large datasets.

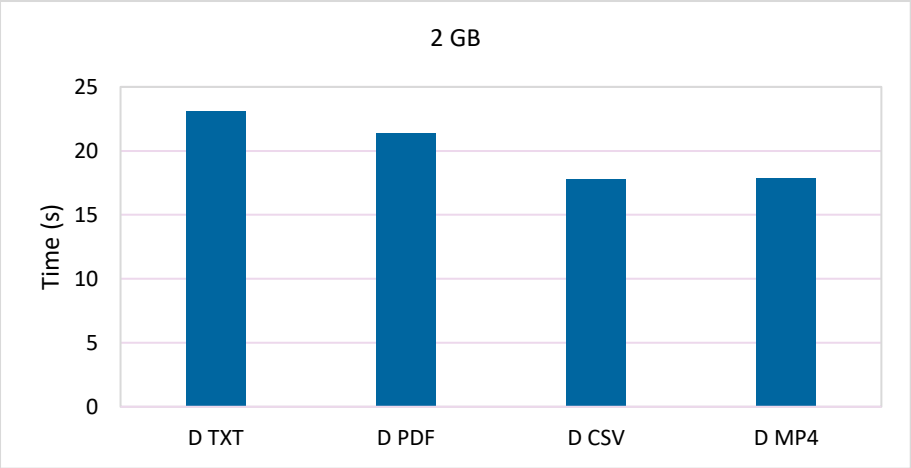

Figure 19 Comparison of decryption times for TXT, PDF, CSV and MP4 files. The measurements are presented in seconds for file sizes of 2 GB.

3.4.7 Future Investment

So far, the development for SEDSS, including the encryption, decryption component, multi-factor information sharing has been implemented. Meanwhile, we have designed APIs deployed to TANGO platform and designed a functional testing workflow to support the pilot users for system testing. In the future, we will do the following tasks. First, our SEDSS APIs will be used for TANGO UI design, in which pilot users with correct credential can login the TANGO system and execute the encryption and decryption functions. Second, we will integrate the SEDSS API into the data transferring process of related usecase, in order to protect the data during the data transmission and local storage. Finally, we will design test cases with simulated and real data from pilot users, to identify potential issues and errors to ensure that the SEDSS meets expected standards in terms of integrity, functionality, performance, reliability, security, and compatibility.

3.4.8 Conclusion

In the task 3.4 SEDSS of the TANGO project, we have implemented self-encryption and decryption modules, as well as a multi-factor information recovery mechanism, ensuring secure data management. The developed APIs facilitate seamless integration into the TANGO framework, providing endpoints for encryption, decryption, key distribution, and key recovery. Comprehensive testing has demonstrated the efficiency and scalability of our software across various file types and sizes. The software's ability to handle different data formats and sizes, combined with secure information sharing mechanisms, ensures robustness and adaptability in real-world scenarios. The support for pilot users through a functional testing toolkit enhances usability and accessibility, enabling non-technical users to effectively utilize the encryption and decryption capabilities of SEDSS. Overall, Task 3.4 has laid a strong foundation for secure and efficient information storage and distribution within the TANGO framework, addressing the critical needs for confidentiality and data protection in diverse applications.

4 Recommendations for secure and privacy-preserving data storage and sharing [T3.5]

4.1 Legal Considerations

In the relevant section of the previous deliverable D3.1, based on the strategies proposed by Hoepman in his book on privacy design strategies,[2] a protocol was created in order to assess the secure and privacy-preserving character of the technologies developed in WP3. An overview of these design recommendations was attached as Annex I in the same deliverable, including the relevant legal provisions as well as examples of practical implementation of these recommendations. Furthermore, an initial assessment of the design recommendations was made an attached as Annex II in the same deliverable, providing compliant and non-compliant features of each technology.

The objective of the present section is to provide a holistic assessment of the technologies developed under WP3, by applying the protocol presented in D3.1 (hereinafter 'WP3 Protocol'), which is valid for both personal and non-personal data in order to guarantee confidentiality. Consequently, each of the subsections of this section offers more detailed view on each of WP3's technologies, exploring all the different elements included therein from a legal perspective.

4.1.1 Blockchain-Based Data Storage and Sharing [T3.1]

4.1.1.1 Identification of personal data stored in FIDES

4.1.1.1.1 Qualification of encrypted data as pseudonymous personal data

An important preliminary consideration to make is that encrypted data is personal data, as data that has been subject to pseudonymisation within the meaning of Article 4, point 5) of 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, and repealing Directive 95/46/EC (General Data Protection Regulation) (hereinafter 'GDPR') [3]. Pseudonymisation is defined by the GDPR as:

> *'[…] the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person'.*

Encryption is a classic example of pseudonymous data, even if there may be other techniques to pseudonymise personal data. Encrypted data is pseudonymous data because, once data has been encrypted, the link to an identity can be re-established by combining the pseudonym (i.e. the encrypted data) and a decryption key. In line with the definition of pseudonymous data, the additional information needed to attribute the data to a specific data subject (i.e. the decryption key) is kept separately and protected by means of technical and organisation measures (which can differ depending on the encryption technology. Without such a key, it is difficult to identify pseudonymised data. However, for those entitled to use the decryption key, re-identification is easily possible, and therefore encrypted data cannot qualify as anonymous data.

The qualification of encrypted data as pseudonymous data means that the GDPR fully applies at any stage of the processes of encryption and decryption. Nevertheless, the qualification of personal data as

[2] Jaap-Henk Hoepman, 'Privacy Design Strategies (The Little Blue Book)' (2022).

[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) 2016 (OJ L 119/1, 4/5/16).

encrypted, and thus pseudonymous, data has legal relevance. Whilst the GDPR fully applies to both pseudonymous and non-pseudonymous data, with no difference between these two categories in terms of applicable provisions, the fact that the processed personal data is pseudonymous can be a determinant factor in assessing whether the data controller has complied with the GDPR. In particular, the controller is required in multiple instances under the GDPR to implement appropriate technical and organisational measures, in order to protect the personal data and implement data-protection principles, and the GDPR explicitly recognises that pseudonymisation could be one of such appropriate measures. As a consequence, applying encryption techniques that render the data pseudonymous can facilitate compliance with certain provisions of the GDPR, for both the controller and the processor. These provisions are as follows:

- Article 6(4) GDPR on processing for a purpose other than that for which the personal data have been collected provides that, in assessing whether processing for another purpose is compatible with the original purpose, the controller shall take into account, *inter alia*, the existence of appropriate safeguards that may include encryption and pseudonymisation.
- Article 25 GDPR imposes on the controller the requirement to ensure data protection by design and by default, by implementing appropriate technical and organisational measures that could, among others, consist of pseudonymisation of the personal data.
- Article 32(1) GDPR requires the controller and the processor to implement measures that ensure a level of security appropriate to the risk including, inter alia, the pseudonymisation and encryption of personal data.
- Article 89(1) GDPR lists pseudonymisation as an appropriate safeguard to implement when processing personal data for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.

In conclusion, while there is no difference between encrypted and non-encrypted data with regard to the applicability of the GDPR, the fact that personal data is encrypted, and thus pseudonymous, shall be taken into account to assess whether a processing operation of such data is in compliance with certain provisions of the GDPR.

4.1.1.1.2 Qualification of individual data chunks separately stored in DRILL as personal data

The data field to be stored in DRILL is broken into separately unintelligible pieces and distributed across DRILL peers in a manner that none of the peers stores all of the data chunks. The bank that onboarded the customer is the only entity in possession of all the encoded data chunks and of the key to decrypt them.

The data chunks separately stored across the peers appear to fall into the definition of pseudonymous data. As a matter of fact, the fact that they are separately unintelligible does not render them anonymous data, given that the data subject may still be re-identified by combining the different data chunks and decrypting them. In particular, at least one data controller (i.e. the bank that onboarded the customer) is at all times able to identify the data subjects, but also the peers or any third party may be able to do so if they combine the data chunks. The process of encrypting the data field and splitting it across peers is clearly not an irreversible process, and thus leads to the creation of pseudonymous data and not anonymous data. As a consequence, the GDPR fully applies to any data stored in DRILL.

4.1.1.1.3 Personal data in plaintext or encrypted (non-hashed) form

In both the banking and public admin pilot, there is personal data to be processed in the platform. In particular, there is information on the names, ID or other identifiers of natural persons, and pictures of the natural persons (e.g. scans of passport pictures).

In processes for KYC and visa checks, the identification of data subjects is one of the purposes for the personal data processing. In this regard, it is essential that the data controller assesses whether there is a processing of biometric data taking place. Biometric data is defined by Article 4(14) of the GDPR as

'personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data'.

Processing of photographs of natural persons does not necessarily entail the processing of biometric data. As stated in **Recital 51** of the GDPR,

'the processing of photographs should not systematically be considered to be processing of special categories of personal data as they are covered by the definition of biometric data only when processed through a specific technical means allowing the unique identification or authentication of a natural person'.

Therefore, it must be considered if the means used for the processing of photographs lead to a processing of biometric data. The EDPB clarified that '*to qualify as biometric data as defined in the GDPR, processing of raw data, such as the physical, physiological or behavioural characteristics of a natural person, must imply a measurement of this characteristics.'*[4] For instance, '*the video footage of an individual cannot however in itself be considered as biometric data under Article 9, if it has not been specifically technically processed in order to contribute to the identification of an individual.'*[5] The same conclusion can be reached in relation to the processing of images of the facial traits of natural persons.

It must be noted that biometric data, further to being personal data, is also a special category of personal data in accordance with Article 9 of the GDPR, when used for the purposes of uniquely identifying a natural person. The legal bases for the processing of special categories of personal data, enumerated in Article 9(2) of the GDPR, are different and more restrictive compared to those for the processing of the other personal data according to Article 6(1) of the GDPR.

4.1.1.1.4 Reference data stored in the blockchain: stamps of approval, timestamps and logs

A preliminary step to conduct concerns the identification of the personal data processed in the shared blockchain of the FIDES platform.

In public and permissionless blockchains, the type of data that is processed in a blockchain are public keys and transactional data. Since FIDES is not a public and permissionless blockchain, but private and permissioned, it must only be assessed whether transactional data processed in FIDES qualifies as personal data. Transactional data are any data that may be used on the blockchain apart from public keys. They are data relating to the transaction. Transactional data could thus qualify as personal data under the GDPR where such data would directly or indirectly relate to an identified or identifiable natural person. Transactional data may be stored in the blockchain in plain text, encrypted form or hashed form.

In the case of FIDES, the shared blockchain is only used for the storage of reference data. Once that the bank decrypts the customer data received through the Norbloc bank portal, it subsequently verifies that the quality of the data is good and creates a stamp in the form of a digital signature on the blockchain. If any changes must be made to the customer file, the same process is repeated with the creation of a new stamp on the blockchain. Therefore, the data processed in the blockchain shared across banks consists of stamps of approval, hashes, timestamps and logs, i.e. reference data.

With regard to stamps of approval, timestamps and logs, their qualification as personal data must be assessed on a case-by-case basis, considering whether they contain any information relating to an identified or identifiable natural person. These data are clearly 'information' within the meaning of Article 4(1) of the GDPR, and may be ''relating to'' the data subject within the meaning of the same Article. On the latter point, it must be ascertained if stamps, timestamps and logs contain any reference to actions of the data subject, e.g. their requests to the bank or any actions taken within the Norbloc platform, or any reference to their files. For instance, if they contain a reference to the request that the

[4] EDPB, Guidelines 3/2019 on processing of personal data through video devices, 2020.
[5] Ibid.

data subject has made to the bank or on the FIDES platform, they could be relating to the data subject. If, on the other hand, they simply contain information about the fact that the storage of the hashes on the blockchain was approved on a given date, without any reference to the actions taken by the data subject, it cannot be considered as relating to a data subject and would not qualify as personal data.

4.1.1.1.5 Reference data stored in the blockchain: hashes

As far as the hashes are concerned, these are the digital fingerprints of encrypted individual data fields, and not of the entire customer/applicant files. The individual data fields are the customer ID, field name, field data, expiry date, verification stamp and verifier ID. The encryption key is a one-shot usage key. If the data is updated, then a new encryption key is issued.

If the individual data field contains personal data (even if pseudonymous), there is a risk that the hash equally qualifies as personal data should it be possible to guess the data to which the hash relates. To determine whether the hash contains personal data, it must first be assessed which are the risks to re-compose the original individual data field from the hash of encrypted data.

The assessment of whether the hashed data is pseudonymous or anonymous must be conducted in light of the '*means reasonably likely to be used*' test under Recital 26 of the GDPR, taking into account the circumstances of the case. This assessment is based on a legal concept that must be conducted analysing the technical possibility of reversing the hash into the plaintext. Whilst the technical analysis of reversing the hash can only be carried out by Norbloc, the following aspects must be taken into account to assess re-identifiability in this specific case:

- A data field is anonymous if it cannot be attributed to a natural person, not even by combining it with additional information. To determine whether a natural person is identifiable, Recital 26 states that *'account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly*'.
- To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.
- When assessing the likelihood of re-identification, it must be verified whether the data controller or any third party would be able to identify the data subject with all the means reasonably likely to be used.
- Data that qualifies as anonymous at a given time may not be anonymous in the future. As clarified by EDPS and AEPB, anonymisation may not be a permanent a status and can be subject to changes over time. For instance, technological advancements could increase the risks of re-identification. For this reason, the data controller shall adequately monitor technological or other developments that could affect the risks of re-identification of a data subject from a dataset.
- The risks of re-identification do not need to be zero, a residual risk of re-identification is possible and does not prevent the data to qualify as anonymous, as affirmed by the AEPD and EDPS. An anonymisation process aims to reduce the re-identification risks below a certain threshold, that will depend on multiple factors such as: i) existing mitigation controls, ii) the impact on individuals' privacy in the event of re-identification, iii) the motives and capacity of an attacker to re-identify the data.
- With specific regard to the reversibility of a hash, hashes aim to be irreversible, but re-identification of a data subject from a hash containing personal data is not impossible. For this reason, Opinion 5/2014 on anonymisation techniques establishes hash functions as pseudonymisation techniques.[6] Whilst it is impossible to reverse the hash function and obtain the original data solely from the hash, it may be possible to link the hash to a data subject by running tests with the hashing algorithm. In particular, the hashing outputs can be guessed from

[6] Article 29 Data Protection Working Party, 'Opinion 05/2014 Anonymisation Techniques' (2014).

known inputs. Thus, as noted by Finck, hashing would be non-invertible only if the number of possible inputs is sufficiently large and unpredictable so that it would be impossible to try all possible combinations. A strategy to hinder re-identification from a hash value can be to use an encryption algorithm to the value, with a confidentially stored encryption key. In this case, however, if the hash can inherently be guessed, the fact that there is an additional layer of obfuscation does not prevent re-identification, given that encrypted data is clearly pseudonymous data. When the hash is obtained from encrypted data, there would be some degree of likelihood of re-identification of the data subject for the person, such as the data controller, that has possession of the encryption key and that knows the encryption algorithm. By applying the encryption algorithm to all the potential inputs and hashing the encrypted form of the plaintext, this person would be able to guess the hash if sufficiently strong computing capabilities are available.

Given that the likelihood of re-identification must be assessed, inter alia, from the perspective of the data controller, it could be concluded that a hash is anonymous data only if the data controller loses their capacity to validate it. Moreover, it cannot be excluded that a hash of personal data could be anonymous data, if the likelihood of re-identification is close to zero because of the difficulty of calculating all the possible inputs, or because the controller has applied strategies that hinder re-identification to the extent that it becomes almost impossible. To conclude that the hash is anonymous data, a risk analysis must be conducted that takes into account the following factors (non-exhaustive list):

- ✓ The level of information entropy in the hash text;
- ✓ The use of single-use salt/random values;
- ✓ Zero links with identifiers, pseudo identifiers and other information, especially in the same record and across records, tables or parallels chains;
- ✓ The implementation of organisational measures that guarantee the removal of any information that allows re-identification for the data controller or any third party;
- ✓ The reasonable guarantee of the system robustness beyond the expected useful life of personal data.

In light of the above, the following can be concluded on the qualification of the hashes as pseudonymous data:

The individual data field whose input is turned into a hash for the blockchain contains personal data. For instance, the customer ID and the verifier ID are pseudonymous personal data which would be subject to the application of the GDPR unless they are properly anonymised. In the current state of technology, it seems that there is always a significant risk that hashed data is linkable to a data subject, even though this must be determined on a case-by-case basis. Given the currently available computing capabilities, it may be hard to have a situation where the number of possible inputs is sufficiently large and unpredictable to prevent the option of trying all possible combinations. In this regard, Edward Felten has argued that hashing is hardly an anonymisation technique, demonstrating that it is fairly easy to identify someone on the basis of hash functions obtained from social security numbers by simply having a computer guess all possible social security numbers for one country and hash them.[7] The risks of re-identification would however depend also on the data field, as guessing the hash of a phone or social security number may be easier than that of a document. Therefore, even if it should not be excluded that hashed data can be anonymous data, it would be highly risky to assume at present.

With regard to the Norbloc's FIDES platform, the hash to be stored on the shared blockchain would be the hash of an encrypted data field. When the hash is obtained from encrypted data, there would be some degree of likelihood of re-identification of the data subject for any person, such as the data controller, that has possession of the encryption key or that knows the encryption algorithm. By applying the

[7] Edward Felten, 'Does Hashing Make Data "Anonymous"?' (*Federal Trade Commission*, 22 April 2012) <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2012/04/does-hashing-make-data-anonymous> accessed 27 January 2025.

encryption algorithm to all the potential inputs and hashing the encrypted form of the plaintext, this person would be able to guess the hash if sufficiently strong computing capabilities are available.

In conclusion, the hash stored on the shared blockchain may be considered pseudonymous personal data subject to the application of the GDPR. However, the assessment over the risks of re-identification is a technical assessment that must be conducted by Norlboc taking into account the legal criteria outlined above. Thus, it is not possible to give a definitive conclusion on whether the hashes qualify as pseudonymous personal data.

4.1.1.2 Application of the GDPR to the processing of personal data in a decentralised environment

4.1.1.2.1 Preliminary considerations

When assessing whether a given processing activity is in compliance with the GDPR, the following steps must be taken:

1. check if the activity falls under the territorial and material scope of application of the GDPR,
2. identify the controller(s) and (if any) the processor(s) of the personal data,
3. assess whether there is a legal basis that can be relied on for the envisaged processing activity, iv) evaluate compliance of the processing activity with the data protection principles in Article 5 of the GDPR and with the other applicable GDPR requirements.

These steps must be taken for the assessment on the legality of personal data processing in different decentralised environments, with the ultimate aim to draw general conclusions on which compliance challenges a decentralised environment can present under the GDPR. As a disclaimer, it must be noted that a case-by-case assessment must always be carried out on the basis of the specific characteristics of a decentralised environment and of the type of processing operations performed in it. The assessment below is by no means intended to provide an exhaustive analysis of all the GDPR-related issues that would arise when processing personal data in any decentralised network, but just general considerations on the compliance challenges presented by decentralised processing of personal data.

4.1.1.2.2 Territorial and material applicability of the GDPR to personal data processing in a decentralised environment

Article 2 of the GDPR outlines its material scope of application. For the GDPR to apply, two conditions must be met: i) there is a processing of personal data, ii) the processing takes place wholly or partly by automated means, or alternatively in a filing system. Therefore, the GDPR applies to the collection, recording and storage of personal data that occurs wholly or partly by automated means. In decentralised environments, data is added, recorded and stored in the decentralised network by automated means. If this data qualifies as personal data within the meaning of Article 4(1) of the GDPR, then the GDPR applies.

The territorial scope of application of the GDPR is described by its Article 3. In the event that entities use a decentralised network to process personal data in the context of their activities of an establishment in the EU, they will be subject to the GDPR. In this regard, it is irrelevant that the processing takes place in the EU or not, as long as it is in the context of operations of an EU establishment the GDPR applies. Therefore, the GDPR could apply also where the entire decentralised network is located outside of the EU. In addition, where entities located outside of the EU process personal data of data subjects who are in the Union, in order to offer goods or services to such data subjects, or to monitor their behaviour that takes place in the Union, the GDPR applies. For instance, if the decentralised network is used by a non-EU entity to provide services to data subjects who are in the Union, and their personal data is processed with the aim to provide such services, this processing activity is subject to the GDPR.

In conclusion, if personal data is processed in a decentralised environment in the context of the activities of a controller or processor established in the EU, as in the case at hand, the GDPR fully applies.

4.1.1.2.3 Identification of the controller(s) and processor(s) for FIDES use in the banking pilot

4.1.1.2.3.1 Identification of the controller

The data controller is the entity responsible for complying with obligations laid down in the GDPR. To be able to enforce its rights under the regulation, each data subject must be able to identify the legal or natural entity it has to address. The identification of data controller in relation to each personal data processing operation in the blockchain environment is thus of fundamental importance.

Under the GDPR, the controller is '*the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data*' (Article 4(7) GDPR). In its guidelines on the concepts of controller and processor in the GDPR,[8] the EDPB identifies the following building blocks of the controller's definition to be assessed when performing the controllership test:

- "the natural or legal person, public authority, agency or other body";
- "determines";
- "alone or jointly with others";
- "the purposes and means";
- "of the processing of personal data".

A central criterion for the identification of the controller in cases of decentralised processing, where multiple entities participate in the processing and storage of personal data, is that the controller must decide on both purposes and means of the processing. The EDPB guidelines on the concepts of controller and processor in the GDPR clarify that determining the purposes and the means amounts to deciding respectively the "why" and the "how" of the processing, and that the controller must decide on both purpose and means of the processing.

As regards the determination of means, the EDPB distinguishes between essential and non-essential means. 'Essential means' are closely linked to the purpose and the scope of the processing and are traditionally and inherently reserved to the controller. Examples of essential means are the type of personal data which are processed ("which data shall be processed?"), the duration of the processing ("for how long shall they be processed?"), the categories of recipients ("who shall have access to them?") and the categories of data subjects ("whose personal data are being processed?"). 'Non-essential means' concern more practical aspects of implementation, such as the choice for a particular type of hard- or software or the detailed security measures which may be left to the processor to decide on.

The identification of the controller(s) is an assessment that must be conducted on a case-by-case basis, taking into account how the decentralised processing takes place and, most importantly, how actors interact within it. When there is decentralised processing of personal data with the nodes of the network being shared across different participants, it is essential to consider the role played by each of these participants to determine whether they act as joint controllers of the personal data processing. As noted by Finck[9], where there is a blockchain-based application with an application layer, it can be argued that the data controller is the legal entity determining the purposes of personal data processing at the application layer level.[10] This conclusion can be considered reiterated also for any other distributed network, and can be useful to identify the (joint) controllers and the data processors. Therefore, if more entities determine the purposes and means of the processing at the application layer level, and other

[8] European Data Protection Board (EDPB), 'Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR' (2021) <https://www.edpb.europa.eu/system/files/2023-10/EDPB_guidelines_202007_controllerprocessor_final_en.pdf> accessed 27 January 2025.
[9] Ibid, 44.
[10] Michèle Finck, 'Cobwebs of Control: The Two Imaginations of the Data Controller in Eu Law' (2021) 11 International Data Privacy Law 333.

entities participate in the processing at the lower levels under their direction, the former would be the controllers, whilst the latter would be the processors for the same processing.

Taking the specific case of FIDES in the banking pilot at hand, FIDES is a multi-layered platform composed of a set of nodes, installed at the premises of each participant. The banks share a common infrastructure where the personal data processing takes place. The banks are the entities that determine the purposes of the processing, i.e. to perform the KYC checks, and decide that such processing shall take place on the FIDES platform, thereby choosing the means. The fact that the technological solution is provided by an entity different from the banks, and the operation of the platform is not managed by the banks is irrelevant for the determination of the controllers. As a matter of fact, it is not necessary that the controllers carry out the actual processing of the personal data, nor that they have access to the personal data,[11] the determinant factor being that they decide the purposes and means of the processing, as is the case with the participating banks.

The qualification of Norbloc as a controller or a processor for the processing operations taking place in FIDES is a more complex assessment. The outcome of this assessment depends on how the collaboration with the banks is established, and how the service is provided. In particular, if Norbloc establishes a collaboration with the banks, determining that FIDES should be used for KYC purposes and participating in the decision-making process on the use of the platform for such purposes, Norbloc may be a (joint) controller alongside the banks. If, on the contrary, Norbloc is merely providing the service to the banks and limits its participation to the technical delivery of the service, Norbloc would likely qualify as a processor. The fact that Norbloc has access to the master encryption keys for the blockchain is irrelevant for the qualification as data controller, as long as Norbloc does not participate in the determination of the purposes and means of processing**.**

4.1.1.2.3.2 Banks as joint controllers

Once that it has been ascertained that the banks can be the controllers, it must be considered whether they can be joint controllers. Besides a single controller, there can also be joint controllers for the same processing operation. According to Article 26 of the GDPR, where two or more controllers jointly determine the purposes and means of the processing, they shall be joint controllers. The assessment of joint controllership shall, as affirmed by the EDPB[12], mirror the assessment of single controllership, and it shall consist of a factual rather than formal analysis.

As clarified by the EDPB, '*the overarching criterion for joint controllership to exist is the joint participation of two or more entities in the determination of the purposes and means of a processing operation*'.[13] Therefore, it must be assessed if the banks that use the shared platform and Norbloc jointly determine the purposes and means of the personal data processing. With regard to the determination of the purposes the EDPB stated, recalling CJEU's caselaw, that '*when the entities do not have the same purpose for the processing, joint controllership may also, in light of the CJEU case law, be established when the entities involved pursue purposes which are closely linked or complementary*'.[14] In particular, this may be the case, for example, when there is a '*mutual benefit arising from the same processing operation, provided that each of the entities involved participates in the determination of the purposes and means of the relevant processing operation*'. As concerns the means, instead, the EDPB stressed that the joint determination of the means does not require that each entity involved needs to determine the means to the same extent. On the contrary, the EDPB affirmed that '*it may also be the case that one*

[11] *Judgement of 5 June 2018, Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH (C-210/16)* (Court of Justice of the European Union); *Judgement of 29 July 2019, Fashion ID GmbH & Co KG v Verbraucherzentrale NRW eV (C-40/17)* [2019] Court of Justice of the European Union (ECLI:EU:C:2019:629).
[12] European Data Protection Board, 'Guidelines 07/2020 on the concepts of controller and processor in the GDPR' < https:// edpb.europa.eu/sites/edpb/files/consultation/edpb_ guidelines_202007_controllerprocessor_en.pdf >."
[13] European Data Protection Board (EDPB), 'Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR' (n 7).
[14] *Judgement of 29 July 2019, Fashion ID GmbH & Co. KG v. Verbraucherzentrale NRW eV (C-40/17)* (n 10); *Judgement of 5 June 2018, Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v. Wirtschaftsakademie Schleswig-Holstein GmbH (C-210/16)* (n 10).

of the entities involved provides the means of the processing and makes it available for personal data processing activities by other entities. The entity who decides to make use of those means so that personal data can be processed for a particular purpose also participates in the determination of the means of the processing'.

By applying the criteria outlined above, the banks that use the FIDES platform for KYC compliance purposes are to be considered controllers for the processing operations on the customer data that can be retrieved on the platform. In particular, the banks pursue purposes that are closely linked or complementary, if not identical, and obtain a mutual benefit from the participation in the platform by facilitating the KYC process thanks to the retrieval of customer information stored in DRILL and the blockchain. Norbloc may be a joint controller alongside the banks if the conditions outlined above in the section on single controllership are met.

If the banks are joint controllers, they should either designate a single legal person as the controller or conclude an agreement setting out their respective responsibilities. This then also allows the data subject to identify the entity they need to contact to enforce their rights and provides a single point of contact for data protection authorities.

4.1.1.2.4 Identification of the processor

If Norbloc does not qualify as data controller, it could be a data processor for the personal data processed in FIDES, if it directly processes personal data and its role is not limited to providing the technical solution to the banks without carrying out any actual processing.

4.1.1.2.5 Compliance with data protection principles of Article 5 GDPR in a decentralised environment

4.1.1.2.5.1 Lawfulness of personal data processing

There are more valid legal bases that could potentially be relied on for the processing of personal data in a decentralised environment. The fact that the processing of personal data takes place in a decentralised environment does not have any relevance for the choice of the legal basis to rely on, in other words it does not preclude or mandate reliance on a specific legal basis. The applicability of a legal basis in practice depends on the circumstances of each specific case.

However, there are some considerations that can be made about how the processing of personal data in a decentralised environment, and with the use of blockchain technology, can render one legal basis more or less adequate and practically viable to rely on.

First, consent may not be a practically convenient legal basis to rely on in a decentralised environment, and in particular when personal data is stored in the blockchain. If consent is relied on as the legal basis for the processing of personal data on the blockchain, it is essential to ascertain that it is freely given, specific, informed and unambiguous. The 'informed' aspect in particular is important, as data subjects must be aware that their personal data will be processed in the blockchain. Detailed information must be provided on how personal data would be processed in the blockchain, and to whom it will be shared. There is at least one problem with using consent for the processing of personal data in the blockchain. Once personal data is included in one of the blockchain's blocks, it will continue to be indirectly processed for as long as the ledger exists. The GDPR, however, provides that the data subject has the right to *'withdraw his or her consent at any time'*. Whilst the withdrawal of consent does not affect the lawfulness of prior processing, the controller would need to rely on another legal basis to continue processing the data in the blockchain. The possibility to continue the processing of personal data after the withdrawal of consent is recognised by Article 17(1)(b) and (3) GDPR. Therefore, consent as a legal basis could be a viable legal basis for processing personal data in the blockchain insofar as the controller can rely on another legal basis to continue the processing of the personal data once that consent is withdrawn. If this is not the case, unless mechanisms are implemented that can halt the processing operation in the event the data subject withdraws consent, Article 6(1)(a) of the GDPR is an inconvenient ground for personal data processing on blockchains, due to the fact that the controller would need to be

prepared to rely on another legal basis to continue the processing and comply with the GDPR requirements stemming from reliance on such different legal basis.

Second, the applicability of the legitimate interest legal basis may be harder to prove when personal data is processed in a decentralised environment. This legal basis can be relied on when the legitimate interests of the controller in processing the data are not outweighed by the interests of the data subjects in not having their personal data processed. Therefore, a balancing between the interests of the data controller and of the data subject must be carried out, taking into account a series of factors, including whether the data subject can reasonably expect at the time and in the context of the collection of the personal data that processing for that purpose may take place.[15] In a decentralised environment, the personal data is subject to an intrusive processing that is split across multiple entities, which could reinforce the interest of the data subject in not having their personal data processed. However, this would depend also on the relationship between the controller and the processor, and on the context of the processing. With regard to the use of FIDES in the banking pilot, for instance, even if consent was not acquired from the data subject, it can be argued that the customer could reasonably expect that its data would be processed in a decentralised environment.

Besides the general considerations on the legal bases that could be relied on for personal data processing in a decentralised environment, it must be noted that, in the banking and public admin pilots, there are two legal bases that would respectively most likely apply: compliance with a legal obligation and performance of a task in the exercise of official authority. In the banking pilot, the processing of personal data to comply with KYC/AML requirements is a classic example where the legal basis in Article 6(1)(c) of the GDPR would apply, as the processing is necessary for compliance with a legal obligation to which the controller is subject. Indeed, the banks are subject to KYC obligations imposed by the applicable EU and national anti-money laundering and financing of terrorism legislation. In the public admin pilot, the processing of personal data would be necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. Thus, the legal basis in Article 6(1)(e) of the GDPR would apply.

However, the possibility to rely on the legal bases in Article 6(1)(c) and (e) in practice must be assessed in light of the circumstances of each case, taking into account how the legal obligation is formulated and how the administrative task is entrusted to the authorities involved in the public admin pilot.

Since the application of these legal bases should not give rise to legal uncertainty and be sufficiently straightforward, it is advisable to rely on them rather than on consent or on legitimate interest.

4.1.1.2.5.2 Purpose limitation of personal data processing

The purpose limitation principle of Article 5(1)(b) GDPR requires that personal data is collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. This principle aims to prevent a processing of personal data in a way of for purposes that data subjects would find unexpected, inappropriate or objectionable. The purpose limitation principle is made of two components: purpose specification (processing only for specified, explicit and legitimate purposes) and compatible use (not further processing in a manner that is incompatible with the purposes).

With regard to respect for the purpose limitation principle in a decentralised environment, it must be noted that the decentralised processing of the data per se is does not pose specific challenges of compliance with purpose limitation. As long as the purpose of the processing is legitimate and communicated clearly and explicitly to the data subject, and the personal data is not processed in an incompatible manner, the decentralised storage of the personal data does not conflict with the purpose limitation principle.

[15] Regulation (EU) 2016/679 of the European Parliament and of the Council - of 27 April 2016 - on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) 2016, Rec. 47.

However, there might be a potential conflict when the processing of personal data takes place in the blockchain, due to the fact that the blockchain is composed of blocks and the data in these blocks continue to be processed as more blocks are added and as long as the blockchain is operational. As noted by Finck, it must be considered whether the further processing of data added to blocks after the execution of the transaction for which it was originally inserted in the ledger can be considered to be compatible with the purpose limitation principle.[16]

This potential conflict can be solved in relation to the two different components of the purpose limitation principle. In particular, the controller should:

- Inform the data subject that the personal data would be processed in the blockchain and that, in order to achieve the purposes intended for the processing, such processing cannot be limited to the first transaction but will continue for as long as necessary for the operation of the blockchain;
- Ensure that processing in the blockchain is not incompatible with the original purposes. Whether this requirement is met must be considered on a case-by-case basis. In any case, the controller could explicitly specify to the data subject that the personal data would be processed for the purposes of storing it in the blockchain and consequently ensure its immutability. Therefore, the processing of personal data beyond the first transaction would fall under the original purpose of storing the data in an immutable ledger.

4.1.1.2.5.3 Data minimisation during personal data processing

The principle of data minimisation of Article 5(1)(c) imposes that personal data subject to processing be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. In substance, this principle requires that a necessity test is carried out, taking into account the purposes of the processing and assessing whether the processing of the personal data is strictly necessary for these purposes. The assessment of necessity entails an analysis of whether there are less intrusive means of achieving the same purposes, without processing the personal data.

In a decentralised environment, there may be a potential conflict with data minimisation if an entire copy of the personal data is stored in multiple databases that compose the network, because there would be a multiplication of personal data processing which may not be strictly necessary for the processing purposes. Moreover, the use of the blockchain may cause concern for respect of data minimisation. In particular, as already noted for purpose limitation, data added to the block continue to the processed over time for the operation of the blockchain, and a full copy of the data is replicated in the distributed networks. This outcome appears intuitively against the concept of minimisation. Nonetheless, the existence of a potential conflict does not imply that processing of personal data in a decentralised environment and in the blockchain is always contrary to data minimisation. It depends, instead, on the circumstances of the case, and most importantly on the purposes of the processing. Purpose limitation, as a cornerstone principle under the GDPR, is of fundamental importance for assessing compliance with data minimisation. If the purposes of the processing require a replication of the personal data across multiple nodes of the network and for a continued amount of time, data minimisation is complied with as the processing would be necessary for these purposes. In particular, the processing of personal data in a decentralized environment and in the blockchain would compliant with data minimisation if the following conditions are met:

1. The purpose of the processing is properly communicated to the data subject, with a clear explanation that the personal data would be stored in a decentralised environment and in the blockchain because such storage is necessary for achievement of the purpose. For instance, this could be the case if the purpose is to achieve data security and immutability of personal data by storing it in the blockchain, thereby establishing an inextricable link between the purpose and the technology used for data storage.

[16] Finck (n 9).

2. The amount of data stored in the network is the data strictly necessary for the specified purpose. The necessity test must be conducted taking account of the characteristics of the technology used, assessing what data must be processed to achieve the intended technical effect sought with the processing (e.g. immutability of a record in the blockchain).

4.1.1.2.5.4 Storage limitation of personal data processing

The principle of storage limitation of Article 5(1)(e) GDPR requires that personal data be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.

Compliance with this principle may be harder to prove for the processing of personal data in the blockchain, for the same reasons outlined above in relation to purpose limitation, as the blockchain requires continued processing of personal data for its operation. In this regard, the same considerations must be reiterated as those made above in relation to purpose limitation. Compliance with storage limitation can be demonstrated by arguing either that the continued processing is necessary for the specified purpose, with a clear explanation that the personal data would be stored in a decentralised environment and in the blockchain because such storage is necessary for achievement of the purpose and/or that the technical characteristics of the blockchain require the continued processing of the personal data in order to achieve the intended technical effects sought with the data processing purpose.

4.1.1.2.5.5 Accuracy

The accuracy principle is enshrined in Article 5(1)(d) of the GDPR and requires that personal data be accurate and, where necessary, kept up to date. It also mandates that every reasonable step be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.

The blockchain can present technical obstacles to ensure compliance with the accuracy principle when data in blocks must be erased or rectified. The removal or modification of data in the blockchain is very difficult to implement due to the interconnections between the blocks. This issue is analysed below in relation to the data subjects' rights, and in particular to the right to modification.

4.1.1.2.5.6 Fairness, transparency, integrity and confidentiality of data processing

These principles are enshrined, respectively, in Articles 5(1)(a) and 5(1)(f) of the GDPR. They require that personal data is processed in a manner that is fair and transparent in relation to the data subject, and that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures.

Processing in a decentralised environment and in the blockchain does not appear to pose any technical limitations to comply with the principles of fairness, transparency, integrity and confidentiality, nor warrant any specific considerations. These principles should be complied with as in any other processing of personal data, without specific hurdles arising in the decentralized environment and the blockchain.

4.1.1.2.6 Compliance with data subjects' rights in a decentralised environment

Pursuant to Article 12(2) of the GDPR, the controller must ensure that data subjects are able to exercise their rights of access, to rectification, to erasure, to restriction of processing, to data portability, to object and not to be subject to automated individual decision-making. These rights are conferred upon data subjects by Articles 15 to 22 of the GDPR.

Given that the controller is accountable for facilitating the exercise of these rights, the controller needs to ensure that the technical and organisational measures are in place to enable data subjects to exercise their rights. This entails that, when the processing takes place in a decentralised environment and in the blockchain, the controller is obliged to make sure that the technologies employed for the processing do

not hinder compliance with Article 12(2) of the GDPR. Whether data subjects' rights can be exercised in practice must be assessed on a case-by-case basis, taking into account the circumstances of each data processing operation. However, there are some general considerations that can be made with regard to compatibility of processing personal data in a decentralised environment and in the blockchain with the exercise of data subjects' rights. These considerations are made solely in regard to the rights whose exercise is at a clear risk of encountering obstacles in decentralised environments and the blockchain, and not for all the data subjects' rights conferred under the GDPR. In particular, the sections below will present considerations in compliance with right of access, right to rectification, right to erasure, right to restriction of processing and right to object.

- **Right of access**

The right of access of Article 15 of the GDPR grants the data subject a right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and to information on the purposes of the processing and the categories of personal data processed.

The processing of personal data in a decentralised environment and in the blockchain does per se present hurdles for the exercise of the right of access. The controller must put in place governance arrangements that enables the data subject to access their personal data. In a decentralised environment, this would all the most important if data is split into data chunks that are distributed across the participants in the network. Given that the separate data chunks are pseudonymous data insofar as re-identification is possible, participants in the network must have arrangements in place to enable the recollection of the separate chunks and access to personal data by the data subject. This is particularly relevant for the personal data stored in DRILL, given that at all time it must be possible to recombine the data chunks and decrypt them to grant access to the data subject.

If there are joint controllers of the personal data participating in the network, these controllers are obliged to determine their respective responsibilities and put in place arrangements to comply with the access request, as is the case also for the other data subjects' rights.

- **Right to rectification**

The right to rectification of Article 16 GDPR confers data subjects the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her. The right to rectification stands clearly in conflict with the technical objectives of immutability and data integrity of the blockchain, and specific precautions must be implemented to ensure the exercise of this right in a decentralised environment even when the blockchain is not used.

As concerns the exercise the right of rectification in the blockchain, it may be hard for the data controller to implement the rectification request, due to the fact that all the relevant blocks would need to be re-hashed. For instance, if the data subject requires rectification of the personal data whose hash is stored on a given block, all the subsequent blocks would have to be re-hashed. Nevertheless, this is not impossible to achieve in a private and permissioned blockchain if the appropriate technical and organisational measures are in place.

If there are joint controllers of the personal data participating in the network, these controllers are obliged to determine their respective responsibilities and put in place arrangements to comply with the rectification request, as is the case also for the other data subjects' rights.

- **Right to erasure**

According to Article 17 of the GDPR, the data subject has the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the grounds listed in Article 17(1) applies. Among the applicable grounds, there is the objection to processing pursuant to Article 21 of the GDPR and the withdrawal of consent. These two grounds pose the biggest challenges for controllers of

processing operations taking place in a decentralised or in the blockchain, as they enable data subjects to demand erasure at any time and are thus outside the control of the controller.

In a decentralised network, the erasure of stored data poses governance challenges, since appropriate arrangements must be in place to ensure that all the nodes comply with the erasure request. In this regard, if there are joint controllers of the personal data participating in the network, these controllers are obliged to determine their respective responsibilities and put in place arrangements to comply with the erasure request, as is the case also for the other data subjects' rights.

In the blockchain, erasure also poses significant technical challenges, for the same reasons outlined above for the right to rectification.

- **Right to restriction of processing and to object**

Articles 18 and 21 provide, respectively, for the right of the data subject to demand the restriction of processing and to object to the processing.

Compliance with these Articles raises the same governance challenges described above for the right to erasure, i.e. coordination between data controllers that govern different nodes in a decentralised environment. As concerns the blockchain, compliance with these rights requires the implementation of technical solutions that enable the halting of personal data processing.

4.1.1.2.7 Compliance with requirements on security of the processing

Article 32 of the GDPR requires the controller and the processor to implement appropriate technical and organisation measures to ensure a level of security appropriate to the risks for the processing of personal data, taking into account '*the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons*'. Therefore, the security obligation must be complied with, taking into account the principle of proportionality, and consequently adopting a risk-based approach on the basis of the circumstances of the processing and the risks faced.

Article 32 GDPR clarifies that the following could be considered, amongst others, as appropriate measures to ensure the security of the processing:

> (a) the pseudonymisation and encryption of personal data;
>
> (b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;
>
> (c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;
>
> (d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

Processing of personal data in a decentralised environment can provide both advantages and disadvantages for compliance with the security requirements of Article 32 of the GDPR. On the one hand, decentralised processing brings about data redundancy which is beneficial for the integrity of the data, and splitting the data in separately unintelligible encrypted data chunks (as is the case in FIDES) can be deemed to be an appropriate measure to ensure security of the data. On the other hand, one disadvantage of decentralised processing is that there is not a single point of control over the data, and the data controller(s) must put in place appropriate arrangements to make sure that the security requirements are complied with in all the nodes.

4.1.1.2.8 Conclusions

Based on all the above considerations regarding personal data processing in a decentralised environment, the following brief answers can be provided:

- When client data can be stored at local physical data centre on premises of the data controller.
  This storage is possible if there is an applicable legal basis for the processing (as would likely be the case in the banking and public admin pilots as explained above) and if all the requirements of the GDPR are complied with (including those that were not specifically analysed above, e.g. transparency requirements towards the data subject). If the entire data field is only stored in the local premises of the controller, without being divided in multiple data chunks with other nodes of a network, the challenges outlined above on the processing of personal data in a decentralised environment do not arise. However, storage of the entire dataset in a single database entails heightened risks from a security standpoint compared to decentralised processing and storage, and thus requires the controller to put in place appropriate measures.
- When client data can be stored at shared virtual premises (in a general-purpose cloud) of the data controller.
  This storage is possible if there is an applicable legal basis for the processing (as would likely be the case in the banking and public admin pilots as explained above) and if all the requirements of the GDPR are complied with (including those that were not specifically analysed above, e.g. transparency requirements towards the data subject). When personal data is stored in a shared virtual premise, the parties having access to the premise may be joint or separate controllers for the processing operation consisting of the storage in the cloud. If they are joint controllers, they must determine their respective responsibilities for the processing and put in place arrangements to ensure that the processing taking place in the cloud is compliant with the GDPR. From a security standpoint, the following steps are recommended: i) encryption of the data stored in the cloud and secure handling of the key, ii) secure management of the access rights to the cloud, including by restricting access only to cases/persons for which it is strictly necessary.
- When client data can be stored at dedicated virtual premises (in a private cloud) of the data controller.
  This storage is possible if there is an applicable legal basis for the processing (as would likely be the case in the banking and public admin pilots as explained above) and if all the requirements of the GDPR are complied with (including those that were not specifically analysed above, e.g. transparency requirements towards the data subject). From a security standpoint, the following steps are recommended: i) encryption of the data stored in the cloud and secure handling of the key, ii) secure management of the access rights to the cloud, including by restricting access only to cases/persons for which it is strictly necessary.
- When client data can be stored at physical/virtual premises of a data processor.
  This storage is possible if there is an applicable legal basis for the processing (as would likely be the case in the banking and public admin pilots as explained above) and if all the requirements of the GDPR are complied with (including those that were not specifically analysed above, e.g. transparency requirements towards the data subject).
  When the processing is carried out by a data processor, specific additional obligations arise under the GDPR for the data controller(s). These obligations are listed in Article 28 of the GDPR and include the following: i) choosing only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of the GDPR, ii) stipulation of a binding contract with the processor that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller, iii) providing instructions to the processor and monitoring activities of the processor to ensure that the GDPR is complies with at all times. The controller remains fully liable for processing carried out by the processor, with the exception of: a) obligations specifically addressed to the processor, b) acts of the processor outside or contrary to the lawful instructions of the controller, c) when the controller proves that it is in no way responsible for the event giving rise to damage.

4.1.2 Elements that can constitute a proof of customer consent to process personal data

4.1.2.1 General considerations

According to Article 7(1) GDPR, the controller must be able to demonstrate that the data subject has given consent to the processing of their data, thus having the burden of proof in case should the validity of consent be contested. This is line with the principle of accountability, that requires the controller to prove compliance with the data protection principles.

The discussion around the demonstration of acquisition of consent imposes to distinguish between consent under Article 6 and explicit consent under Article 9 of the GDPR. Explicit consent is required only when the controller processes the special categories of personal data indicated in Article 9 of the GDPR. Given that consent must be acquired in different forms in these two cases, the requisite elements to prove also differ. The difference lies in the fact that simple consent can consist of any affirmative action by the data subject, whereas explicit consent requires an explicit acknowledgement by the data subject that he or she is consenting to the processing.

Valid consent must be informed, free, specific and unambiguous. The controller must therefore be able to prove that all these four requirements are complied with when consent is collected from the data subject. The GDPR does not impose any limitation on the forms of acquisition of consent, and does not require a formal signature nor imposes any other formalities. As clarified by the EDPB,[17] '*controllers are free to develop methods to comply with this provision in a way that is fitting in their daily operations*'. Moreover, the EDPB stated that '*it is up to the controller to prove that valid consent was obtained from the data subject. The GDPR does not prescribe exactly how this must be done'*. Therefore, any element can be used to prove acquisition of consent.

4.1.2.2 Elements to be used to prove consent and general factors to take into account

Whilst any element can be used to prove consent, the following can certainly constitute proof of consent:

- A signed document in paper form;
- An electronically signed document;
- A ticked box on a sign-on page, as long as the box was not pre-ticked and the data subject carried out an affirmative action that unambiguously shows that he or she consented to the processing. In this case, electronic records that demonstrate that the data subject ticked the box can be used to prove acquisition of consent, such as a screenshot of the webpage and a timestamp showing the date and time of consent;
- The data submitted and the timestamp of an electronically filled template indicating consent;
- Any other electronic record of an affirmative action that unambiguously shows that the data subject consented to the processing.

Apart from the elements enumerated above, there are other elements that can be used to prove consent on a given case, and their validity will depend on the circumstances of the case. The following factors can be taken into account to ensure that acquisition of consent can be demonstrated:

- The documentation or other elements proving the acquisition of consent should be kept at least for the entire duration of the processing activity;
- The processing of personal data necessary to demonstrate consent shall be limited to what is strictly necessary. After the processing activity ends, proof of consent should be kept no longer then strictly necessary for compliance with a legal obligation or for the establishment, exercise or defence of legal claims;
- To prove that consent was informed, the controller must keep the documentation that was provided to the data subject before the acquisition of consent. For instance, when consent is collected in paper form, this may be easy to prove by asking the data subject to sign the same

[17] European Data Protection Board (EDPB), 'Guidelines 05/2020 on Consent under Regulation 2016/679' (2020) <https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf>.

document that contains the information provided. On the other hand, if consent is acquired electronically, the controller may keep a copy of the webpage containing the information provided alongside the consent of the data subject;
- To prove that consent was free, the controller shall ensure that there are no elements indicating that the data subject could not take a free choice, e.g. due to a situation of imbalance of power, because acceptance of consent was bundled with other conditions, or for fear of a detriment;
- To prove that consent was specific, i.e. for one or more specific purposes, the controller must exhibit the documentation submitted to the data subject showing that the data subject was able to give specific consent;
- To prove that consent was unambiguous, the controller must possess elements proving an affirmative action of the data subject unambiguously showing the wish to consent to the processing;
- To prove explicit consent, the controller should acquire and keep a statement by the data subject explicitly consenting to the processing. This statement may be in the form of a document signed electronically or on paper, of an email, of a recording of oral consent, of a filled electronic form, etc.

4.1.3 Re-evaluation of the previous legal advice on FIDES

We were asked to evaluate the legal advice provided by the law firm Lindahl in 2016 on compliance by the FIDES platform with (Swedish) data protection law. In this regard, we concluded the following:

- The legal advice was provided before the date of application of the GDPR, and is therefore based on a different legal framework, i.e. Swedish data protection law at the time. As a consequence, the conclusions reached at the time are certainly different from those to be reached at present under the GDPR;
- The section of the legal advice dedicated to data protection is very short to and does not analyse many relevant aspects that would now be relevant under the GDPR;
- We overall agree with the conclusions of Lindhal, especially on the identification of the banks as controllers when FIDES is used for KYC purposes. There is only one point on which we would not agree. In particular, for the way Norbloc's participation is envisaged, we are not certain that it would not be a data controller nor a data processor. At the very least, if it is Norbloc that continues to operate the platform as a provider of the service to the banks, Norbloc would qualify as a processor.

4.1.4 WP3 Protocol compliance and recommendations

As a logical continuation of the assessment of design recommendations and their implementation in WP3, as included in Annex II of D3.1, the following table builds upon those findings and provides the final version of this assessment. Moreover, based on the non-compliant features and/or gaps as identified, some recommendations can be provided in order to ensure that the Blockchain-based data storage and sharing technology follows the WP3 Protocol.

Recommendations for technological design (protocol)	Compliant features	Non-compliant features/Potential gaps	Recommendations
Data mapping (predetermination of the type of data to be processed)	Predetermination of all types of data that may be stored and shared through DRILL. No pre-assumptions about the type of	Possible risks in maintaining a clear classification of data nature across all nodes in case of P2P decentralised storage.	Create a comprehensive data classification framework to identify and categorize the types of data processed across all FIDES nodes.

Recommendations for technological design (protocol)	Compliant features	Non-compliant features/Potential gaps	Recommendations
	data, the most stringent requirements are assumed. Explicit consent management via DnAMS.		
Data obfuscation (disguising the data with the aim to render it unlikable or unobservable)	Encryption and hashing by DRILL. Control of access to encryption keys and guarantee of release only upon valid 'proof of consent' by DnAMS.		Symmetric encryption used in DRILL up-to-date with the latest standards. Implementation of additional obfuscation layers.
Data minimisation (limiting the processing of the data)	Inclusion solely of data for which client consent was given in plaintext data of every participant node.	Inherent conflict with decentralised storage/blockchain technology. Risks of data storage beyond required retention periods w/o automated data deletion periods.	Introduction of mechanisms for automatic deletion of data chunks stored on DRILL nodes once the retention period expires. Minimisation of the information stored in the blockchain by ensuring that only essential data are stored thereon.
Data abstraction (limiting the detail in which personal data is processed)	Storage of as little information as possible in logs.		Inclusion of aggregation/summarisation tools to limit processing in detail.
Data separation (separating, logically or physically, the processing of data)	Separation in data chunks by DRILL only when combined with self-encryption and decryption. Separation of keys from data. Key management of keys by DnAMS independently from DRILL nodes.	Reintroduction of centralised of failure in case of improper configuration of DnAMS.	Introduction of automated validation checks during FIDES node deployment to detect misconfigurations that could reintroduce centralised storage or access points. Testing and documenting of DRILL chunking and separation logic to confirm that no individual node can reassemble data without

Recommendations for technological design (protocol)	Compliant features	Non-compliant features/Potential gaps	Recommendations
			proper permissions and that logical separation is maintained for different data types.
Security (protection against unauthorised or unlawful processing or against accidental loss, destruction or damage, of the data)	Security by design.		Create detailed incident response plans to handle data breaches within FIDES, including logging and monitoring to detect breaches and defined protocols to meet GDPR's 72-hour breach notification requirement.
Control over the data (ability to change the processing conditions of the data at any time during the processing)	Explicit consent management through DnAMS. Dynamic update of sticky policies.	Inherent contradiction of blockchain's immutability with the right to erasure (Art. 17 GDPR)	Development of a solution to handle GDPR's right to erasure despite blockchain immutability (e.g. by storing only references or pointers in the blockchain while keeping the actual data in mutable storage).
Record-keeping and demonstrability (recording of how processing of the data takes place in order to demonstrate compliance with the relevant legal framework)	Automated immutable audit logs in the blockchain. Facilitation of auditing through storage of proofs and metadata on the blockchain.	Hinderance of proof of accountability in case of lack of granularity or accessibility of audit logs.	Leveraging of the immutable audit logs on the blockchain to generate automated reports demonstrating compliance with Art. 30 GDPR.

4.2 Trustworthy Data Sharing [T3.2]

As already noted in D3.1. in Annex II, this technology is not directly involving the processing of personal data, its assessment against the Protocol would prove unsuccessful. Nonetheless, a brief evaluation of certain aspects of the components of these technologies against the security-related aspect of the Protocol, consisting in highlighting the strong privacy-preserving points therein, is essential for aligning with the overall spirit of this subsection.

As far as the Privacy Risk Scoring component is concerned, privacy risks before sharing data are evaluated, thus reducing exposure to unauthorized processing. Simulated attacks can estimate the impact of data breaches, help assess vulnerabilities and inform decisions on protective measures, whereas the trustworthiness assessment considers partners' past compliance, limiting access to trustworthy entities. Similarly, the Trustworthiness Scoring Module (TSM) establishes security attributes as protection objectives for data access, uses a weighted scoring model to determine whether an entity should be trusted for data access through the implementation of threshold-based security policies defining minimum security standards for data access, and employs sticky policies (SP) which can act as safeguards to prevent unauthorised use. As such, all components strengthen protection against unauthorised or unlawful processing, accidental loss, destruction, or damage.

With regard to network security, UPCV (Ubiquitous Personal Context Vectors) uses Volatile Random Numbers (VRNs) instead of storing personal data, implements decentralised information exchange, reducing exposure to centralised breaches, and by enabling peer-to-peer communication, reduces dependency on a single vulnerable hub.

Cryptographic measures and incident management and logging are not present in this technology, other maybe than TSM’s tracking of security compliance, which may help detect anomalies. These elements could be further taken into consideration, without it strictly necessary.

4.3 Confidentiality and Privacy by Design [T3.3]

4.3.1 Confidentiality and Privacy by Design mechanisms

4.3.1.1 Sticky policies (CPE-ABE)

Several characteristics of sticky policies (SP) are relevant not only for GDPR compliance, but also with Regulations 2024/1183 on digital identities (hereinafter 'eIDAS')[18] and Directive 2016/943 on trade secrets (hereinafter 'TSD').[19]

Potential processing of personal data and trade secrets in the data to which sticky policies are attached, but sticky policies as such do not contain any other personal data. Furthermore, the components for data lifecycle management, policy enforcement, policy-based encryption and decryption, confidentiality and integrity insurance and privacy compliance should contribute to ensure compliance with the GDPR and the ePrivacy Directive,[20] and to protect trade secrets within the meaning of Article 2 of the TSD. They implement data protection by design and default in accordance with Article 25 of the GDPR. The EDPS expressed support for the use of personal information management systems.[21]

Sticky policies reflect access and usage conditions as defined at a given point in time, based on customer consent or other legal bases for personal data processing. However, the conditions and legal basis for the processing of personal data may change over time, for instance if the data subject withdraws consent, requests erasure of the personal data or restriction to the processing, or objects to the processing. In these cases, there must be a technical enabler to ensure that access to, and usage of, the relevant data would change accordingly. Changes in the access and usage rights may also be needed to protect trade secrets, for instance when there is a change in trade secrets' policy or changes in the circumstances surrounding the data sharing that endanger the secrecy of information, should it be shared with certain parties. To be also noted that sticky policies can be used potentially to define access rights for trade secrets.

Another consideration that is highly relevant for understandability and transparency of sticky policies is to the possibility to translate the latter into natural language, so that parallel descriptions of the policies both in natural and machine-readable language can be made. This is necessary to enable the provision of information to data subjects in compliance with transparency requirements. Finally, an audit mechanism must be available to verify ex post that the data was properly handled, thus ensuring that the data controller can demonstrate compliance with the principle of accountability under the GDPR (burden of proof is on the controller).

4.3.1.2 User Consent (CPE & XAMCL)

Usage control is an extension to traditional access control. It is about the specification and enforcement of restrictions regulating what must (not) happen to data. Thus, usage control is concerned with requirements that pertain to data processing (obligations), rather than data access (provisions).

XACML (eXtensible Access Control Markup Language) Standard is used to introduce commonly used terms in the field of access control. XACML is a policy language to express ABAC rules. The main building blocks of the language are subject, action, resource and environment. The subject describes who is accessing a data asset (e.g., a user). The action describes what the subject wants to perform on

[18] Regulation (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework (OJ L, 3042024).
[19] Directive (EU) 2016/ 943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-How and Business Information (trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure (OJ L 157/1, 15/6/16).
[20] 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) 2002.
[21] European Data Protection Supervisor (EDPS), 'Opinion 09/2016 on Personal Information Management Systems' (2016).

the data asset (e.g., read, write). The resource describes the data asset. Finally, the environment specifies the context (e.g., time, location).

Access control in the IDS is a resource-centric regulation of access requests from subjects (i.e., IDS participants) to resources (i.e., data services). Resource owners define attribute-based access control policies for their endpoints and define the attribute values a subject must attest in order to grant access to the resource. These attributes may include:

- Specific identity of connector(s) (only access requests from a specific connector or specific connectors will be granted).
- Connector attributes (only access requests from a connector that possesses specific attributes will be granted).
- Security profile requirements (only access requests from a connector that fulfils specific security feature requirements will be granted, e.g., having a TPM >= 1.2 and doing application isolation)

In contrast to access control, where access to specific resources (e.g., a service or a file) is restricted, the IDS architecture additionally supports data-centric usage control. In general, the overall goal is to enforce usage restrictions for data after access has been granted. Therefore, the purpose of usage control is to bind policies to data being exchanged and to continuously control the way how messages may be processed, aggregated, or forwarded to other endpoints. This data-centric perspective allows the user to continuously control data flows, rather than accesses to services. At configuration time, these policies support developers and administrators in setting up correct data flows.

As regards the question of signature of the data, we refer to the analysis provided under D2.4 on eIDAS and SSI.

4.3.2 WP3 Protocol compliance and recommendations

As a logical continuation of the assessment of design recommendations and their implementation in WP3, as included in Annex II of D3.1, the following table builds upon those findings and provides the final version of this assessment. Moreover, based on the non-compliant features and/or potential gaps as identified, some recommendations can be provided in order to ensure that the Confidentiality and Privacy by Design technology follows the WP3 Protocol.

Recommendations for technological design	Compliant features	Non-compliant features/Potential gaps	Recommendations
Data mapping (predetermination of the type of data to be processed)	Predetermination of different data types can be partially achieved through CP-ABE and XAMCL modules.	Processing of unknown data types or lack of clear mechanisms to flag data that do not meet the preconditions set could render the predetermining of the nature of data impossible.	Implementation mechanisms to predefine and classify all types of data being processed or shared (e.g. metadata tagging, data categorisation frameworks, etc.)
Data obfuscation (disguising the data with the aim to render it unlikable or unobservable)	CP-ABE module obfuscates data to render it unlikable.	Current reliance on centralised key authorities for CP-ABE introduces	Decentralisation of CP-ABE key authority.

Recommendations for technological design	Compliant features	Non-compliant features/Potential gaps	Recommendations
	Signature modules ensure integrity while maintaining privacy. Decentralised Identifiers (DIDs) and Verifiable Credentials (VCs) with zero-knowledge proofs (ZKP) minimise unnecessary data exposure.	potential points of vulnerability.	Integration data anonymisation alongside encryption.
Data minimisation (limiting the processing of the data)	CP-ABE and XAMCL modules allow and enforce granular access, limiting exposure. Zero-knowledge proofs (ZKP) support minimisation by enabling verification without full data disclosure.		Implementation of automated deletion protocols and anonymisation procedures for personal data that is no longer necessary for processing.
Data abstraction (limiting the detail in which personal data is processed)	Attribute-based decisions made on by technologies such as the CP-ABE inherently include aggregation and summarisation. Selective disclosure and granular access are implemented via the CP-ABE and XAMCL modules.	No explicit mention of aggregation tools or mechanisms to abstract detailed data before processing.	Inclusion of aggregation/summarisation tools to limit processing in detail.
Data separation (separating, logically or physically, the processing of data)	CP-ABE ensures encrypted data can only be decrypted by authorized users with specific attributes, effectively separating data by policy. Data-sharing architectures in the	Physical separation is not explicitly described.	Implementation of physical separation techniques.

Recommendations for technological design	Compliant features	Non-compliant features/Potential gaps	Recommendations
	FIWARE connector support logical separation.		
Security (protection against unauthorised or unlawful processing or against accidental loss, destruction or damage, of the data)	Encryption through CP-ABE and digital signatures protect against unauthorised access.	No mention of other security mechanisms beyond encryption (e.g. firewalls, etc.)	Expansion of network security measures (e.g. firewalls, advanced logging systems, etc.)
Control over the data (ability to change the processing conditions of the data at any time during the processing)	CP-ABE and XAMCL modules allow dynamic updates to processing conditions and ensure users can manage how their data is processed.	Insufficient details on user-initiated updates during active data processing.	Enabling of real-time updates to consent by data subjects.
Record-keeping and demonstrability (recording of how processing of the data takes place in order to demonstrate compliance with the relevant legal framework)	Audit logs align with demonstrability requirements. The use of Verifiable Credentials (VCs) ensures traceability of user consent and policy enforcement.	No explicit reference to how audit logs will be maintained securely or updated for purposes of long-term compliance demonstration.	Development of secure audit logs that describe in detail data access, processing and sharing activities.

4.4 Self-Encryption and Decryption Techniques with Multi-factor Information Recovery Mechanism [T3.4]

4.4.1 Material and territorial applicability of the GDPR

The material scope of the GDPR is defined in Article 2(1). Based on the definition of the material scope, the GDPR would apply to the self-encryption of personal data, as this process clearly entails the '*processing of personal data wholly or partly by automated means'*.

Article 2(2) of the GDPR also sets out a list of situations in which the Regulation does not apply, among which there is the processing of personal data ''by a natural person in the course of a purely personal or household activity''. This exclusion concerns personal or household activities that have no connection to professional or commercial activity. Given that self-encryption may be used for different purposes and in different contexts, it cannot be excluded that certain uses of self-encryption may fall outside of the material scope of application of the GDPR because they pertain to purely personal or household activity. For instance, a natural person may decide to make use of self-encryption to send sensitive

information to friends or family members. Whilst this information may contain personal data relating to multiple data subjects, if it is encrypted only in the course of a purely personal or household activity the GDPR does not apply.

However, in the uses of self-encryption envisaged in the pilots ''autonomous vehicles'', ''smart manufacturing'' and ''retail'' (altogether, hereinafter referred to as the ''relevant pilots''), the processing of data would never be only in the course of a purely personal or household activity. Therefore, it must be assumed that the processing of personal data during self-encryption would fall in the material scope of the GDPR.

The perimeter of the territorial scope of the GDPR is defined in Article 3, and is delineated by: i) the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, ii) the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, provided that certain conditions are met. Given that all the processing activities at hand take place within the Union, they would certainly fall within the territorial scope of the GDPR.

In conclusion, the GDPR would apply to any processing of personal data effectuated during a self-encryption procedure carried out in the context of one of the relevant pilots. The next steps of the analysis require the identification of the data controller(s) and the data processor(s) in all the processing activities taking place during self-encryption and, subsequently, the assessment on the existence of a legal basis for processing and compliance with the applicable GDPR requirements.

4.4.1.1 Encrypted and hashed data as pseudonymous data

An important preliminary consideration to make is that encrypted data is always personal data, and hashed data is generally personal data except for specific situations where the hash is not reversible (as also analysed above under 4.1.1.5.).

Encrypted data is data that has been subject to pseudonymization within the meaning of Article 4, point 5) of the GDPR. Encrypted data is pseudonymous data because, once data has been encrypted, the link to an identity exists in the form of the pseudonym plus a decryption key. Without such a key, it is difficult to identify pseudonymised data. However, for those entitled to use the decryption key, re-identification is easily possible, and therefore encrypted data cannot qualify as anonymous data. The qualification of encrypted data as pseudonymous data means that the GDPR fully applies at any stage of the processes of encryption, data recovery and decryption.

As concerns the hashes, these are the digital fingerprints of data fields. If the data field contains personal data (even if pseudonymous), there is a risk that the hash equally qualifies as personal data should it be possible to guess the data to which the hash relates. To determine whether the hash contains personal data, it must first be assessed which are the risks to re-compose the original individual data field from the hash of encrypted data. The computing capabilities currently available make it quite easy to reverse the hash and obtain the original data field in many circumstances, therefore it should not be assumed that hashed data is anonymous data. However, a circumstantial assessment must be conducted on a case-by-case basis.

4.4.1.2 Identification of the personal data processed in the self-encryption process in the pilots

In the autonomous vehicles pilot, there is personal data to be processed, relating to the user profile on the smartphone app and to the location of the user while the autonomous vehicle is driving. This data is uploaded to the cloud in encrypted form, together with data on user's ID, and subsequently sent to the autonomous vehicle. In all these steps, there is self-encryption of the data. In the autonomous vehicles pilot, the following processed data may qualify as personal data: 1) the user ID, insofar as the user ID consists of an identifier of the user, even in the form of a pseudonym (such as a code), 2) data on the user location when the user first starts using the vehicle and while driving it.

In the retail and smart manufacturing pilots, the user ID may qualify as personal data if the user is a natural person and the ID relates to a natural person. There may be other personal data to be processed, but this depends on which type of data must be shared, which at present seems not be clarified from the description of the pilots. The documents descripting the pilots state that the data to be shared is anonymous.

In each of the pilots, there would be personal data to be processed during self-encryption. Hence, the GDPR applies.

4.4.2 Identification of data controller(s) and data processor(s) in the self-encryption process

During self-encryption, and subsequent data recovery and decryption, there are multiple parties that participate in the processing of personal data. For each of these parties, their role in the processing of self-encrypted data must be analysed to determine whether they are data controllers, joint controllers, data processors or sub-processors, or simply recipients of the data. This is essential to determine their qualification under the GDPR and, consequently, the specific provisions applicable to them.

The identification of the data controller(s) and processor(s) is highly circumstantial and depends on how the processing is structured and on its purposes. Therefore, a separate assessment must be conducted for each of the pilots:

In the autonomous vehicles pilot, the following actors have the following roles under the GDPR:

- **User (or data owner)**: certainly not the controller of the processing, but simply the data subject of such processing. Considering the circumstances of the case, the processing operations taking place during self-encryption are not determined by the user, as the user does not have a control over them stemming from either legal provisions or factual influence. The role of the user is confined to giving consent to the processing of its personal data, or to exercising its rights as data subjects, while using the service.
- **Operator**: responsible for the provision of the entire service consisting in the functioning of autonomous vehicles likely qualifies as the data controller. This could be the operator that provides the smartphone app from which the user can book an autonomous vehicle, and that manages the whole system by making use of the TANGO technologies (in the pilot, it may be IDIADA depending on the practical responsibilities entrusted to it). This operator will be the controller of all the processing operations on the personal data processed for the purposes of providing the service, including the processing taking place during self-encryption, unless there is another entity deciding in its stead. As a matter of fact, it is this operator that decides on the purposes and means of the processing. The EDPB guidelines on the concepts of controller and processor in the GDPR clarify that determining the purposes and the means amounts to deciding respectively the "why" and the "how" of the processing, and that the controller must decide on both purpose and means of the processing. The determination of a controller must be made for each processing operation on a set of personal data. The operator would be the controller for: i) the collection of personal data in the smartphone app, ii) the self-encryption of the personal data and subsequent transfer of personal data to the cloud servers, iii) the storage of personal data in the cloud servers in encrypted form and any potential decryption operation, iv) the transfer to the car and vice-versa. As a matter of fact, this operator decides whether, and how, personal data is to be processed for the purposes of providing the service. The design of the service is not determined by the passenger, who merely subscribes to a service and accepts its terms of use. Therefore, the operator is the entity that determines the purposes and (essential) means of the processing of the data in line with the definition of controller in Article 4(7) of the GDPR. This can be affirmed on the basis of the following two considerations: i) the data owner decides, alone or jointly with others, that the data must be subject to self-encryption, thus determining that the data is to be processed for the purposes of encrypting them, ii) the data owner decides that the self-encryption software must be used as a means to encrypt the data. If these two

conditions are met, the operator is the controller of both the data and of the ID, whose hash will be used as the encryption key. In this regard, it is irrelevant that the data owner has actual access to the data (Fashion ID C-40/17[22] and Wirtschaftsakademie C-210/16), as the determinant factor is that he/she decides the purposes and means of the processing. If the operator is the controller of the processing of personal data, it will be subject to a series of obligations and a liability regime under the GDPR.

- **Provider of the cloud servers**: in this case it's AWS (Amazon Web Services) where the exchanged data is uploaded is the processor of the personal data on behalf of the operator, the controller.
- Any other entity that participates in the functioning of the secret sharing scheme may qualify as a processor of the personal data.
- **Any other party**: if the personal data is transmitted to a third party who is not operating under the instructions of the controller or of the processor, that third party may an independent or a joint controller of the personal data. In any case, the recipient will need to have a separate legal basis for the processing of any transmitted personal data.

In the retail pilot, the following actors have the following roles under the GDPR:

- **Data owner**: initiates self-encryption and injects his or her ID. If the ID is an identifier of a natural person, then the ID is personal data and the data owner is also a data subject under the GDPR. Data on customers' behaviour may be transmitted, in this case it is important to check if this data is anonymised and, if not, the GDPR may apply if customers are identified or profiled in the data. If this is the case, the customers are data subjects under the GDPR;
- **Operator**: that manages the entire system is the data controller for all the personal data processed while using self-encryption for data sharing. In the pilot this would be METRO. However, it is not clear if there would be any processors of the data.
- **Any other party**: if the personal data is transmitted to a third party who is not operating under the instructions of the controller or of the processor, that third party may an independent or a joint controller of the personal data. In any case, the recipient will need to have a separate legal basis for the processing of any transmitted personal data.

In the smart manufacturing pilot, the same considerations can be done as for the retailer pilot.

4.4.3 Compliance with data protection principles of Article 5 GDPR

4.4.3.1 Identification of a legal basis for the processing

There may be multiple legal bases applicable to carry out self-encryption in the pilots, to be determined on a case-by-case basis and taking into account the purposes of the processing.

If the data owner is also the data subject to whom all the personal data to be self-encrypted pertains, by initiating self-encryption there is a clear affirmative action giving consent to the processing, which could be relied on as a legal basis for the processing pursuant to Article 6(1)(a) of the GDPR. Consent must be valid according to Article 7 of the GDPR. However, whether it qualifies as valid consent also depends on the circumstances of the case, including: 1) which information is presented to the data subject before he or she initiates self-encryption, 2) whether the data subject understands the consequences of initiating self-encryption for the processing of their personal data, 3) to what extent the initiation of self-encryption can be considered as an unambiguous manifestation of consent to the processing of personal data, considering the circumstances of the case.

[22] *Judgement of 29 July 2019, Fashion ID GmbH & Co. KG v. Verbraucherzentrale NRW eV (C-40/17)* (n 10); *Judgement of 5 June 2018, Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v. Wirtschaftsakademie Schleswig-Holstein GmbH (C-210/16)* (n 10).

In this regard, the considerations made above with regard to consent must to be taken into account in order to assess whether consent is valid in a given case.[23]

4.4.3.2 Display of users' IDs and data minimisation

The principle of data minimisation of Article 5(1)(c) imposes that personal data subject to processing be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. In substance, this principle requires that a necessity test is carried out, taking into account the purposes of the processing and assessing whether the processing of the personal data is strictly necessary for these purposes. The assessment of necessity entails an analysis of whether there are means of achieving the same purposes that are less intrusive for data protection.

As clarified by the EDPB in the Guidelines 4/2019 on Article 25 Data Protection by Design and by Default[24], data minimisation must be implemented by design and by default by adopting appropriate technical and organisational measures. Data minimisation must be ensured with regard to the amount of personal data collected, to the extent of their processing, to the period of their storage and to their accessibility. In all these dimensions, the controller must minimise personal data processing.

The Guidelines list, by way of example, the following key design and default data minimisation techniques:

- Data avoidance – Avoid processing personal data altogether when this is possible for the relevant purpose.
- Limitation – Limit the amount of personal data collected to what is necessary for the purpose;
- Access limitation – Shape the data processing in a way that a minimal number of people need access to personal data to perform their duties, and limit access accordingly.
- Relevance – Personal data should be relevant to the processing in question, and the controller should be able to demonstrate this relevance.
- Necessity – Each personal data category shall be necessary for the specified purposes and should only be processed if it is not possible to fulfil the purpose by other means.
- Aggregation – Use aggregated data when possible.
- Pseudonymisation – Pseudonymise personal data as soon as it is no longer necessary to have directly identifiable personal data, and store identification keys separately.
- Anonymization and deletion – Where personal data is not, or no longer necessary for the purpose, personal data shall be anonymised or deleted.
- Data flow – The data flow should be made efficient enough to not create more copies than necessary.
- "State of the art" – The controller should apply up to date and appropriate technologies for data avoidance and minimisation.

All these criteria must be applied to the self-encryption solution to determine whether there is data minimisation by design and by default. Based on the available information, the following conclusions can be reached prima facie on data minimisation by design and default in self-encryption:

- Pseudonymisation seems to be used to the largest extent possible, by means of encryption, thus appropriately limiting accessibility. The fact that all data chunks are encrypted along with the keys, except for the key of the last chunk, is in line with data minimisation as the processing of non-encrypted personal data is limited to what is strictly necessary.

[23] European Data Protection Board (EDPB), 'Guidelines 05/2020 on Consent under Regulation 2016/679' (n 16).
[24] European Data Protection Board (EDPB), 'Guidelines 04/2019 on Article a25 Data Protection by Design and by Default' (20 October 2020) <https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf>.

- In self-encryption, there is extensive and continued processing of the IDs of data owners, which are visible to whoever decrypts the data chunks, and the IDs of all those who initiated self-encryption continue to be processed even in subsequent transfers of the data through self-encryption.
  The continued processing of the IDs of data owners may be in compliance with data minimisation, but subject to conditions and with some precautions to take into account:
  - A link should be established between self-encryption and the purposes of the processing. This can be done either by determining that the purpose of the processing is self-encryption per se, or indicating that self-encryption is necessary to ensure a certain level of security, and that this level of security is instrumental for the achievement of the data sharing purposes.
  - The controller must be able to demonstrate at any time that the level of security sought with self-encryption can only be achieved with such continued processing of the IDs in the subsequent transfers, and that there are no less intrusive means available.
  - The IDs are merged into each data chunk, which leads to a multiplication of personal data processing. It must be demonstrated that this is necessary to attain a certain level of security.
  - The data subject(s) must be informed beforehand that the ID or other personal data are to be subject to self-encryption, that they will continue to be processed in subsequent transfers and potentially shared with processors and third parties.

In conclusion, as a practical recommendation, the programmer of the self-encryption software should, when designing the software, make sure that the personal data contained in the ID processed during self-encryption is strictly necessary to achieve the technical effect of self-encryption. This precaution in the design of the software may ensure, or at least facilitate, the minimisation of personal data processing and compliance with Articles 5(1)(c) and 25 of the GDPR.

4.4.3.3 Continued processing of user IDs and data ownership records: compliance with storage limitation

According to the principle of storage limitation of Article 5(1)(e) of the GDPR, the controller must ensure that personal data is kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data is processed. This principle is about the amount of time for which personal data is stored.

The EDPB Guidelines 4/2019 on Article 25 Data Protection by Design and by Default list the following key design and default storage limitation elements:

- Deletion and anonymization – The controller should have clear internal procedures and functionalities for deletion and/or anonymization.
- Effectiveness of anonymization/deletion – The controller shall make sure that it is not possible to re-identify anonymised data or recover deleted data, and should test whether this is possible.
- Automation – Deletion of certain personal data should be automated.
- Storage criteria – The controller shall determine what data and length of storage is necessary for the purpose.
- Justification – The controller shall be able to justify why the period of storage is necessary for the purpose and the personal data in question, and be able to disclose the rationale behind, and legal grounds for the retention period.
- Enforcement of retention policies – The controller should enforce internal retention policies and conduct tests of whether the organization practices its policies.
- Backups/logs – Controllers shall determine what personal data and length of storage is necessary for back-ups and logs.
- Data flow – Controllers should beware of the flow of personal data, and the storage of any copies thereof, and seek to limit their "temporary" storage.

These criteria must be applied to all the processing taking place during self-encryption, including the continued processing of user IDs. Based on the available information, the following conclusions can be reached prima facie on storage limitation by design and default in self-encryption:

- The available information refers to a central server (the audit log?) that keeps track of which file the data chunk belongs to, and who owns the chunk. The storage of ownership information, which would be personal data if referred to a natural person, and any other personal data stored in the central server must be limited in accordance with the criteria mentioned above. The controller may implement automated deletion and must be able to demonstrate at all times that storage is necessary for the processing purposes, and that the data is deleted when it is no longer necessary.
- As already observed for data minimisation, the continued processing of the IDs must be justifiable as necessary for the purposes of the processing, and substantiated proof to this end must be readily available.

In conclusion, as a practical recommendation, any data stored for record-keeping purposes should be either anonymised or deleted when the storage is no longer necessary. Moreover, it is essential that proof can be provided as to the necessity of storing such data as well as to the continued processing of the IDs.

4.4.3.4 Security requirements

One of the general principles governing personal data processing is integrity and confidentiality, of Article 5(1)(f) GDPR, which requires that appropriate security of personal data is ensured during the processing. Moreover, the GDPR lays down security requirements for the processing of personal data in its Article 32, demanding both controllers and processors to put in place appropriate technical and organisational measures to ensure a level of security appropriate to the risks to the protection of personal data.

Both provisions require the implementation of appropriate technical and organisational measures to ensure security of the personal data. The only difference lies in the fact that Article 32 extends the security requirements also to processors, whereas Article 5(1)(f) only applies to controllers. Article 32 is often regarded as a more practical specification of what the principle in Article 5(1)(f) entails, and the two provisions can be intended as imposing the same requirements.

The security obligations of the GDPR impose controllers and processors to, first, gauge the level of risks posed by the processing operation for data subjects and, taking into account the circumstances of the case (including, besides the risks, the state of the art, the characteristics of the processing and the costs of implementation), appropriate technical and organisational measures to ensure a level of security appropriate to the risk. In this regard, it is important to note that this is an obligation of means, and not of results, with the consequence that, as long as risks have been assessed and appropriate measures have been implemented, there will be no infringement of Article 32, even when there is a data breach.

Article 32 of the GDPR provides a non-exhaustive list of measures that could be considered appropriate to ensure a level of security proportionate to the risks. These measures are:

- the pseudonymisation and encryption of personal data.
- the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services.
- the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident.
- a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

To ensure compliance with the security obligations, the controller and the processor shall implement the appropriate measures by design and by default. In this regard, Recital 78 of the GDPR states that such measures could consist of ‘enabling the controller to create and improve security features’. Moreover,

the EDPB has provided guidance on how to comply with the security requirements in its Guidelines on Data Protection by Design and by Default[25]. The Guidelines provide a non-exhaustive list of key design and default integrity and confidentiality elements, i.e.:

- Information security management system (ISMS) – Have an operative means of managing policies and procedures for information security.
- Risk analysis – Assess the risks against the security of personal data by considering the impact on individuals' rights and counter identified risks. For use in risk assessment, develop and maintain a comprehensive, systematic and realistic "threat modelling" and an attack surface analysis of the designed software to reduce attack vectors and opportunities to exploit weak points and vulnerabilities.
- Security by design – Consider security requirements as early as possible in the system design and development and continuously integrate and perform relevant tests.
- Maintenance – Regular review and test software, hardware, systems and services, etc. to uncover vulnerabilities of the systems supporting the processing.
- Access control management – Only the authorized personnel who need to should have access to the personal data necessary for their processing tasks, and the controller should differentiate between access privileges of authorized personnel.
    - Access limitation (agents) – Shape the data processing in a way that a minimal number of people need access to personal data to perform their duties, and limit access accordingly.
    - Access limitation (content) – In the context of each processing operation, limit access to only those attributes per data set that are needed to perform that operation. Moreover, limit access to data pertaining to those data subjects who are in the remit of the respective employee.
    - Access segregation – Shape the data processing in a way that no individual needs comprehensive access to all data collected about a data subject, much less all personal data of a particular category of data subjects.
- Secure transfers – Transfers shall be secured against unauthorized and accidental access and changes.
- Secure storage – Data storage shall be secure from unauthorized access and changes. There should be procedures to assess the risk of centralized or decentralized storage, and what categories of personal data this applies to. Some data may need additional security measures than others or isolation from others.
- Pseudonymization – Personal data and back-ups/logs should be pseudonymised as a security measure to minimise risks of potential data breaches, for example using hashing or encryption.
- Backups/logs – Keep back-ups and logs to the extent necessary for information security, use audit trails and event monitoring as a routine security control. These shall be protected from unauthorised and accidental access and change and reviewed regularly and incidents should be handled promptly.
- Disaster recovery/ business continuity – Address information system disaster recovery and business continuity requirements to restore the availability of personal data following up major incidents.
- Protection according to risk – All categories of personal data should be protected with measures adequate with respect to the risk of a security breach. Data presenting special risks should, when possible, be kept separated from the rest of the personal data.
- Security incident response management – Have in place routines, procedures and resources to detect, contain, handle, report and learn from data breaches.
- Incident management – Controller should have processes in place to handle breaches and incidents, in order to make the processing system more robust. This includes notification

[25]ibid.

procedures, such as management of notification (to the supervisory authority) and information (to data subjects).

Based on the above, the following features of the solution contribute to ensure compliance of self-encryption with the security obligations of the GDPR:

- The threat landscape made by DUT is an essential first step for the assessment of the level of risks posed by the processing operations for data subjects, and it must be taken into account by controllers and processors that will use the solution in the future. The more the risks in the threat landscape will be addressed by DUT, the easier it will be for controllers and processors to comply with the security requirements of the GDPR. Further to the factors listed in the threat landscape, that are specific to the self-encryption solution, controllers and processors will of course need to assess any other risks that arise in the specific circumstances of the processing.
- Before transmission of the data, the data are split in data chunks, hashed and encrypted, leaving only the key to the last data chunk unencrypted. This feature contributes to security of the processing by design and by default, implementing one of the measures explicitly listed in Article 32 and in Guidelines 4/2019 (pseudonymisation and encryption).
- The multi-factor data recovery mechanism, with the division of the key into multiple pieces, is a feature that contributes to the security by design and by default of the transfer and of the storage.
- Records are kept of data chunk ownership and stored in a central server (Audit log?). This feature can contribute the ongoing integrity of the data and to the security by design and by default, as Guidelines 4/2019 explicitly list the implementation of backups/logs.

Moreover, the following recommendations can be provided to ensure security by design and by default when using self-encryption:

- Perform a risk analysis to identify the risks for the security of personal data involved in the processing, and develop a threat landscape.
- The self-encryption solution must be designed in a way that the controller can easily, and at any time, make updates or enhancements, being able to improve its security features whenever needed. The GDPR requires the controller to continuously assess whether it has implemented appropriate measures to ensure security of processing and evaluate whether there are new vulnerabilities to counter. To this end, the controller may have to make changes whenever needed in light of the evolving threat landscape.
- Ensure security of the central server/audit log in the connector if it contains personal data. Should the audit log contain personal data, e.g. through the records on data ownership, appropriate measures must be put in place against potential attacks. The records kept on the audit log contribute to data integrity, and its security is thus instrumental to ensure that the self-encryption as a whole meets GDPR requirements.
- Ensure that the technical features of self-encryption enable the controller to implement policies for security incident response, incident management and disaster recovery.
- Implement appropriate access control management, especially as concerns the data stored in the audit log and the secret key sharing scheme followed for decryption.
- The controller would need in any case to take care of maintenance of the system and consider to put in place an information security management system.
- The controller can consider implementing ISO 27001 to comply with the security requirements of the GDPR.

It is essential to note that the recommendations set out above are not meant to be exhaustive, as the controller of the processing taking place during self-encryption should always assess whether additional measures would be needed in the circumstances of the case. Such assessment would always be based on an appraisal of the specific risks to personal data processing identified in a given situation, including, but not limited to, the threat landscape relating to self-encryption per se.

4.4.3.5 Exercise of data subjects' rights

Pursuant to Article 12(2) of the GDPR, the controller must ensure that data subjects are able to exercise their rights of access, to rectification, to erasure, to restriction of processing, to data portability, to object and not to be subject to automated individual decision-making. These rights are conferred upon data subjects by Articles 15 to 22 of the GDPR.

Given that the controller is accountable for facilitating the exercise of these rights, the controller needs to ensure that the technical and organisational measures are in place to enable data subjects to exercise their rights. This entails that, when the processing takes during self-encryption, the controller is obliged to make sure that the technologies employed for the processing do not hinder compliance with Article 12(2) of the GDPR. As a result, the self-encryption solution must be designed in a way that enables the data controller to comply with any request of data subjects in exercise of their rights.

More particularly, the following can be noted in relation to self-encryption:

- Compliance with access requests should not be hindered by the self-encryption technology. On the one hand, records on data ownership allow to identify the data pertaining to a specific person and, on other hand, there are no technical obstacles preventing the control to grant access to the processed data. In granting access, the controller must ensure that the appropriate security measures are in place.
- Compliance with rectification, erasure, restriction requests, as well as with objections from the data subjects, may be hindered or not by the features of self-encryption depending on whether the continued processing of user IDs for subsequent transfers is strictly necessary for the functioning of the solution. Should it be necessary, the solution me be designed in a way that the modification or erasure of the IDs attached to the data chunks is technically feasible without drastically affecting the operation of the solution.
- Compliance with data portability requests may encounter technical obstacles in the self-encryption solution. In this case, the controller must ensure that the appropriate security measures are in place, managing the risks posed by the fact that the ID would be shared with third parties.

4.4.4 WP3 Protocol compliance and recommendations

As a logical continuation of the assessment of design recommendations and their implementation in WP3, as included in Annex II of D3.1, the following table builds upon those findings and provides the final version of this assessment. Moreover, based on the non-compliant features and/or potential gaps as identified, some recommendations can be provided in order to ensure that the Confidentiality and Privacy by Design technology follows the WP3 Protocol.

Recommendations for technological design	Compliant features	Non-compliant features/Potential gaps	Recommendations
Data mapping (predetermination of the type of data to be processed)	Predetermination of data to be encrypted and shared. No pre-assumptions on the type of data to be processed, the most stringent requirements are assumed.		

Recommendations for technological design	Compliant features	Non-compliant features/Potential gaps	Recommendations
Data obfuscation (disguising the data with the aim to render it unlikable or unobservable)	State of the art encryption (AES-256) and hashing (SHA-256).		Regular updates of encryption algorithms to ensure compliance with evolving security standards.
Data minimisation (limiting the processing of the data)	Minimal exposure of data by file-splitting in chunks.	No encryption/pseudonymisation of metadata.	Metadata encryption/pseudonymisation.
Data abstraction (limiting the detail in which personal data is processed)	Processing at a lower level (binary chunks) rather than full plain text file.	No explicit mention of data abstraction when data is summarised in output files.	Implementation of aggregation techniques for summary files.
Data separation (separating, logically or physically, the processing of data)	Logical and physical separation of data chunks. Secret sharing schemes usage.	Compromise of the threshold system recovery in case where k_min participants are fewer than expected.	Implementation of an adjustable k_min threshold.
Security (protection against unauthorised or unlawful processing or against accidental loss, destruction or damage, of the data)	Robust security via encryption and hashing (AES-256 and SHA-256).	No reference to incident management response in case of data compromise.	Creation of a detailed incident response plans to handle data breaches within FIDES, including logging and monitoring to detect breaches and defined protocols to meet GDPR's 72-hour breach notification requirement.
Control over the data (ability to change the processing conditions of the data at any time during the processing)	Control via parameter specification (e.g. number of data chunks, user ID, etc.).		
Record-keeping and demonstrability (recording of how processing of the data takes place in order to	Separate summary file containing hashing/encryption algorithms, chunk	No specification of logs to full lifecycle of data processing activities.	Regular review and update of logs to demonstrate ongoing compliance.

Recommendations for technological design	Compliant features	Non-compliant features/Potential gaps	Recommendations
demonstrate compliance with the relevant legal framework)	information as audit log for compliance.		

4.5 On the ePrivacy Directive

Following the analysis provided in D3.1 with regard to the ePrivacy Directive, some additional elements are juxtaposed below.

4.5.1 General confidentiality obligation

4.5.1.1 Applicability of the obligation to personal and non-personal data

The ePrivacy Directive, as implemented in the Member States, requires the confidentiality of communications and related traffic data transmitted in public communications networks and through publicly available electronic communications services. This requirement is focused on the confidentiality of communications on certain networks, and applies only to data transmitted in public electronic communications networks and through publicly available electronic communications services.

An important aspect to consider is whether the confidentiality obligation of Article 5(1) of the ePrivacy Directive also applies only to communications containing personal data, or also to communications containing non-personal data. On the one hand, the Directive protects all types of traffic and location data, relating to both natural and legal persons. As a matter of fact, the legitimate interests of legal persons are also protected by the Directive, in relation to multiple provisions that refer to the 'subscriber' and not just the 'user', where the subscriber can be both a natural and legal person, whilst the user can only be a natural person according to the legislative definition in Article 2, letter a), of the ePrivacy Directive. Moreover, the legislative definition of communication and traffic data is broad and seemingly not limited to personal data, as it includes, respectively, any information exchanged or conveyed between a finite number of parties and any data processed for the purpose of the conveyance of a communication. On the other hand, Article 1 of the Directive is explicit in delimiting the scope of application of the Directive to personal data, and such delimitation of the scope is even more explicit when reading the text of Article 3 that describes the services concerned by the Directive. Given the clear literal interpretation that such provisions are conducive to, it would be sensible to interpret all the references to data and communication in the Directive as referring to personal data.

Due to the fact that a strict literal interpretation of the Directive must be preferred, the confidentiality obligation is hereinafter intended as referring only to personal data.

4.5.1.2 Applicability of the obligation to the activities of the platform

In order to determine whether Article 5(1) of the ePrivacy Directive is applicable to the activities of the platform, there are two aspects to consider:

i) if any of the activities of the platform are machine-to-machine services within the meaning of Article 2, point 4), letter c) of the Directive (EU) 2018/1972 establishing the European Electronic Communications Code (hereinafter 'ECCC'),[26] because they consist wholly or mainly in the

[26] Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code (recast) (OJ L 321 17122018, p 36).

conveyance of signals such as transmission services used for the provision of machine-to-machine services and for broadcasting,

ii) ii) if this is the case, whether the services are 'public' within the meaning of the EECC. The two points must be analysed separately.

Based on the current definition of electronic communications service provided in the ECCC, this definition covers also services consisting wholly or mainly in the conveyance of signals such as transmission services used for the provision of machine-to-machine (M2M) services and for broadcasting. As to M2M services, this category covers services consisting wholly or mainly of conveyance of signals between network termination points of electronic communications networks. The EECC does not define further this category but gives two examples: the transmission services used for the provision of machine-to-machine services and for broadcasting. With regard to M2M communications, only the transmission element falls within the scope of the EECC, whereas M2M services at the applications layer remain outside its scope (cfr. Garzaniti). This fine balance ensures the necessary protection at the transmission layer while being open for innovative M2M services, irrespective of the undertaking providing them. Moreover, IoT users are in principle not subject to the rules of the EU regulatory framework.

It is not necessary that the operator itself conveys signals to be considered as a provider of an electronic communications service. What matters is that the operator is responsible vis-à-vis the end-users for transmission of the signal which ensures that they are supplied with the service to which they have subscribed.[27] The interpretation of mainly implies a weighting of the respective value of the elements of the service that are conveyance and that are not. Technical and "functional" characteristics, i.e. demand-side related aspects such as the end user perspective with regard to the OTT-provider's contractual liability vis-à-vis the end-user, may also be relevant.

As to the qualification of a M2M service as ''public'', the EECC does not provide clarification of the circumstances under which a service is considered to be publicly available, but according to the case law[28], a service must be considered publicly available when any part of the public may choose to make use of the service offered. Even if a service is made available only to the subscribers of a particular undertaking, it is considered to be publicly available where there is no limit placed on the number of potential subscribers and any part of the public may, de facto, make use of the service by becoming a subscriber.

The ePrivacy Directive does not further flesh out the confidentiality obligation, which may however be better clarified by national laws. Therefore, at the EU level, it must be intended as a general requirement of confidentiality comparable to that imposed under the GDPR.

It is essential to note that the confidentiality obligation does not cover only the content of the communication, but also the related traffic data. Traffic data is defined by Article 2(1)(b) of the ePrivacy Directive as any data processed for the purpose of the conveyance of a communication on an electronic communications network or for the billing thereof.

As far as traffic and location data are concerned, according to Article 6 of the ePrivacy Directive, traffic data relating to subscribers and users processed and stored by the provider of a public communications network or publicly available electronic communications service must be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication.

[27] *Judgement of 30 April 2014, UPC DTH Sàrl v Nemzeti Média- és Hírközlési Hatóság Elnökhelyettese (C-475/12)* Court of Justice of the European Union (ECLI:EU:C:2014:285).

[28] *Judgment of 22 December 2016, Fjarskipti hf v The Icelandic Post and Telecom Administration (E-6/16)* Court of Justice of the European Union (2017/C 133/06). This reasoning is consistent with recital 55 of the Citizens' Rights Directive and an earlier position of the Commission that "The term 'for the public' … must be understood in its common sense: a service for the public is a service available to all members of the public on the same basis. Particular examples of services which should not be considered 'for the public'… are those provided over corporate networks and/or to closed user groups. Corporate networks and closed user groups (CUGs)":.

In relation to location data other than traffic data, Article 7 of the ePrivacy Directive states that, when it relates to users or subscribers of public communications networks or publicly available electronic communications services, such data may only be processed when they are made anonymous, or with the consent of the users or subscribers to the extent and for the duration necessary for the provision of a value-added service. A value-added service is defined by Article 2(1)(g) of the ePrivacy Directive as any service which requires the processing of traffic data or location data other than traffic data beyond what is necessary for the transmission of a communication or the billing thereof.

4.5.1.3 Applicability to the activities of the platform and interplay with the GDPR

The GDPR and the ePrivacy Directive both prescribe confidentiality and security requirements for the processing of personal data in publicly available electronic communications networks. This leads to an overlap in the material scope of the two legislations, as explicitly recognised by the EDPB[29]

In some cases, the ePrivacy Directive and the GDPR converge towards the imposition of essentially the same requirements, whereas in other cases the ePrivacy Directive supersedes the GDPR by virtue of the principle *lex specialis derogate legit generali*.

One example of where the ePrivacy Directive supersedes, and particularises, the provisions of the GDPR can be found in article 6 of the ePrivacy Directive, which concerns the processing of traffic data. For the processing of traffic data, the lawful grounds provided by article 6 GDPR cannot be applied by the provider of an electronic communications service, because article 6 ePrivacy Directive explicitly limits the conditions in which traffic data, including personal data, may be processed.

Overall, it can be said that the combined framework of the GDPR and the ePrivacy Directive prescribes the same security and confidentiality requirements, except for the processing of traffic data, and location data other than traffic data, for which the ePrivacy Directive lays down stricter requirements.

4.6 Ethical considerations

Privacy is considered a fundamental human right under various international declarations and legal frameworks (e.g., the Universal Declaration of Human Rights and the European Union's General Data Protection Regulation, or GDPR). Privacy-preserving technologies support the notion that individuals should have control over their personal data and its use.

An integral part of the TANGO action is the development of privacy-preserving components which aim to ensure the protection of the personal data of the user, through the provision of information to the users, the implementation of privacy protocols and mechanisms and the ensurance of secure data transfers within the platform. As an addition to the previous section referring to the legal aspects related, with the current section we will provide an overview of the relevant WP3 components on an ethical perspective, both in terms of privacy and socially.

4.6.1 General privacy aspects

Prior to the examination of the matters related to the secure data storage and sharing, we must ensure the compliance of the TANGO platform to the privacy related provisions from its development phase. Under the GDPR, the overall processing operations on personal data shall adhere to the fundamental principles presented in its article 5 and their extent shall be addressed at any case. Starting from the initial processing operation, which is most of the time the collection of data, and throughout its operation, the TANGO platform shall be properly developed in order to adhere to the “privacy by

[29] European Data Protection Board (EDPB), ‘Opinion 05/2019 on the Interplay Between the ePrivacy Directive and the GDPR, in Particular Regarding the Competence, Tasks and Powers of Data Protection Authorities’ (2019) <https://www.edpb.europa.eu/sites/default/files/files/file1/201905_edpb_opinion_eprivacydir_gdpr_interplay_en_0.pdf>.

design" principle and ensure protection of the rights and freedoms of the data subjects and data security and integrity.

In terms of lawfulness of the processing operations, the provision of prior consent shall be examinded as the appropriate legal basis for the processing of personal data, especially in the cases where the platform will be used commercially. During the consent provision process, the data subjects shall thoroughly receive information through the consent form regarding the purpose of the processing operations on their personal data, their right to withdraw from the provided consent at any given time freely, the data controller(s) and if applicable, any data processor(s), the categories of personal data to be processed, the processing operations, the storage period and their data protection rights and how to perform them. Any additional information required by the data subjects shall be provided by the relevant partner(s) during the course of the project related activities. Additionally, consent shall be provided by the data subjects freely, through an intelligible and easily accessible form, using clear and plain language. Due to the accountability principle, the data controller shall be able to demonstrate that the data subject has provided his/her consent according to the provisions foreseen in article 7 of the GDPR.

Apart from the provision of consent, it is envisaged that alternative legal basis for the processing of personal data may be applicable, depending on the nature of the data controllers, the processing operations and the relative service provided through TANGO. For instance, in the case of the VISA provision sector, the relevant partner, as a public authority in charge of the granting procedures of resident permits, it is expected that the relevant legal basis for the processing operations will be "processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller", according to article 6 par. 1e of the GDPR. Additionally, due to the nature of the processing operations and the provided services, a different legal basis may apply to some of the processing operations, for instance in the case of the smart hospitality sector, some processing operations may be based in the legal basis of "processing is necessary for the performance of a contract", while the rest may be based on consent.

In any case, the provision of information to the data subject is considered a matter of outmost importance, both during the development phase of the platform and during the use of the end product. Relevant privacy policies shall be conducted in order to comply to this obligation, according to articles 12-14 of the GDPR. This information shall be provided in a concise,

transparent, intelligible and easily accessible form, using clear and plain language. The data subject shall have enough information to be able to identify the overall processing operations and simultaneously maintain control on his/her data at any given time.

Furthermore, the processing of data of vulnerable people should be examined. This refers to the processing of personal data of people such as children, patients and people under judicial support. Any processing operations on personal data of children shall follow the relevant provisions of the GDPR, especially article 8 of the GDPR, according to which the processing can be deemed lawful only to the extent that consent is provided or authorised by the holder of parental responsibility over the child, according to the nationally applicable laws. In the case of the rest of the cases, especially in terms of the consent acquisition process, the data controllers shall be able to demonstrate that the aforementioned circumstance related to the consent provision process, ensuring that human integrity is respected.

Data minimisation should be also ensured through the use of the TANGO platform. Depending on each service provided, the platform and the organisation using it should request and receive for the data subjects the categories of personal data that are absolutely necessary for the purposes of the processing, following a relevant prior assessment of the data controller's needs and relevance to the processing purpose. This fact does not only refer to the eventually requested personal data but additionally, to the interface, through which the data subject will be providing his/her personal data. The latter should be developed so that the data subject will not be able to provide any additional information than the categories required for the relevant service. That way, the data controllers will avoid unnecessary data

collection and reduce potential risks of data exposure, especially in the cases where data deletion or alteration may not be possible due to the nature of a TANGO component (e.g. the blockchain storage).

In terms of data security, according to article 32 of the GDPR, "all data controllers and processors to implement appropriate technical and organisational measures to ensure a level of data security that is commensurate to the risks faced by the data subjects in the event of unauthorised access to, or disclosure, accidental deletion or destruction of, their data". This refers to the risk-based approach of the GDPR, which links the implemented measures to the potential issues that the controller or the processor may face as a result of the processing operations. The higher the risk (for instance in the case of processing of special categories of personal data, according to article 9 of the GDPR), the more efficient the measures shall be. The data controller shall also provide the necessary information related to the technical and organisational measures implemented to ensure the protection of the personal data.

Moreover, in order to ensure to the purpose limitation principle, the implementation of a deletion or archiving plan shall be examined. Since the processing of personal data is related to a specific purpose that is furthermore linked to the relevant legal basis, by the time the purpose is accomplished and there is no applicable legal basis to ensure lawful processing of personal data, the data controller shall proceed either to the deletion of the personal data or to the implementation of anonymization techniques, in case the anonymized data may be further used in the future. Following their obligation to provide the necessary information to the data subject, the data controllers shall inform the data subjects regarding the potential deletion or implementation of anonymization as the last processing operation, in order to ensure that the data subject retain their right to have control on their data. Since though deletion may not be possible in some cases, especially in the case of storage of personal data on a blockchain platform, anonymization could be considered the best option in order to ensure compliance, given especially the fact that the contents of the blocks on a blockchain are protected through the use of encryption (hash keys).

Finally in terms of the data transfers, it should be noted that the data controllers shall be very careful in the cases of transfers of personal data to a non-EU country. In principle, the data controllers shall refrain from any processing operations as such since issues may arise, in the

cases especially that a non-EU country does not possess a relevant legislation for the protection of personal data. In case though such transfers may be deemed absolutely necessary, in order to ensure compliance with the relevant legislation of the GDPR, the explicit consent of the data subject regarding this transfer shall be acquired and additionally, the data shall be transferred either to a country which has been granted with an "adequacy decision" by the European Commission or to organizations with which the data controller conducts a data-transfer agreement containing EC standard contractual clauses giving effect to EU data protection law or maintains binding corporate rules covering both sender and recipient, approved by a national supervisory authority.

4.6.2 Regarding the blockchain-based Data Storage and Sharing

Through the WP3 action, matters related to the secure data storage and sharing are addressed, highly crucial and sensitive due to their importance as processing operations. As a result of the action of T3.1, a blockchain-based platform will be developed to ensure secure data storage and sharing. Having into consideration the basic principles of the GDPR, the data controllers shall ensure that any data transfers or the storage period shall be limited to the extent required either by the purpose of the processing or any applicable legal obligation.

The use of the blockchain technology is considered a very efficient measure in order to ensure data security, through the use of the encryption and especially in the case of use of private blockchain systems, where only predefined people or organizations have access. Through encryption, the access to the data contained in each block is impossible without the encryption key that could provide access to the contents of the block. On the other hand though, due to the nature of this technology, the exercise of specific rights by the data subjects is deemed impossible, such as the right of erasure or the right of

rectification, in case personal data is contained to the blocks, since the latter cannot be altered in any way following their creation.

Having this into consideration, and as an addition to the aforementioned legal considerations presented above, the use of a blockchain platform for storage and sharing purposes of personal data raises significant privacy issues. Each data controller, who is projected to use this blockchain platform, shall prior procced to the necessary data protection impact assessment, having especially in mind the provision of article 35 of the GDPR, according to which an assessment is required especially whenever "a type of processing in particular is using new technologies". As a first step, the developer of this component shall identify whether information stored in the blockchain could be considered personal data, having in mind the definition of personal data provided in article 4 case 1 of the GDPR ("personal data means any information relating to an identified or identifiable natural person") and the potential identifiability of a data subject by any third party through the information stored in it. This assessment is mostly technical and shall be conducted by the developer of the platform in order to identify and examine potential issues in terms of the potential identifiability of the data subjects, according to their obligation to demonstrate their compliance to the provisions of the GDPR.

Moreover, following this assessment and in case it is identified that personal data will be stored in the FIDES platform, the end users shall be thoroughly informed in a clear and plain language, understandable to them, regarding the potential processing of their personal data through the use of a component which utilises the blockchain technology. The end users should be especially informed about the implications relevant to the processing of their personal data through the use of a blockchain-based storage environment, with reference especially to their potential inability to exercise some of their rights.

It should be also noted that even the prior provision of consent by the data subjects is not sufficient in this case since the data subject will be also unable to successfully withdraw from it, provided that the blocks cannot be altered. Having this into consideration, the end users shall be provided with the option to request the exclusion of the FIDES component during the

use of the TANGO platform and the use of alternative means for the purposes of storage and transfer of their personal data. In the contrary, in case this option is not provided to them, the end users may feel forced to accept the processing operations through this platform, thus directly affecting the freedom of provision of consent or the maintenance of control of the data subjects on their personal data.

To conclude, the secure option in this case is the storage of data in the blockchain platform that are not considered personal data, either as a result of their nature or their level of anonymity. The developer of this component, alongside the organizations that will be using the end product in order to provide their services, shall proceed to a relevant data protection impact assessment according to articles 35 and 36 of the GDPR and even consult in this case their regional supervisory authority regarding this matter.

5 Conclusions

This deliverable outlines the final outcomes of the software components developed within the framework of TANGO WP3 tasks. In addition to outlining the overarching vision and goals of WP3, the document elucidates the current status of tool implementation for AI-related activities in each task.

For each task and component, the document furnishes a comprehensive overview of the implemented features, internal architecture, expected support for pilots, associated software artifacts, potential future work, and references to relevant software/demos. A summary of the conclusion for each task and component is provided in the subsequent table.

T3.1 highlights the platform's readiness through its integration into pilot use cases, including public organizations and financial institutions. The advancements achieved since D3.1 reflect a focused effort to enhance functionality, scalability and compliance. As TANGO progresses, Fides will continue to evolve, incorporating insights from pilot activities and addressing emerging challenges in privacy and security.

T3.2 reported the design of three components of task T3.2, considerably progressed towards availability for the use cases in the recent period. Each of them provides its functionality and is ready for integration while further alignment with the use cases may be needed for final validation.

T3.3 provides a robust and privacy-preserving approach to data access management. By enforcing strict policy-based control and embedding security within the data itself, this dual-layered mechanism ensures confidentiality, integrity, and privacy. Even after data is shared, access control policies remain effective, making this solution well-suited for secure and distributed environments.

T3.4 successfully developed and integrated self-encryption, decryption, and multi-factor information recovery mechanisms, ensuring secure, efficient, and scalable data management while enhancing usability for both technical and non-technical users.

T3.5 investigated the techniques involved in work package 3 and draw the conclusion that storing non-personal or anonymized data on a blockchain platform is the secure option, requiring a data protection impact assessment and potential consultation with regional supervisory authorities to ensure GDPR compliance.

In conclusion, this deliverable presents the final outcomes of the software components developed within TANGO WP3, detailing their implementation, architecture, and potential future enhancements. Each task has contributed to the overarching goal of providing secure, scalable, and efficient AI-driven solutions. Collectively, these developments ensure that TANGO WP3 is well-positioned to support pilot use cases and address emerging challenges in data security, privacy, and AI integration.

6 References

[1] Article 29 Data Protection Working Party, 'Opinion 05/2014 Anonymisation Techniques' (2014)

[2] European Data Protection Board (EDPB), 'Opinion 05/2019 on the Interplay Between the ePrivacy Directive and the GDPR, in Particular Regarding the Competence, Tasks and Powers of Data Protection Authorities' (2019) <https://www.edpb.europa.eu/sites/default/files/files/file1/201905_edpb_opinion_eprivacydir_gdpr_interplay_en_0.pdf>

[3] European Data Protection Board (EDPB), 'Guidelines 05/2020 on Consent under Regulation 2016/679' (2020) <https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf>

[4] European Data Protection Board (EDPB), 'Guidelines 04/2019 on Article 25 Data Protection by Design and by Default' (20 October 2020) <https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf>

[5] European Data Protection Board (EDPB), 'Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR' (2021) <https://www.edpb.europa.eu/system/files/2023-10/EDPB_guidelines_202007_controllerprocessor_final_en.pdf> accessed 27 January 2025

[6] European Data Protection Supervisor (EDPS), 'Opinion 09/2016 on Personal Information Management Systems' (2016)

[7] Felten E, 'Does Hashing Make Data "Anonymous"?' (*Federal Trade Commission*, 22 April 2012) <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2012/04/does-hashing-make-data-anonymous> accessed 27 January 2025

[8] Finck M, 'Cobwebs of Control: The Two Imaginations of the Data Controller in Eu Law' (2021) 11 International Data Privacy Law 333

[9] Jaap-Henk Hoepman, 'Privacy Design Strategies (The Little Blue Book)' (2022)

[10] *Judgement of 5 June 2018, Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH (C-210/16)* (Court of Justice of the European Union)

[11] *Judgement of 29 July 2019, Fashion ID GmbH & Co KG v Verbraucherzentrale NRW eV (C-40/17)* [2019] Court of Justice of the European Union (ECLI:EU:C:2019:629)

[12] *Judgement of 30 April 2014, UPC DTH Sàrl v Nemzeti Média- és Hírközlési Hatóság Elnökhelyettese (C-475/12)* Court of Justice of the European Union (ECLI:EU:C:2014:285)

[13] *Judgment of 22 December 2016, Fjarskipti hf v The Icelandic Post and Telecom Administration (E-6/16)* Court of Justice of the European Union (2017/C 133/06)

[14] 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) 2002

[15] Directive (EU) 2016/ 943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-How and Business Information (trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure (OJ L 157/1, 15/6/16)

[16] Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code (recast) (OJ L 321 17122018, p 36)

[17] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) 2016 (OJ L 119/1, 4/5/16)

[18] Regulation (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework (OJ L, 3042024)

Annexes

A.1 Encryption API of SEDSS

The user can access the encryption API via HTTP POST. Our Encryption API designed is described in Table 5.

Table 5 The description of Encryption API design.

Property	Description
n_chunks	the number of data chunks to be generated after encryption. (the user will get n+1 data chunks, the extra one will be used to store necessary encryption information. So if n_chunks=3, you will get 4 chunks in the HTTP response).
n_secret_pieces	The number of key parts to be generated. Classical encryption methods generate only 1 key after encryption. In SEDSS, the key is divided into several pieces for decryption. Presenting a certain number of pieces results in successful key recovery.
security_level	The minimum number of key pieces required to recover the original key. For example, if n_secret_pieces=3 and security_level=2, then any two of the three key pieces will result in a successful key recovery, and therefore a successful data decryption.
userName	the name of the data owner (string, with a maximum length of 16bytes). This info is optional. If pilot users do not wish to use it, please fill in any placeholder here (such as N.A.).
userID	the ID of the data owner (string, with a maximum length of 16bytes). This info is optional. If pilot users do not wish to use it, please fill in any placeholder here (such as N.A.).
useLocalStorage	False (a string). If we eventually need to store the encrypted data and key in the SEDSS docker image, we switch to True (or true, TRUE).
sharingUnits	optional. This parameter is intended for property-based encryption. For example, after generating 3 key pieces, users do not want to remember them. Then the key pieces will be mapped to some properties (e.g. apple, banana and kiwi). In the decryption process, if the three properties are presented to the decryption algorithm, the decryption will be successful. Since pilot users have no need for this feature, and we don't have space to store the mapping table at the moment, please leave this parameter empty.
file	The file for encryption.

Returned JSON:

- transaction_number: the unique number (string) for the encryption instance.

- file-x: the content of the x-th encrypted data chunk
- filename-x: the filename of the x-th encrypted data chunk
- useLocalStroage: Should always be 'False' if we can find a reasonable way to store data rather than storing in SEDSS docker.
- n_chunks: the number of encrypted chunks
- key-y: the y-th key piece

A.2 Decryption API of SEDSS

To use the decryption API, the user is supposed to send a HTTP POST request with a json file containing the following required information and file(s):

- transaction_number
- file-x: the x-th encrypted data chunk
- filename-x: the filename of x-th encrypted data
- useLocalStorage: 'False'
- n_chunks: the same as that returned by the encryption process (i.e., the number of chunks included in the json)
- key-y: the y-th key
- userID: the id of user who intend to decrypt the data. Can be N.A. if not needed.
- UserName: same as userID, can be N.A. if not needed.
- n_keys: number of keys for decryption.

The returned JSON contains the following data:

- transaction_number
- decryptedFile: the content of decrypted file. To access the decrypted file, please decode this string with base64decoder and write to local disk in binary mode.
- decryptedFileName: the name and extension name of decrypted file
- re_enc_key-x: the x-th key pieces of regenerated encrypted data
- re_enc_file-y: the y-th regenerated encrypted data chunk
- re_enc_filename-y: the file name of y-th regenerated encrypted data chunk
- user_history: the name and ID of users who previously encrypt this file. Not sure if this function will be used for any pilots.

A.3 On the Encryption and Decryption Efficiency of SEDSS for Real world Data

In this section, we present the analysis of the encryption and decryption times for four typical types of files with varying sizes. The files tested include TXT, CSV, PDF, and MP4 files, and the measurements were taken using an encryption-decryption algorithm in SEDSS. The results are depicted in Figure 20, Figure 21, Figure 22 and Figure 23 respectively. The analysis aims to provide insights into the performance and efficiency of the algorithm for different file types and sizes.

The encryption and decryption times for TXT files were measured across different file sizes ranging from 1 MB to 2 GB. For smaller file sizes (1 MB to 100 MB), the encryption times showed slight variations but generally remained under 6 seconds. As file sizes increased to 1 GB and 2 GB, the encryption times increased significantly to approximately 10.5 seconds and 21.5 seconds, respectively. Similarly, decryption times for smaller files were close to the encryption times and followed a similar pattern of increase with file size, reaching up to approximately 23 seconds for the largest file size tested.

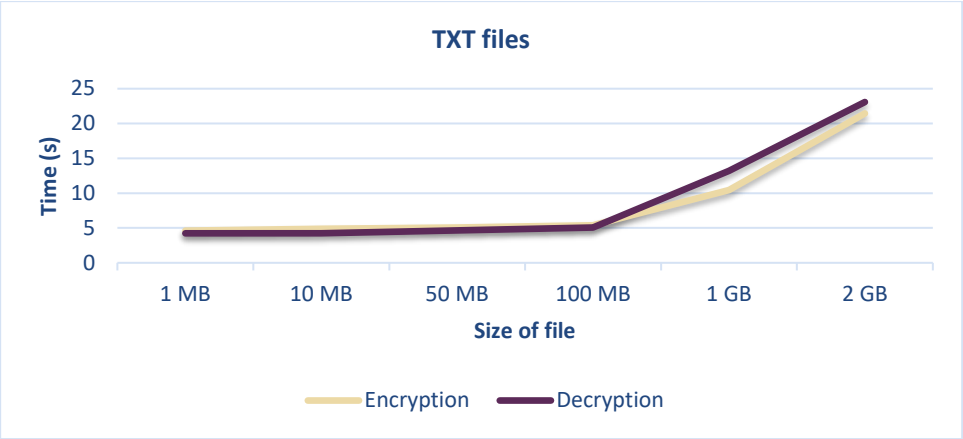

Figure 20 Encryption and decryption times for TXT files of varying sizes, ranging from 1 MB to 2 GB.

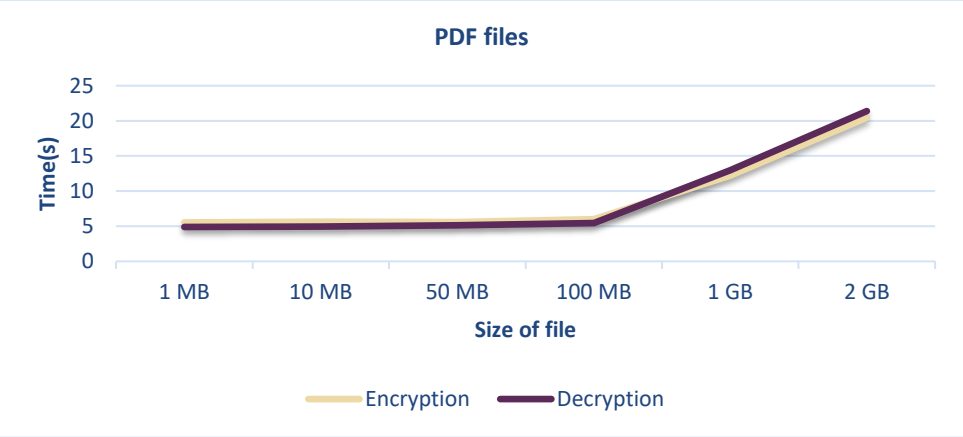

Figure 21 Encryption and decryption times for PDF files of varying sizes, ranging from 1 MB to 2 GB.

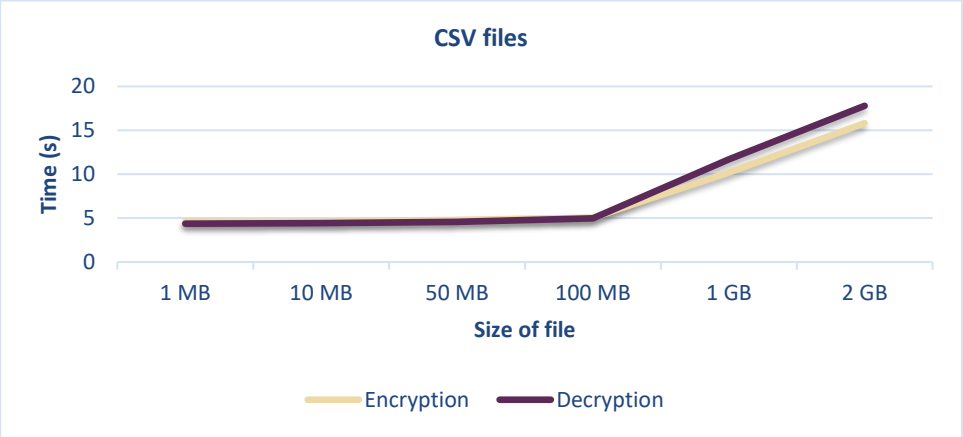

Figure 22 Encryption and decryption times for CSV files of varying sizes, ranging from 1 MB to 2 GB.

Encryption and decryption of PDF files followed a slightly different trend. For smaller file sizes, encryption times were higher than for other file types, with times around 5.5 seconds for 1 MB files and increasing up to approximately 6 seconds for 100 MB files. Decryption times were slightly lower but followed a similar upward trend. For 1 GB files, encryption times were approximately 12.2 seconds,

and for 2 GB files, the times were around 20.5 seconds. Decryption times for these large files were slightly higher, peaking at around 21.4 seconds.

The analysis of CSV files revealed that the encryption and decryption times exhibited a consistent pattern. For file sizes between 1 MB and 100 MB, the encryption times ranged from around 4.6 seconds to 5 seconds, while decryption times were slightly lower, staying under 5 seconds. As the file size grew to 1 GB, the encryption and decryption times increased significantly, with encryption times reaching up to around 10.2 seconds and decryption times peaking at approximately 11.7 seconds. The largest file size of 2 GB resulted in encryption and decryption times of approximately 15.8 seconds and 17.9 seconds, respectively.

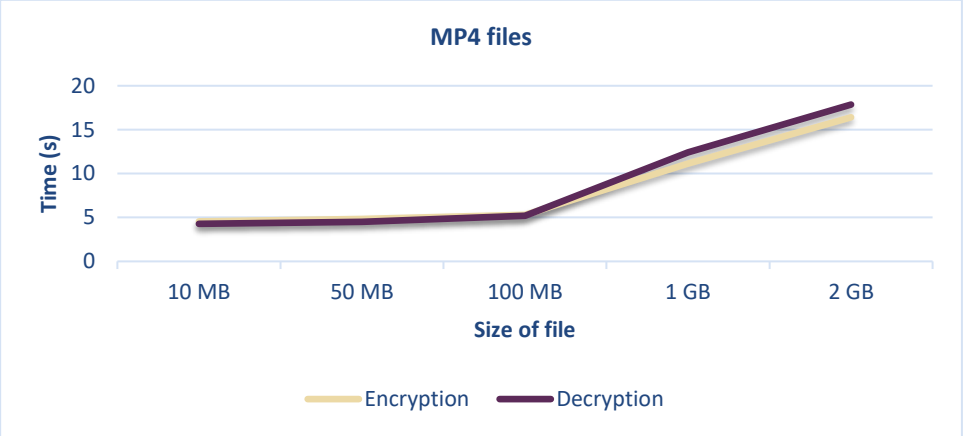

Figure 23 Encryption and decryption times for MP4 files of varying sizes, ranging from 1 MB to 2 GB.

MP4 files exhibited a different behavior, particularly for larger file sizes. The encryption and decryption times for smaller MP4 files (1 MB to 100 MB) were relatively close, ranging from approximately 4.6 seconds to 5.3 seconds for encryption and slightly lower for decryption. However, for the 1 GB file, encryption times were around 11.1 seconds and increased to approximately 16.4 seconds for the 2 GB file. Decryption times followed a similar pattern, with the 1 GB file taking around 12.4 seconds and the 2 GB file peaking at approximately 17.9 seconds.

These results illustrate the correlation between file size and encryption/decryption times across different file types, with processing times increasing as file size grows. Variations are observed based on the specific characteristics of each file type. However, since pilots in TANGO rarely need to encrypt very large files (e.g., those exceeding 100 MB), the SEDSS remains efficient for most use cases within the project. Furthermore, the distributed storage implemented in D3.1 enables large files to be divided into smaller data chunks for encryption, ensuring optimal efficiency. This analysis offers valuable insights into the performance and scalability of the encryption-decryption algorithm for handling different file sizes and formats.