D3.1 Distributed Privacy-preserving Data Management and Storage Intermediate Version

Document Identification			
Status	Final	**Due Date**	31/01/2024
Version	1.0	**Submission Date**	31/01/2024

Related WP	WP3	**Document Reference**	D3.1
Related Deliverable(s)	Insert Related Deliverables	**Dissemination Level**	PU
Lead Participant	DUT	**Lead Author**	Kaitai Liang Dazhuang Liu
Contributors	NOR, DUT, UTH, UMU, VTT, FUJ_GE, KUL, DBC, FHG	**Reviewers**	Antonios Chronakis(SVI) Vitalii Demianets (NOR)

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

Document Information

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

Document History			
Version	Date	Change editors	Changes
0.1	15/11/2023	Kaitai Liang (DUT) Dazhuang Liu (DUT)	Table of Contents to be approved
0.2	15/01/2024	Kaitai Liang (DUT) Dazhuang Liu (DUT)	Merge the drafted deliverable of each component
	19/01/2024	Dazhuang Liu (DUT)	Reformat the document and check missing information
0.3	22/01/2024	Kaitai Liang (DUT)	Polish the document
0.4	23/01/2024	Dazhuang Liu (DUT)	Version ready for internal review
0.5	24/01/2024	Dazhuang Liu (DUT)	Update based on internal review feedback
1.0	31/01/2024	Kaitai Liang (TUD)	FINAL VERSION TO BE SUBMITTED

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	Kaitai Liang (DUT)	30/01/2024
Quality manager	Jürgen Neises (FSDE)	30/01/2024
Project Coordinator	Tomás Pariente Lobo (ATOS)	31/01/2024

Table of Contents

Document Information 2
Table of Contents 3
List of Tables 5
List of Figures 6
List of Acronyms 7
Executive Summary 8
1 Introduction 9
1.1 Purpose of the document 9
1.2 Relation to other project work 9
1.3 Structure of the document 9
2 Overview of WP3 10
2.1 Overview of WP3 main components 11
3 Description of components 13
3.1 Blockchain-based Data Storage and Sharing [T3.1] 13
3.1.1 Introduction 13
3.1.2 Current development 13
3.2 Trustworthy Data Sharing [T3.2] 18
3.2.1 Introduction 18
3.2.2 Usage Control – Privacy Risk Scoring 20
3.2.3 Trustworthiness Scoring 22
3.2.4 Ubiquitous Personal Context Vectors (UPCVs) 26
3.3 Confidentiality and Privacy by Design [T3.3] 30
3.3.1 Introduction 30
3.3.2 Current development 30
3.4 Selft-encryption and Decryption Techniques with Multi-Factor Information Recovery Mechanisms [T3.4] 35
3.4.1 Introduction 35
3.4.2 Self-encryption and decryption 35
3.5 Recommendations for secure and privacy-preserving data storage and sharing [T3.5] 42
3.5.1 Purpose of the recommendations 42
3.5.2 Scope and structure of the recommendations 43
3.5.3 Mapping of the data 44
3.5.4 Data obfuscation 45
3.5.5 Data minimisation 51
3.5.6 Data abstraction 54
3.5.7 Data separation 54

3.5.8 Security of data processing 55
3.5.9 Control over the data 57
3.5.10 Record-keeping and demonstrability 58
4 Conclusions 59
5 Bibliography 61
6 Annexes 62
Annex I – OVERVIEW OF DESIGN RECOMMENDATIONS 62
Annex II - ASSESSMENT OF DESIGN RECOMMENDATIONS IMPLEMENTATION IN WP3 TECHNOLOGIES 63
Annex III – Legal framework relevant to T3.5 65

List of Tables

Table 1: WP3 components _____ *11*
Table 2 API of Decryption module _____ *41*
Table 3 API of Encryption module _____ *41*

List of Figures

Figure 1 WP3 components relation 10
Figure 2 Fides Design overview 14
Figure 3 DRILL data flow 14
Figure 4 DnAMS data flow. 15
Figure 5 Fides implementation for data sharing use case (here shown for KYC/KYB data) 16
Figure 6 Fides institution A example 17
Figure 7 Fides institution B example 18
Figure 8 IIC Trustworthiness Radar 19
Figure 9 Architecture of the Privacy Risk Module 21
Figure 10 Singling Out Risk Algorithm 22
Figure 11 TSM Modules 23
Figure 12 Trustworthiness configuration 24
Figure 13 Swagger API documentation of TSM 26
Figure 14 User interest and exchange of Volatile Random Numbers by the UBEM 27
Figure 15 User Recommendations 27
Figure 16 Synchronizing VRNs across vendors 28
Figure 17 Recommendation process 28
Figure 18 VRN exchange process after showing interest in an item. 29
Figure 19. Service Authorization process within the TANGO (FIWARE connector) architecture. 31
Figure 20 ABE module within FIWARE connector reference architecture 32
Figure 21 ABE toolset processes 35
Figure 22 The internal architecture and dataflow of self-encryption and decryption 37
Figure 23 The workflow of encryption module 37
Figure 24 The workflow of decryption module 38
Figure 25 The workflow of the multi-factor information sharing mechanism. 39
Figure 26 Map factors to participants or features for factor sharing. 39
Figure 27 Recover the original key to the encrypted data from given key pieces. 40

List of Acronyms

Abbreviation / acronym	Description
ABE	Attribute-based Encryption
AES	Advanced Encryption Standard
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
M2M	Machine to Machine
MFIR	Multi-Factor Information Recovery
IIC	Industrial Internet Consortium
OKVS	Oblivious Key-Value Storage
P2P	Peer-to-peer
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PIP	Policy Information Point
RRI	Robot Revolution Initiative
SaaS	Software as a Service
SED	Self-Encryption and Decryption
SSI	Self-Sovereign Identity
TSD	Trade Secrets Directive
TSM	Trsutworthiness 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

This deliverable signifies the initial phase of developing software components within the framework of TANGO WP3, dedicated to Distributed Privacy-preserving Data Management and Storage. The document meticulously details the progress achieved across the five tasks within the work package, emphasizing the implementation aspects, and elucidates the primary components and features that have been successfully implemented to date. Accompanying this documentation are pertinent software artifacts and demos intended for seamless integration within the diverse architecture layers of the TANGO platform under WP3.

The document initiates by presenting a comprehensive overview of the primary vision behind the work package, elucidating the overarching objectives of this release. The emphasis is largely placed on delivering an initial set of components designed for testing across diverse pilot scenarios.

Secondly, the document offers a detailed breakdown for each task and component, encompassing the description of the component, its internal architecture, features implemented to date, anticipated support for pilots, a depiction of the associated software artifacts, and potential avenues for future enhancements in subsequent releases. This comprehensive information is presented for all tools developed within WP3, including blockchain-based data storage and sharing, tokenization techniques for trustworthy data sharing, confidentiality and privacy by design, self-encryption and decryption techniques featuring a multi-factor information recovery mechanism, and lastly, ePrivacy mechanisms, protocols, and processes. Where applicable, the document includes pointers to software and demos, with the expectation that these tools will seamlessly integrate into the TANGO platform, thereby becoming accessible for other components, pilots, and end-users.

Finally, the document concludes with a summary of the work done so far and the work to be done for the integration of the tools and the remaining aspects to be covered in following releases.

1 Introduction

1.1 Purpose of the document

The objective of this deliverable is to present an intermediate release, serving as a demonstrator, featuring functional implementations of components developed within the framework of WP3 – Distributed Privacy-preserving Data Management and Storage. This document, categorized as a demonstrator, offers the initial outputs in terms of the implementation of all tasks outlined in WP3. It is complemented by the inclusion of software prototypes developed up to M17 of the project, marking January 2024.

This document is in consonance with the overarching objectives delineated in the Description of Action (DoA) and the project roadmap. Concentrating on pivotal elements such as data sharing, trustworthiness, confidentiality, distributed data storage, privacy, ePrivacy, and energy efficiency, the document thoroughly explores the intricacies and challenges associated with these essential facets of data management.

Within the document, each task uniformly presents its outcomes to ensure a streamlined approach, offering a swift overview of the results. This standardized reporting method is designed to facilitate readers in easily locating the implemented features, accompanied by additional technical details and their relevance to pilot initiatives.

The document delineates the present perspective regarding future work for each component, recognizing that this outlook may necessitate adjustments based on feedback garnered from integration processes, pilot initiatives, and user experiences in the forthcoming months.

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 several key sections. Chapter 1 outlines the purpose, relation to other project work, and overall structure. Chapter 2 details the vision of WP3, its interrelation among components, and its placement in the TANGO architecture, featuring a figure illustrating component relationships. Chapter 3 lists the objectives of WP3 and assesses their achievement in this release. In Chapter 4, relevant provisions of the Data Governance Act and Trade Secrets Directive are covered, including their applicability to TANGO and the 'reasonable steps' requirement in the TSD. Chapter 5 summarizes main conclusions, addressing challenges, results, and tasks requiring input, while outlining the next project steps. Lastly, Chapter 6 provides an executive summary, offering a concise overview of the document's content, key findings, and conclusions.

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:

- To create a decentralised data-sharing ecosystem using blockchain technology
- To design and develop of a dynamically configurable trustworthiness module
- To assess the security context through risk levels for data sharing or storing
- To allow users to maintain ownership throughout the whole lifecycle of the data sharing process
- To protect data confidentiality and achieve data traceability through distributed encryption/decryption
- To expand the notion of confidentiality to non-personal data.

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

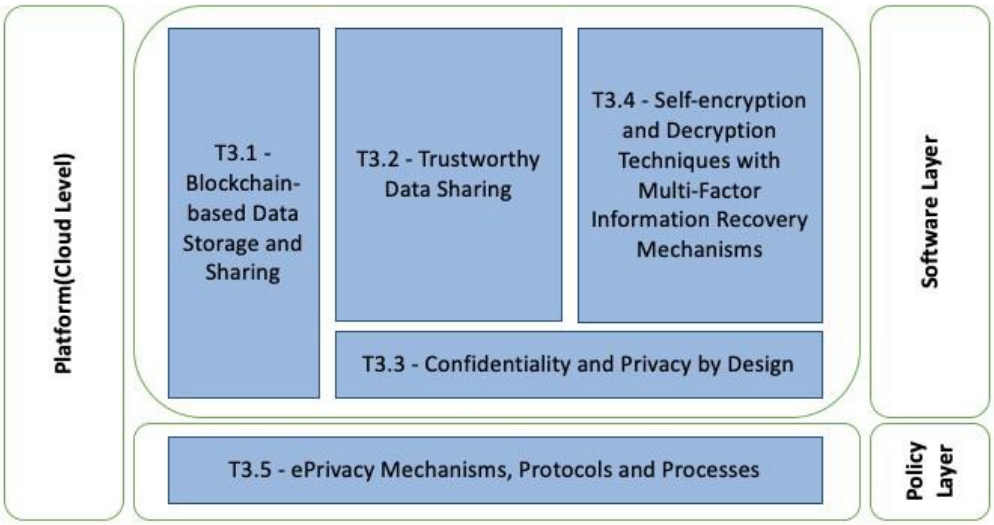

Figure 1 WP3 components relation

2.1 Overview of WP3 main components

The main components of WP3.

Table 1: WP3 components

Task	Component name	Short description
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.
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.
	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 Behavior 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).
	Privacy risk score	The last sub-task is to establish a technical privacy risk score for the data to be shared, based on simulated 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.
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.
	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.
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.

Task	Component name	Short description
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.

In the next sections, the document provides a detailed view of the current status of the implementation of these components for the intermediate release.

3 Description of components

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

3.1.1 Introduction

norbloc is a Regtech company with HQ in Sweden, offering solutions (software) for data sharing and onboarding. Our solution Fides, which is a data-sharing platform, will be utilized in Tango, in Pilot 5 (T7.6 – Public organizations) to share VISA data between VISARIGHT and one or more German authorities.

3.1.2 Current development

The Fides platform offers decentralized storage of regulated data / attestations together with verification stamps and sharing stored data across interested parties.

The full solution is truly decentralized and offers legal compliance, great data redundancy, system resilience and enforced data verification upon entry with support for independent data verification entities.

Fides is a holistic solution which can be implemented independently or can be used as a part of another project. The latter is the case in the realm of TANGO project; in this case, some of the internal components of Fides will not be fully utilized as their functions will be either redundant or irrelevant to the goals of TANGO. Every such case of reduced utilization is indicated in the description of the respective internal Fides component (please see more details in D2.3).

3.1.2.1 Short description of the component

Fides functionalities that are exploitable by TANGO project include:

- **Explicit consent management.** This functionality is delivered by DnAMS internal component, see section 3.1.1.1.2, In the scope of TANGO project other components requiring reliable decentralized data storage will be issuing “proofs of consent” / “data access permission proofs” to facilitate data sharing to the data consumers.
- **Private data sharing in regulated environments.** This functionality is delivered by Fides design. TANGO project will include Fides component for the use cases that require data sharing in regulated environments and include two or more independent parties. In other use cases where there are no strict privacy regulations, or the data pieces being shared do not constitute private data, or there is only one independent party – other TANGO components shall be used.
- **Immutable audit logs.** This functionality is delivered by blockchain Fides component. A set of metadata for every data access is immutable stored, ready for external/internal audit.

3.1.2.2 Internal architecture

Figure 2 presents 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 – fully utilized in TANGO
- Proprietary norbloc access management system: DnAMS – partially utilized in TANGO
- Blockchain (any private blockchain can be used). By default, Fides offers Hyperledger
- Fabric blockchain – fully utilized in TANGO

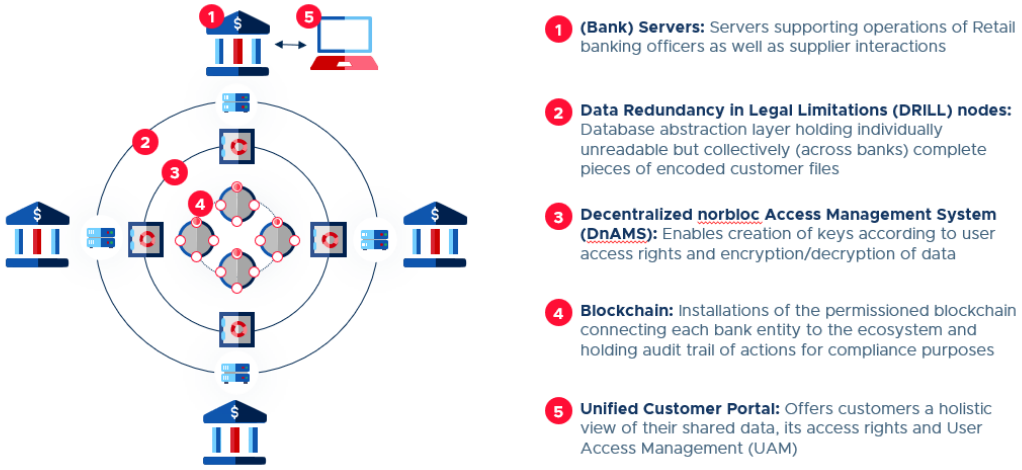

Figure 2 Fides Design overview

3.1.2.2.1 Internal component: DRILL

In essence DRILL is a peer-to-peer (P2P) network comprising DRILL nodes, which are an essential part of every Fides node. DRILL is designed to store individual data pieces in an environment where conformance to privacy regulations is an essential requirement. To facilitate privacy regulations, DRILL includes symmetric data encryption as a part of its design.

Figure 3 describes the main steps of DRILL data flow. Each piece of data is independently encrypted, split in several chunks and those chunks are distributed between DRILL peers in such a way that no one of the peers stores all chunks. Therefore no one of the peers is storing private data from a GDPR perspective, as individual chunks are unintelligible.

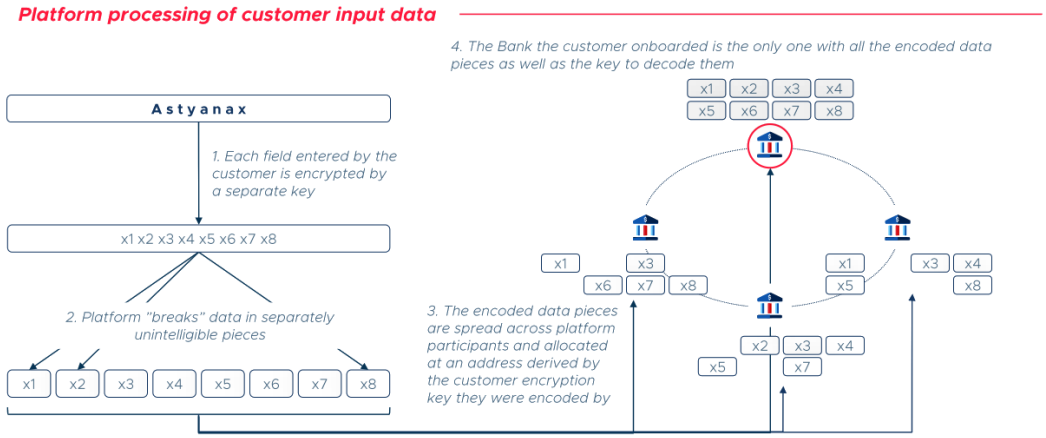

Figure 3 DRILL data flow

When DnAMS component is presented with a "data access permission proof" it releases the necessary information from blockchain storage to a DRILL peer, which allows the DRILL peer to retrieve all chunks from other peers and reconstruct the full encrypted data. Then DnAMS component on the basis of the same "data access permission proof" releases the encryption key to decrypt the reconstructed encrypted data.

3.1.2.2.2 Internal component: DnAMS

DnAMS is a decentralized access management system. It manages symmetric data keys used by DRILL encryption of individual data pieces; also, it manages private keys of data subjects, thus removing a need in maintaining separate external wallets. DnAMS manages access rights by managing respective

encryption keys. DnAMS allows release of a key only upon being presented with a valid “data access permission proof” ("proof of consent”).
Figure 4 describes the main steps of DnAMS data flow.

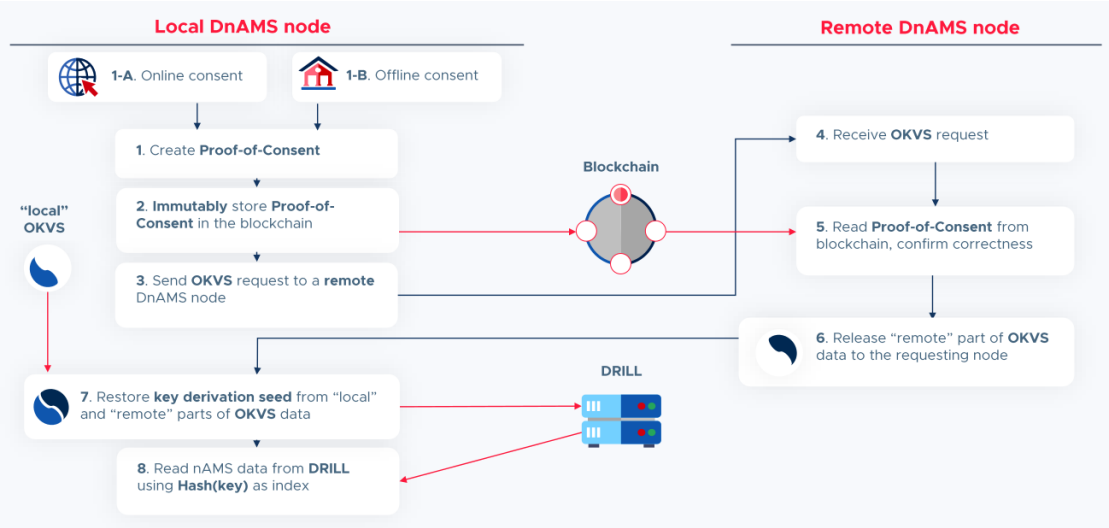

Figure 4 DnAMS data flow.

Features that distinguish DnAMS from other key management/access management systems are:

- Explicit consent management (see Figure 4).
- True decentralization – all DnAMS node are of the same rank, all of them together form a flat P2P DnAMS network.
- Local obliviousness: By exploring only local data storage it is impossible to deduce any knowledge about keys or data stored, including even simple boolean fact if a certain user data is present in the system or not. This is achieved by utilizing norbloc proprietary technology OKVS – Oblivious Key-Value Storage.

3.1.2.2.3 Internal component: Blockchain

A blockchain is a helper component that brings in such properties as immutable storage, immutable audit log and reliable timestamping. It is heavily used by other components; DnAMS and DRILL:

- DRILL stores in blockchain data hashes, thus ensuring data integrity;
- DnAMS stores in blockchain “proofs of consent” / “data access permission proofs”, preventing certain Fides participants form manipulating the proofs to their advantage.

3.1.2.3 Features implemented

Fides platform will be instantiated as a set of Fides nodes. These nodes will be installed at each participant’s premises. For the case of TANGO pilots, it will be considered to install the nodes at the hosting environment provided by INTRA TANGO partner. The Fides nodes will be communicating with each other, thus creating three peer-to-peer (P2P) networks shown in Figure 5 below:

- DRILL P2P: decentralized distributed legally compliant data backend
- DnAMS P2P: decentralized access management system
- Blockchain P2P: decentralized blockchain – a glue bringing properties of immutability and reliable timestamping to the system.

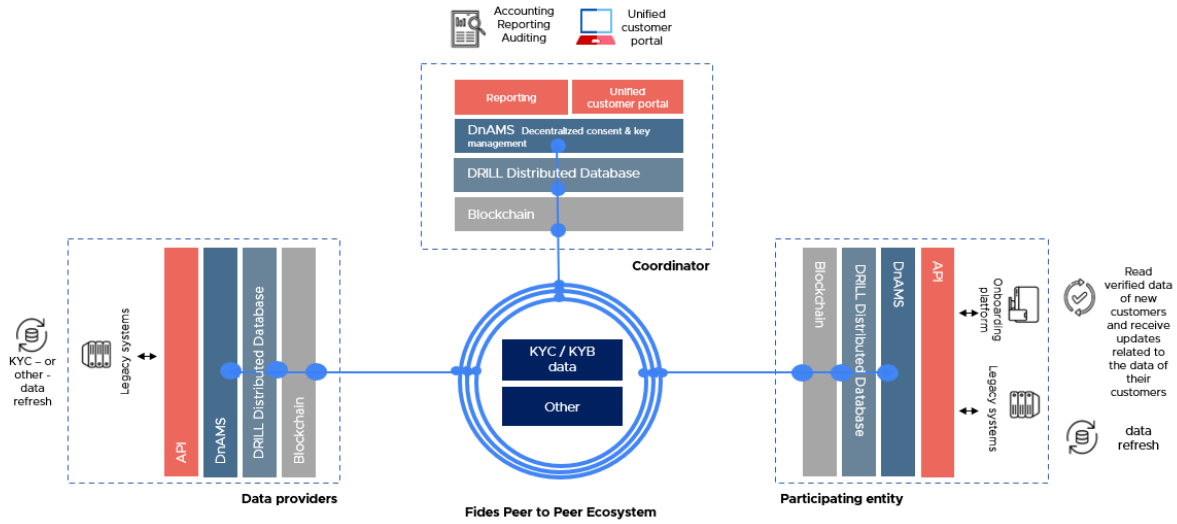

Figure 5 Fides implementation for data sharing use case (here shown for KYC/KYB data)

3.1.2.4 Software artifacts

Fides is delivered in the form of docker containers, designed to be run under Kubernetes cluster orchestration. Fides is a closed-source software, so no source code is going to be exposed.
The list of artifacts includes both docker images and Kubernetes configuration files (using YAML language):

Docker images

- Fides-3 (A main application offering API connectivity);
- BC-connector (A blockchain connector image);
- DnAMS (A DnAMS node image);
- DRILL (A DRILL node image);
- Fabric-orderer (A Hyperledger Fabric orderer node blockchain image);
- Fabric-peer (A Hyperledger Fabric peer node blockchain image);
- Crypto-svc (A helper image offloading encryption/decryption tasks).

Kubernetes YAML configuration files

- Fides-3:
  – fides-3-deployment.yaml (A Deployment config)
  – fides-3-configmap.yaml (A ConfigMap config)
  – fides-3-cluster-ip-service.yaml (A Service config of type “ClusterIP”)
  – fides-3-ingress.yaml (An Ingress config)
- BC-connector:
  – bc-connector-deployment.yaml (A Deployment config)
  – bc-connector-configmap.yaml (A ConfigMap config)
  – bc-connector-cluster-ip-service.yaml (A Service config of type “ClusterIP”)
- DnAMS:
  – dnams-deployment.yaml (A Deployment config)
  – dnams-configmap.yaml (A ConfigMap config)
  – dnams-ip-service.yaml (A Service config of type “ClusterIP”)
- DRILL:
  – drill-deployment.yaml (A Deployment config)
  – drill-configmap.yaml (A ConfigMap config)
  – drill-cluster-ip-service.yaml (A Service config of type “ClusterIP”)
  – drill-ingress.yaml (An Ingress config)
  – drill-persistent-volume-claim.yaml (A Persistent Volume Claim config)
- Fabric-orderer:

– fabric-orderer-deployment.yaml (A Deployment config)
– fabric-orderer-configmap.yaml (A ConfigMap config)
– fabric-orderer-cluster-ip-service.yaml (A Service config of type “ClusterIP”)
– fabric-orderer-loadbalancer.yaml (A Service config of type “LoadBalancer”)
– fabric-orderer-persistent-volume-claim.yaml (A Persistent Volume Claim config)
- Fabric-peer:
– fabric-peer-deployment.yaml (A Deployment config)
– fabric-peer-configmap.yaml (A ConfigMap config)
– fabric-peer-cluster-ip-service.yaml (A Service config of type “ClusterIP”)
– fabric-peer-loadbalancer.yaml (A Service config of type “LoadBalancer”)
– fabric-peer-persistent-volume-claim.yaml (A Persistent Volume Claim config)
- Crypto-svc:
– crypto-svc-deployment.yaml (A Deployment config)
– crypto-svc-configmap.yaml (A ConfigMap config)
– crypto-svc-cluster-ip-service.yaml (A Service config of type “ClusterIP”)

Note: Kubernetes cluster secrets cannot be included in the list of artifacts as they will be containing the actual credentials of the pilot users.

3.1.2.5 Screenshots of Fides

The below screenshots show the officer portal of two different institutions (blue and orange logo) – the login and then the overview of a customer file with the relevant stamps/verifications.

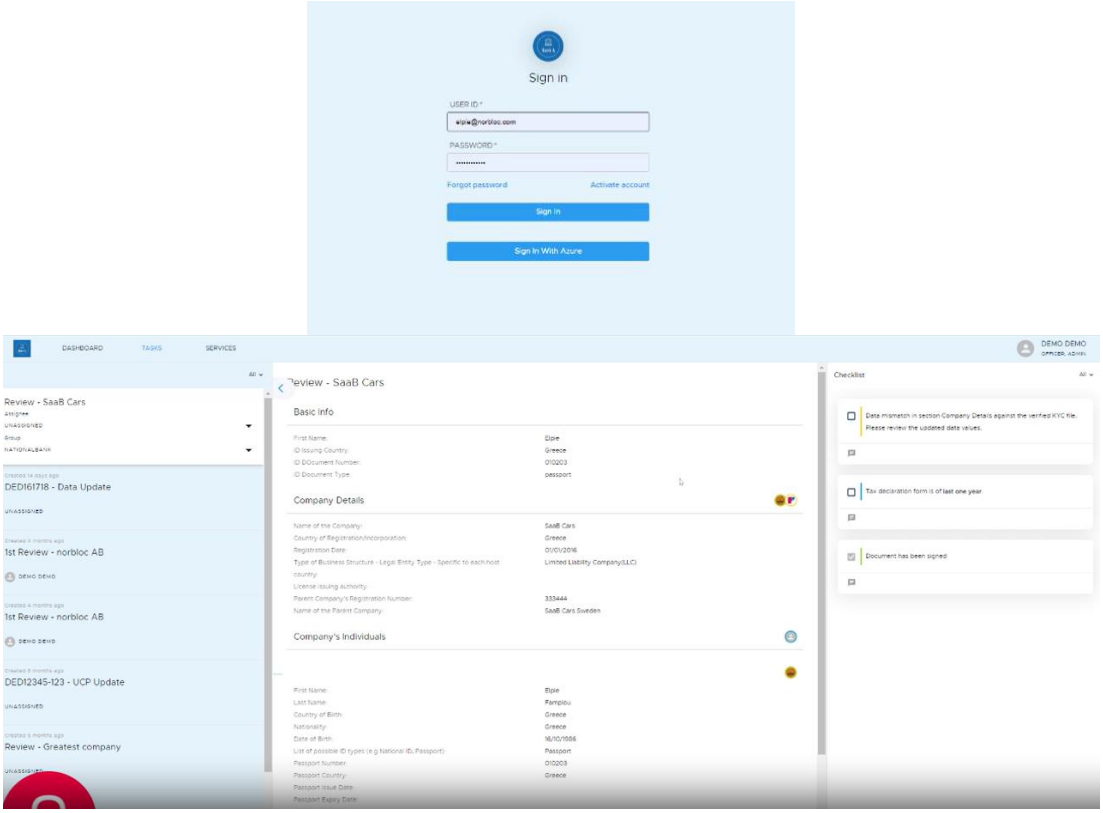

Figure 6 Fides institution A example

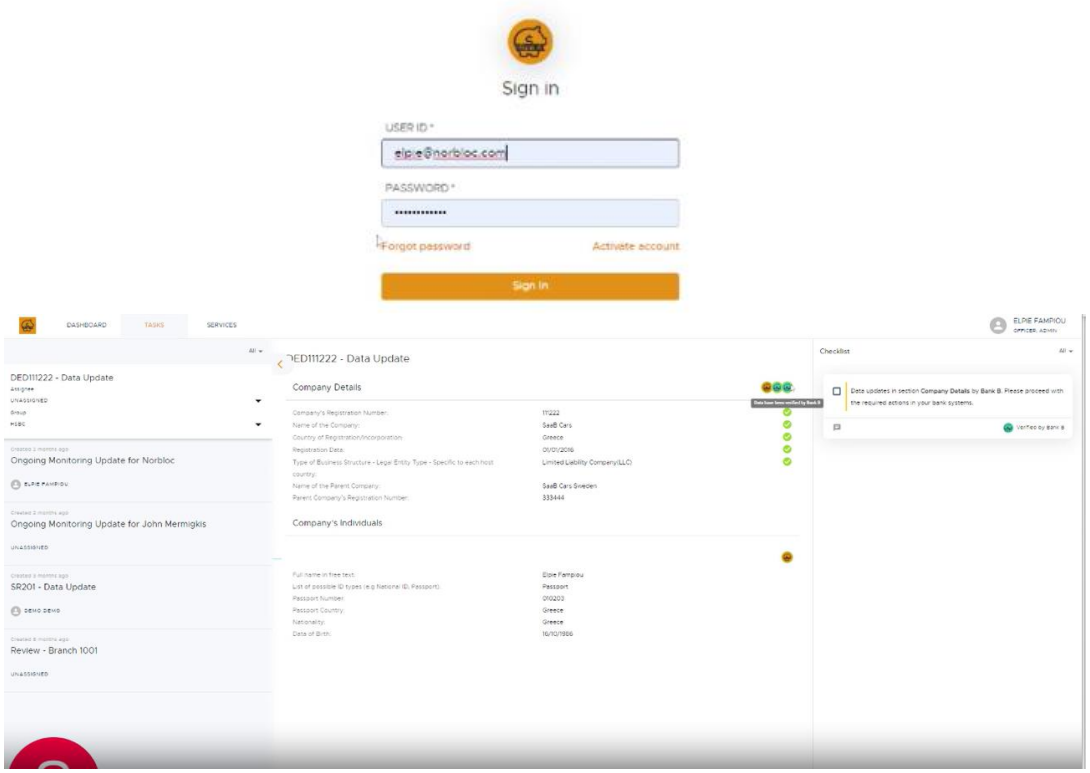

Figure 7 Fides institution B example

3.1.2.6 Future work

Fides component will be used as a decentralized data storage and sharing platform by other components requiring such storage and/or sharing. For other components to use Fides, two sets of APIs will be developed:

1. Data API. A set of APIs to be utilized by other components to store data in Fides and request data from in Fides DRILL;
2. Consent API. A set of APIs to be utilized by other components to manage data access permissions in Fides DnAMS.

The exact format of the data stored is not limited by Fides and can be made compatible with IDS/GAIA-X/DID specifications.

3.2 Trustworthy Data Sharing [T3.2]

3.2.1 Introduction

Trustworthy Data Sharing is a common issue in all use cases and in data sharing ecosystems. It can be considered as a set of tools supporting policies on access and usage control of data. It does not enforce usage policies on its own, but it contributes to a controlled usage of data across entities. Hence, addressing trustworthiness is a critical aspect of data sharing policies within a data sharing ecosystem. In this context it is related to creating an environment where participants can have confidence in a controlled secure, privacy preserving use of shared data. Incorporating trustworthiness-focused policies into the data sharing ecosystem, fosters participants confidence in the trustworthy use of shared data and fosters collaboration and maximizing the value derived from shared information.
Therefore, policy driven data access and usage should be supported by enforcement of policies to data and facilitate control how data may be processed, stored, aggregated, or forwarded. Within T3.2 three major components will be developed. They are briefly described in the following subsections.

3.2.1.1 Usage Control – Privacy Risk Scoring (FHG)

The Usage Control module establishes a technical privacy risk score for the data to be shared on the Tango platform, based on simulated attacks and metrics rooted in information theory. The privacy 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. While privacy usually refers only to individuals, the privacy risk score can also be used to estimate the risk of intellectual property disclosure.
In addition, one aspect of the privacy risk score is the privacy characteristics of the data sharing partners (consumer and provider). As this is also an aspect of the trustworthiness assessment, it will be decided at a later stage whether parts of the privacy risk score will be used within the trustworthiness assessment. For now, the privacy score and the trustworthiness score should be considered as independent components with the possibility that the privacy risk score becomes a component of the trustworthiness score.
Finally, it should be noted that Usage Control is a pipeline for enforcing policies on data to be shared. Thus, a critical aspect of usage control is the application of technical changes to the data, which is outside the scope of this task.
Therefore, we propose to rename this task to Privacy Risk Assessment Module.

3.2.1.2 Trustworthiness Scoring (FTSDE)

The Trustworthiness Scoring Module (TSM) shall evaluate the Trustworthiness of data sharing partners (consumer and provider). It is based on the IIC trustworthiness model utilizing system characteristics, which are evaluated as weighted sum of the degree of attributes' fulfilment, which defines protection objectives or properties. The attributes represent criteria measuring the fulfilment of policies. The characteristics may also be aggregated to a single Trustworthiness score in the same way as a weighted sum.

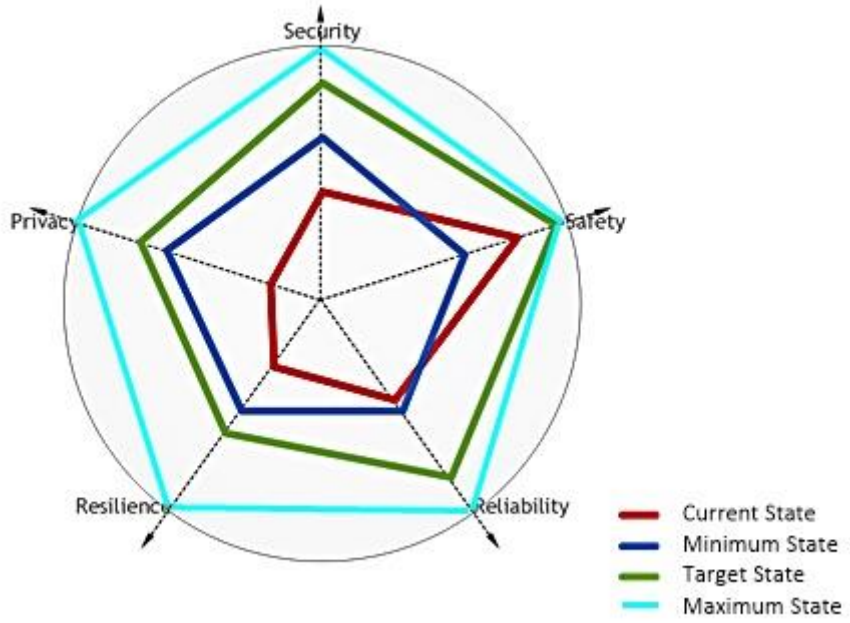

Figure 8 IIC Trustworthiness Radar

This method realizes a transparent and simple scoring, which may be used in a data access control or as sticky policy for data sharing. However, the complexity of this method lies in the specification of technical data representing policies and thus is closely linked to the TANGO use cases.
Alignment with use case on the availability and selection of relevant and technically measurable system characteristics, attributes and protection objectives is ongoing. The presented module therefore is a

showcase demonstrating the configuration and the scoring method based on synthetic data. The visualization of the Trustworthiness scores illustrates the way the scoring module works.
The user may select out of a predefined set of characteristics. In the demo case the characteristics are selected as specified in the underlying model of the Industrial Internet consortium (see Figure 8), select weights and required minimum values as thresholds.
Further development shall integrate this module within the Pilot 2 Autonomous Vehicles and Pilot 3 - Smart Manufacturing Case 3. With this, it is planned to provide the Trustworthiness scores for the policy engine on TANGO contributing to the decision of granting access to data.

3.2.1.3 Ubiquitous Personal Context Vectors (UPCVs) (VTT)

The privacy preserving UPCV technology is designed for personalized recommendations, service personalization and advertising, based on user behaviour. One of the obvious applications is to recommend content or services, based on findings by people with similar behaviour in the past. Therefore, UPCV is categorized as a pure collaborative recommender. In general, recommenders aim at predicting user actions, therefore the technology can be assumed to apply for business intelligence as well, e.g., for defining user groups and item clusters. Unique as a collaborative recommender UPCV, does not record any user history as such – it is based on exchanging Volatile Random Numbers (VRN) that have no association with personal actions – and can be implemented into a decentralized or distributed architecture. All this enables information sharing between different vendors without breaching users' privacy or violating data protection regulation.
As a nutshell of the method, each user and item maintain VRN collections, between which VRNs will be copied, whenever the user has indicated interest in the item. Consequentially, copies of each VRN can potentially be found everywhere, hence the privacy. Items that same users had an interest on, tend to get same VRNs, while users interested in same items tend to have same VRNs, too. Recommendations are based simply by comparing VRN collections – best matches will be recommended.

3.2.2 Usage Control – Privacy Risk Scoring

3.2.2.1 Current development

The privacy risk scoring module is currently in the development stage.
Important steps being worked on are the identification of attributes that influence the privacy score (e.g. attacker capabilities, privacy violation history of the data provider or data consumer, or simulation models).
In addition, previous projects will be reviewed to see if source code and other concepts can be reused. For example, previous master's theses on privacy risk assessment supervised by FHG could be an asset.

3.2.2.2 Short description of the components and the internal architecture

The purpose of the Privacy Scoring Module is to calculate a privacy risk score for data to be exchanged on the Tango platform. The architecture of the module is depicted in Figure 8. To calculate the privacy risk-score the data itself, the data consumer and the data provider must be modelled (data sharing partners). The data consumer is the recipient of the sensitive data, whereas the data provider is the holder of such information. Note, the data consumer can be a third party or a software module.
In addition to that, an adversarial model needs to be defined which estimates the capabilities of a potential attacker. On top of that, an information model must be defined. It describes characteristics about the data to be exchanged.
After this, the privacy risk can be calculated before or after the use of privacy enhancing technologies to safeguard personal data. With that it is possible to measure the effective protection of a privacy enhancing technology.
Ways to assess the privacy risk score can be purely information theory-based approaches that approximate how much sensitive information is contained in the data.

However, the more common way of measuring privacy risks is simulated attacks that can also partially be based on information theory-based principles. Nevertheless, most of them are performed by combining background knowledge of a modelled adversary with the data to be exchanged.

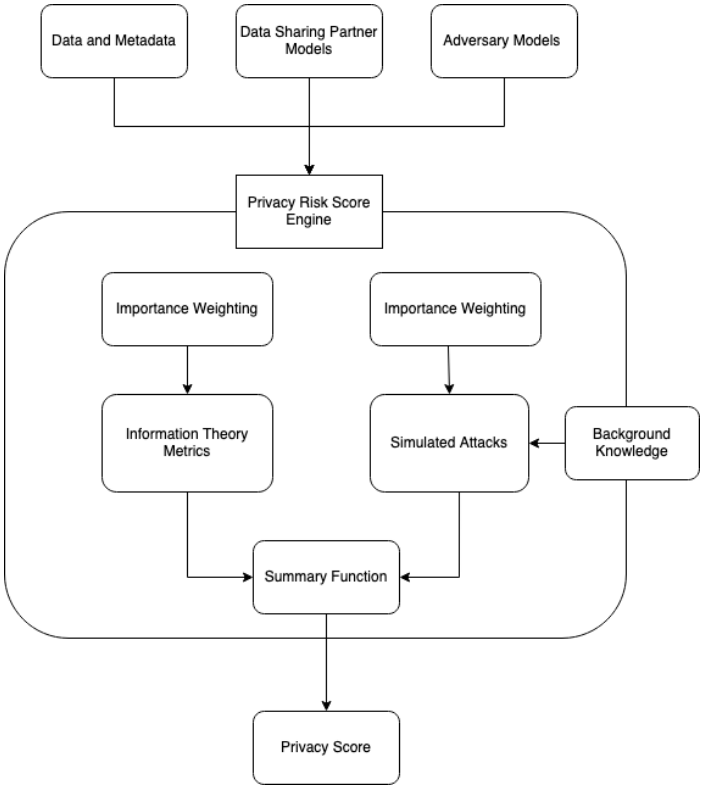

Figure 9 Architecture of the Privacy Risk Module

To illustrate the concept, we provide an example of one such risk estimation from Manjari Chaudhri's Master's thesis titled “Estimation of Individual Privacy Risk in Data Sharing using Predictive Models.” This work was supported by Fraunhofer FIT (FHG) staff involved in the project, including Felix Hermsen and Mehdi Akbari Gurabi. It's important to note that the thesis includes multiple risk estimations; here, we present the simplest one.

The thesis primarily focuses on the privacy risk estimation of medical data in a tabular format. A key consideration is the Article 29 Data Protection Working Party of the EU[1], which states: 'An effective anonymization solution prevents all parties from singling out an individual.' Based on this, one approach was to compute an approximation of a privacy risk score for tabular anonymization. Figure 9 displays the source code used to calculate the number of data points that can be singled out when the adversary has some background knowledge. Following this, the privacy risk score is estimated based on the number of records a potential adversary can single out using different sets of background knowledge giving more weight to counts when less background knowledge is required.

[1] https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Require: Dataset D with n columns
Ensure: A list of counts *Counts* for each row, where each count represents the number of ways a row can be singled out using combinations of n attributes

```
Initialize an empty list of counts, Counts
for each row in D do
    Initialize an empty list, rowCounts
    for r from 1 to n do
        Compute the number of combinations of r attributes that can single out the current row, combinations
        Append combinations to rowCounts
    end for
    Append rowCounts to Counts
end for
return Counts
```

Figure 10 Singling Out Risk Algorithm

3.2.2.3 Features implemented

No features have been implemented yet.
The development plan is to select one pilot use case and implement the structure of the previously defined architecture alongside it. The architecture will be implemented in such a way such that it supports arbitrary metrics and can support all candidate pilots.
Implementation will begin with the selection of appropriate metrics for each data domain considered. For instance, calculating a privacy score for image data from an autonomous car is different than calculating a privacy risk score for transactional banking data in a tabular or graph format.

3.2.2.4 Support for pilots

Currently, we are actively engaged in discussions with IDIADA regarding their autonomous driving pilot, which is described in Chapter 2 of the Tango deliverable D2.2. IDIADA is interested in using the Tango platform to upload sensor data collected during the car's operation to the cloud, while ensuring certain privacy and security guarantees. In relation to this task, our plan involves assigning a privacy risk score of sensitive data that has been modified using privacy-enhancing technologies (for example, face blurring in collected video data). This is to ensure that the data collected and stored in the cloud maintains sufficient privacy.

3.2.2.5 Software artifacts

As a software artifact, we will provide a code repository that can be used to estimate a privacy score for data specific to the associated use cases and is integrable in the Tango platform.
In addition, as mentioned earlier, the privacy scoring module could be integrated into the trustworthiness scoring module. Thus, software produced within the privacy risk scoring module will also be compatible with trustworthiness scoring module.

3.2.2.6 Future work

The next milestone will be a baseline demonstrator usable for privacy risk estimation in a dataspace scenario.

3.2.3 Trustworthiness Scoring

3.2.3.1 Current development

The Trustworthiness Scoring Module (TSM) depends on data on the fulfilment of user specified protection objectives, which are combined in a weighted sum to a score of system characteristics, which themselves are combined to a Trustworthiness score.
The current development of the TSM provides 3 docker containers:

- tsm_frontend, delivering a frontend for configuration, visualization, and the generation of synthetic data,
- tsm_backend, calculating the scores based on the currently synthetic data package and the configuration stored in a PostGres database,
- tsm_database, which serves as persistent storage of the users' configurations of characteristics, attributes, weights and thresholds (i.e. minimum values)

This is illustrated in Figure 11:

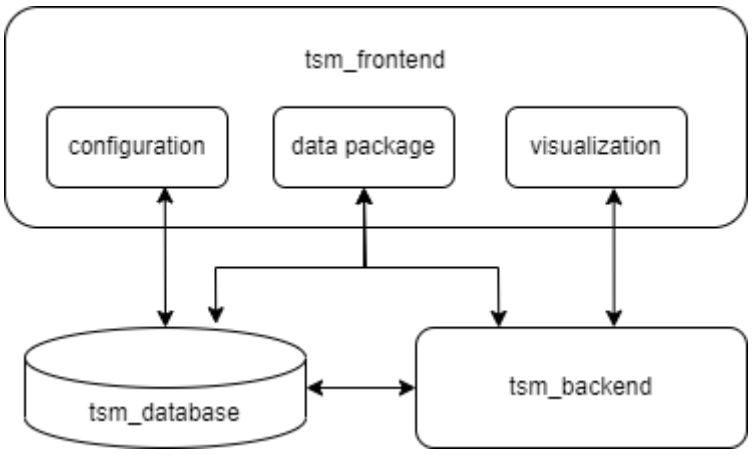

Figure 11 TSM Modules

3.2.3.2 Short description of the components

The TSM frontend facilitates a user specifying configurations of the Trustworthiness scoring by the configuration module using a graphical user interface (GUI). The screenshots show a Fujitsu branded Demo version. This will be adapted to a TANGO branded version in the next steps.
The GUI allows the selection of a set of system trustworthiness characteristics and related attributes. Moreover, weightings and required minimum values as thresholds can be set by the user. In the current version the 5 characteristics specified by the IIC and 5 generic attributes are selectable for each characteristic. However, the user may change the weights and experience the influence to the score.
A further screen of the GUI visualizes the trustworthiness score per characteristic in a radar diagram. The trustworthiness score is visualized by color of the area within the area defined by the characteristics' values. For this moment a simple color schema has been selected: the Trustworthiness Score pentagon is coloured green if the score is greater of equal the threshold and red if it is smaller than the threshold.
Finally, another screen enables the generation of synthetic data as placeholders for use case data. The user may define the duration of the synthetic trustworthiness measurement in second with one data per second. The data is generated by a random number generator and values are in between 0 and 1. The series may be smoothed by a simple moving average selected by the user. This way more or less smoothing can be applied and the effect of trustworthiness scores can be assessed by the user within this demo.
The preliminary version is a standalone demonstrator of the Trustworthiness scoring. It illustrates the work principles and shall instruct the use case owners and may be used for dissemination and training purposes. However, the core modules shall provide data for access policies based on the criticality of data and observations by the sharing party.

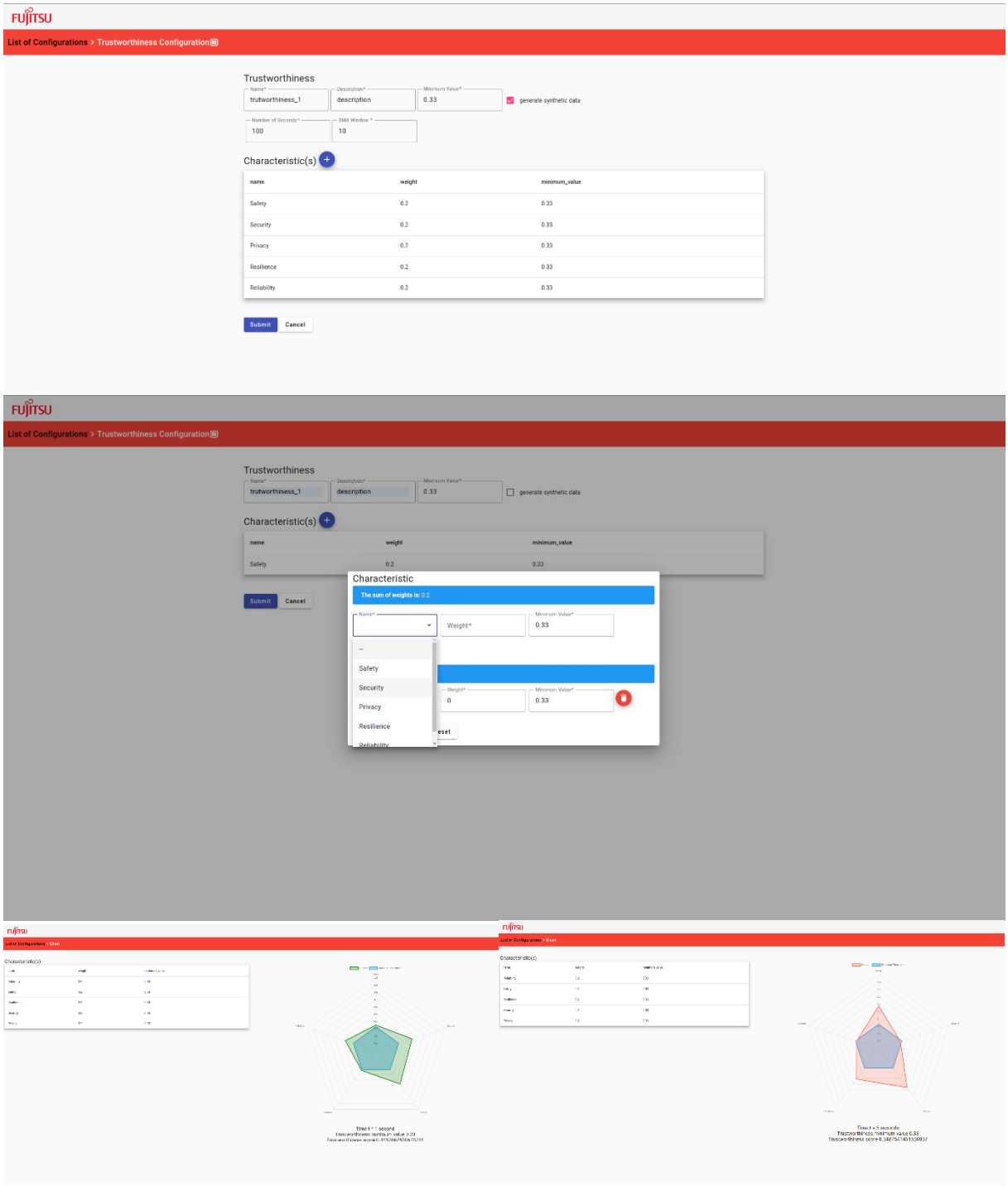

Figure 12 Trustworthiness configuration

3.2.3.3 Internal architecture

The TSM internal architecture has been illustrated in the sections above and in Figure 11.
The TSM comprises 3 components provided as docker containers:

- The tsm_frontend is based on angular. Chart.js is used for the visualization part.
- The tsm_database is a Postgres database. It contains the user's Trustwortiness configurations and a timeseries of scores for visualization or future analytical purposes, which are not yet envisioned.

- The tsm_backend is a python code accessing the configuration from the tsm_database, reading the data package and calculating the scores of the characteristics and the Trustworthiness using the staggered schema of weighted sums.

Up to now the TSM is a standalone demonstration tool. However, it shall gather measurement data from the use cases. Hence, the final component depends on the integration with use case systems. Currently, kafka is the most promising approach for a 1st integration.
Moreover, the TSM shall provide scores to the TANGO policy engine. Hence, it needs to push scores and policies shall apply the thresholds. The design of this part is open yet. Sticky policies using the thresholds and WEB:DIDs seem to be appropriate approaches.

3.2.3.4 Features implemented

With the TSM demonstrator the configuration and scoring components have been implemented. Additionally, the visualization component enables the inspection of the scoring by an operator. This way, adaptation of the weights of the attributes and the characteristics can be fine-tuned by the pilots after integration.

3.2.3.5 Support for pilots

The TSM is targeted to support the pilots

- Pilot 2 Autonomous Vehicles and
- Pilot 3 – Smart Manufacturing Case 3

In first alignment sessions, the principles of the TSM and the configuration were explained and further enablement sessions for detailing the available data and the configuration of the TSM are scheduled. In collaboration with use case 2 the specific data for scoring privacy, security and safety of a vehicle or the other components of the solution are currently assessed.

3.2.3.6 Software artifacts

As outlined above, the TSM is contributed by 3 docker containers. The tsm_frontend is illustrated in the screenshots in section 3.2.3.2.
The swagger documentation of the TSM's standalone demo version internal APIs can be accessed by the user on the local machine where the demo is installed by visiting the link http://localhost:8000/api/docs. The result is illustrated in Figure 13 below:

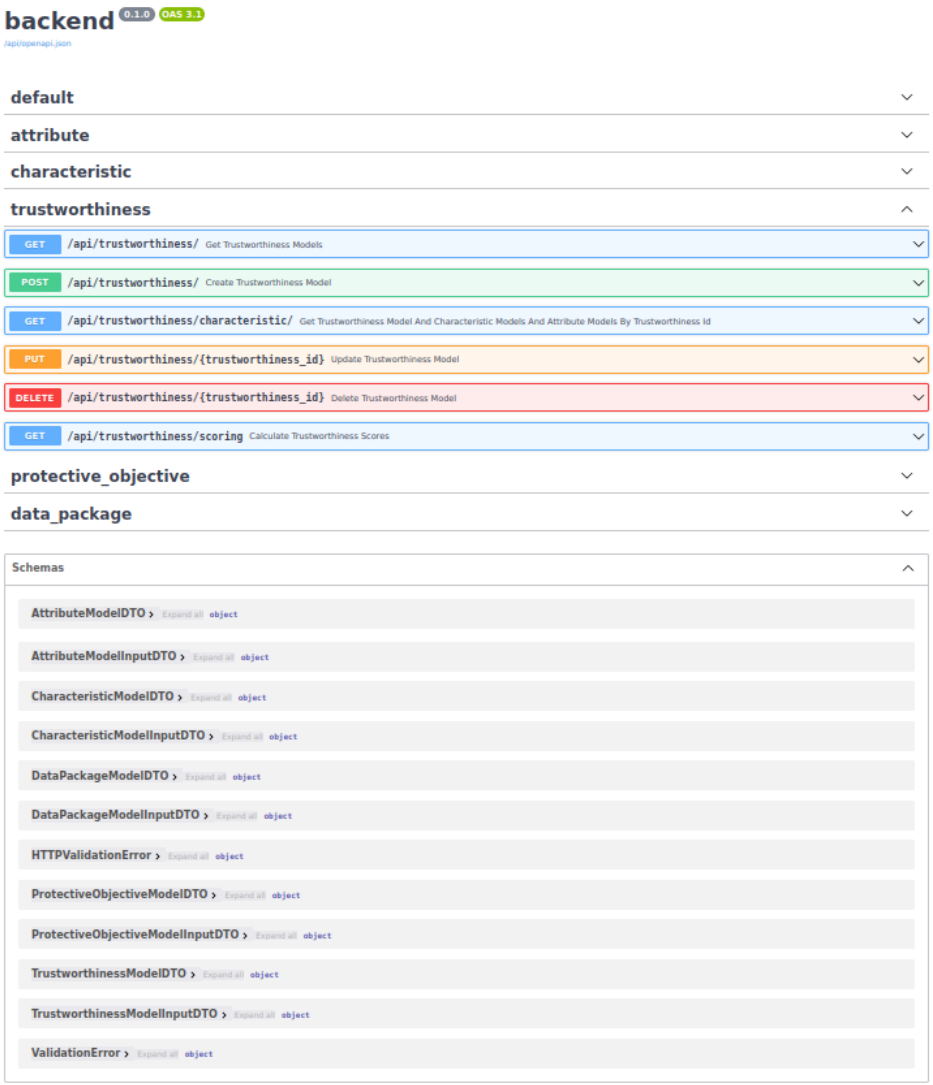

Figure 13 Swagger API documentation of TSM

3.2.3.7 Future work

The demo module currently is designed as a conceptual tool running on a local machine facilitating the assessment of the Trustworthiness Scoring by the use cases and for dissemination. It shall enable use cases' understanding of the way it works and supporting discussion of relevant data and the integration of data sources. The alignment with pilots on data, data sources and integration is ongoing.
In addition, the existing module lays the foundation for integration into the TANGO environment and the policy engine. Here we envision policy description in XACML and WEB:DIDs or other appropriate means for scores' evaluation.

3.2.4 Ubiquitous Personal Context Vectors (UPCVs)

3.2.4.1 Current development

UPCV algorithm is VTT's background to the project, with several patent families, having a centralized C++ Windows server implementation and a WIN32 desktop test application as VTT background in the project. These source codes are currently under porting to LINUX, from where different functionalists can be extracted and re-packaged to TANGO usage.

3.2.4.2 Short description of the components

In TANGO T3.2, UPCV technology will be introduced in a distributed architecture, in which each user and each vendor have their own instances of a User Behavior Exchange Module (UBEM), which is a symmetric component implementing the algorithm - similar on both user and vendor (service) side.
Regarding 'Smart Hospitality' use case, each user has a personal UBEM in a mobile device, integrated with a simple mobile app (UI) capable of displaying offerings ("items"; e.g. services, tourist activities) which the vendors in the location (e.g. the hotel or activity providers, respectively) have available. Each vendor also has their UBEMs that can communicate with the user UBEMs. UBEMs on vendor side operate with open IP addresses and are connected to a simple service that provides contents associated with each item.
UBEM has an additional functionality of creating recommendations. They will be generated on users' UBEMs and the respective contents will be displayed on the UI.

3.2.4.3 Internal architecture

While the UPCV technology can be implemented in various topologies, a distributed implementation will be introduced in TANGO. All communications between a user and a vendor take place on one-to-one basis. In Figure 14, the user on the right indicates an interest to an Item (* on the left), after which vendor's UBEM and user's UBEM exchange VRNs between the respective VRN collections, as described earlier.

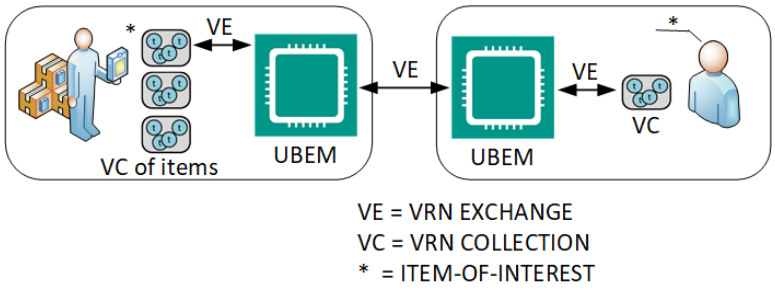

Figure 14 User interest and exchange of Volatile Random Numbers by the UBEM

In Figure 14, the user is provided by item VRN collections, among which the user selects the collection(s) that has the best match and receives item identifier(s) of the respective recommendation(s).

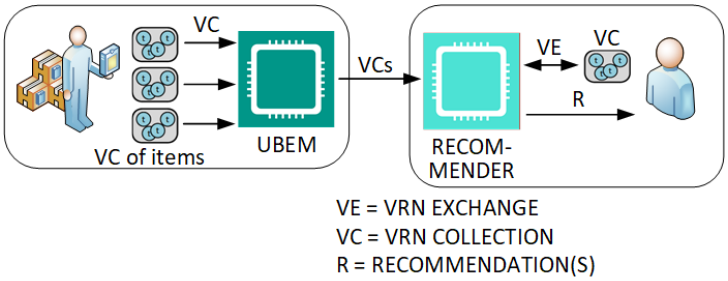

Figure 15 User Recommendations

Assuming that there would be several vendors providing same or similar items, they can co-operate by synchronizing the VRN collections of said items, enabling an ecosystem which have no centralized hub. The Figure 15 illustrates this.

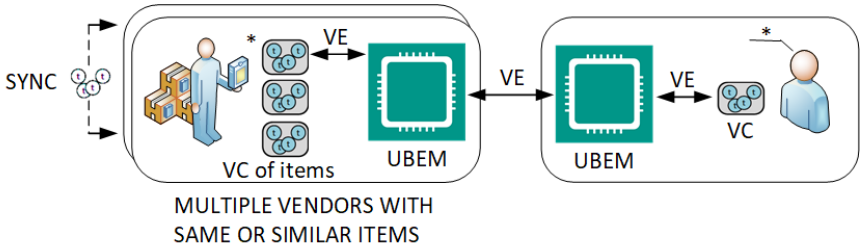

Figure 16 Synchronizing VRNs across vendors

As per the current design choice, recommendation engine will be provided as a compiled WebAssembly package with JavaScript wrapper for each client. The main functionalities are based on two different use cases; requesting recommendations and showing interest in an item. The use-cases can be depicted in a sequence diagram as follows.

In the first use-case, user requests recommendations. This can be, for example, through clicking a button 'Activities for me'. The client gets a list of items and their respective VCs from the server through API request. With the item information, recommendation engine and the users private VC, the client can determine recommendations for the user.

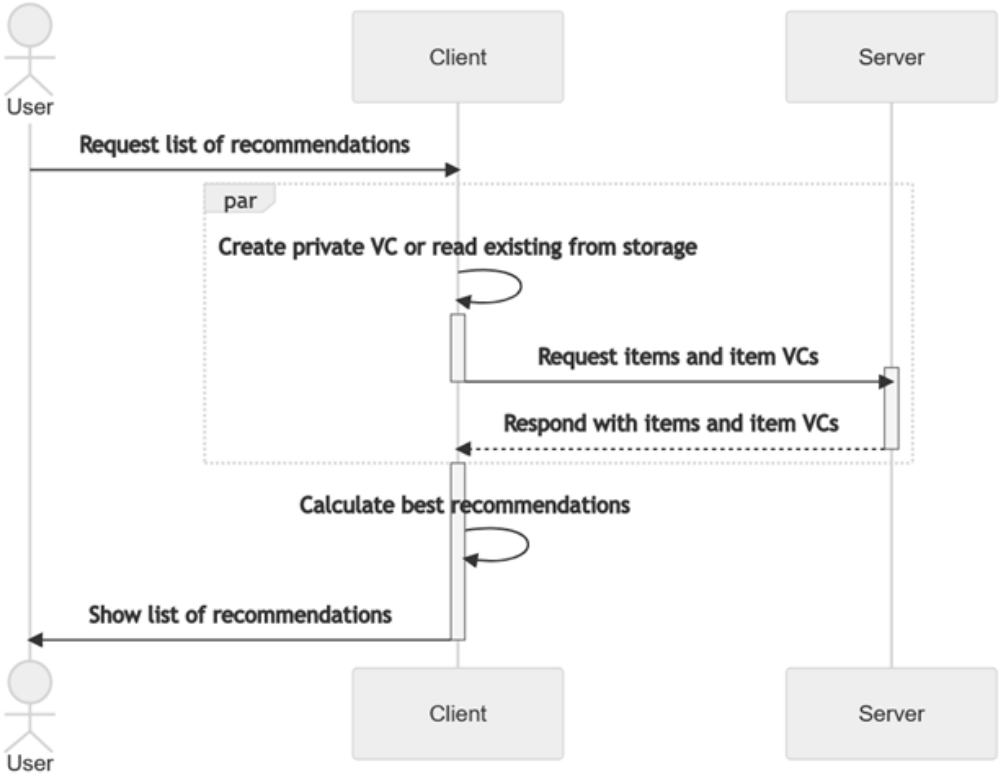

Figure 17 Recommendation process

In the second use-case, the user indicates interest in a specific item. This triggers VRN exchange flow.

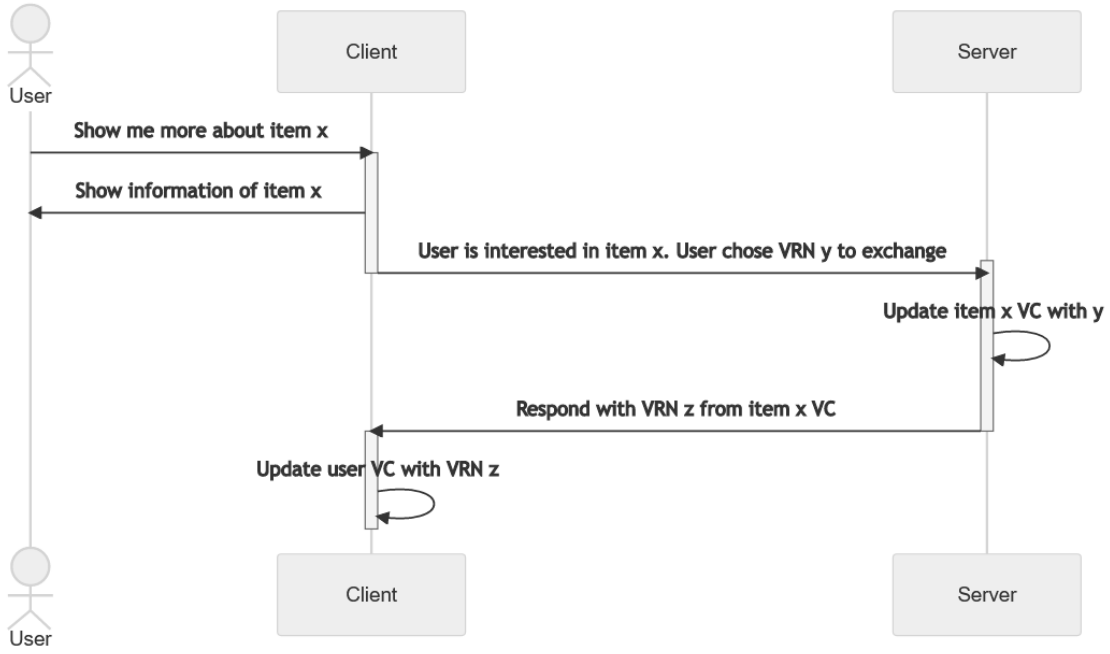

Figure 18 VRN exchange process after showing interest in an item.

In 'Retail' pilot, the existing UPCV test application will be used for analysis, while subsequent decisions will be based on the results.

3.2.4.4 Features implemented

TANGO will exploit UPCV technology that provides user-item recommendations for hotel guests in 'Smart Hospitality' pilot. For 'Retail' pilot, the desktop test application will be used for shopping basket analysis, primarily as an item-item recommender.

3.2.4.5 Support for pilots

Pilots represent use cases describes in a previous deliverable D2.1. UPCV will be applied in 'Smart Hospitality' pilot as a personalized recommender for offering by one or more vendors and data sharing between them, as illustrated in Figure 16. In 'Retail' pilot shopping baskets will be analysed, with eventual goal to enable analytics between corporate branches in Greek and Cyprus.
At the time of writing this, the status of the pilots is the following: For 'Smart Hospitality' pilot Alcudia Garden Aparthotel in Mallorca has been selected as the location. For 'Retail' pilot, the first shopping basket dataset of 131862 user-item transactions has been received from partner METRO a few weeks ago, waiting for analysis.

3.2.4.6 Software artifacts

At the time of writing this, no software artifacts have been released yet.

3.2.4.7 Future work

For 'Smart Hospitality' pilot, two main components containing the UBEM have been planned: a Docker image of a vendor side server and a installable mobile applications and/or a web application for the users. For 'Retail' pilot, decision for further actions will be based on the results of analysing the first data set.

3.3 Confidentiality and Privacy by Design [T3.3]

3.3.1 Introduction

Confidentiality and Privacy by Design entail embedding confidentiality and privacy considerations into the design and development of the system, adopting a proactive approach rather than treating them as a compliance requirement at later stages of the development. It involves anticipating events that affect data privacy before they take place, so that privacy becomes the default setting. Additionally, privacy must be guaranteed throughout the lifecycle of the data- adequate measures must be implemented in every stage of data processing, following visibility and transparency principles. Importantly, Privacy by Design has now transitioned from a best practice to a legal requirement under many privacy regulations, including GDPR. These above principles follow the specifications defined for Privacy by Design in ISO 31700-1:20232 to consider in the design process of data lifecycle.
To this end, TANGO aims to adopt the two main mechanisms followed:

- **Sticky policies**:
  Machine-readable policies are "sticky", attached to the data they govern, they travel with the data, ensuring that the specified restrictions or permissions are consistently enforced throughout its lifecycle, even as it moves across different systems, platforms, or organizations.
  The way in which sticky policies are instantiated is through CP-ABE. Data is encrypted according to defined attribute-based policies (ABE) set by the data owner, this allows giving the user the power to control how and with whom their data is being shared, being enforced during the whole data lifecycle.
  Additionally, data can be optionally signed providing authentication and integrity to the protected data.
- **User Consent (UC)**:
  Policy-based access control will be used to address user consent. Attribute-based policies will be used to enable fine-granular checks based on identity attributes. It plays a crucial role in ensuring compliance with privacy regulations (e.g., GDPR), fostering trust between the user and the data provider.
  In conjunction with CP-ABE, we will leverage the standardized XACML[3] framework, which consists of different components that work together to enable fine-grained access control and establishes a process of policy-based decision-making. XACML's attribute-based policies can integrate identity attributes which are associated to user preferences, permissions, or consent choices. Besides, policies can be tailored to incorporate specific conditions related to user consent, allowing for granular control over data access.

3.3.2 Current development

Currently, some subcomponents that define atomic features (e.g., ABE toolset, PDP/PEP functionalities) have already been developed and they are being tested. However, they have yet to be integrated into the architecture and into the different processes, which are defined in the following section.

3.3.2.1 Short description of the component

The process of granting a consumer access to the requested data can be described through the identification of two parallel sub-processes, following the different mechanisms established above to fulfil confidentiality and privacy by design requirements.
On the one hand, leveraging the use of Verifiable Credentials (Verifiable Presentations together with zero-knowledge proofs, sourced from the SSI module, task 4.1.) and the XACML framework (comprising different modules, mainly: Policy Enforcement Point, PEP; Policy Decision Point, PDP and

[2] https://www.iso.org/standard/84977.html
[3] https://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html

Policy Information Point, PIP) the user/consumer is granted access to a certain asset. This process involves the following steps:

1. User authenticates at the connector (with a previous step of registration).
2. User (consumer) sends request to producer (provider connector) for accessing an asset, initiating the authentication process with the SSI wallet.
3. Verifier requests the consumer for VCs that claim that the user owns credentials connected to the attribute-based policies defined to access the requested asset.
4. (Optional) Wallet checks if the verifier belongs to a participant in the connector dataspace and the consumer sends the VPs.
5. Verifier verifies whether:
    a. Consumer is a trusted participant of the dataspace.
    b. VCs were issued by a trusted issuer in the dataspace.
6. If everything is OK, a token is sent to the user.
7. Consumer uses token to access X asset.
8. PEP proxy and PDP verify if the claims included in VPs extracted from the token is authorized to access the request.
    a. PIP will request additional external attributes so that PDP can verify every defined policy (e.g., trust scores defined in T3.2).
9. If authorization is OK, the request is forwarded, and (optional) consumer receives an authorization token coming from provider PEP/PDP.

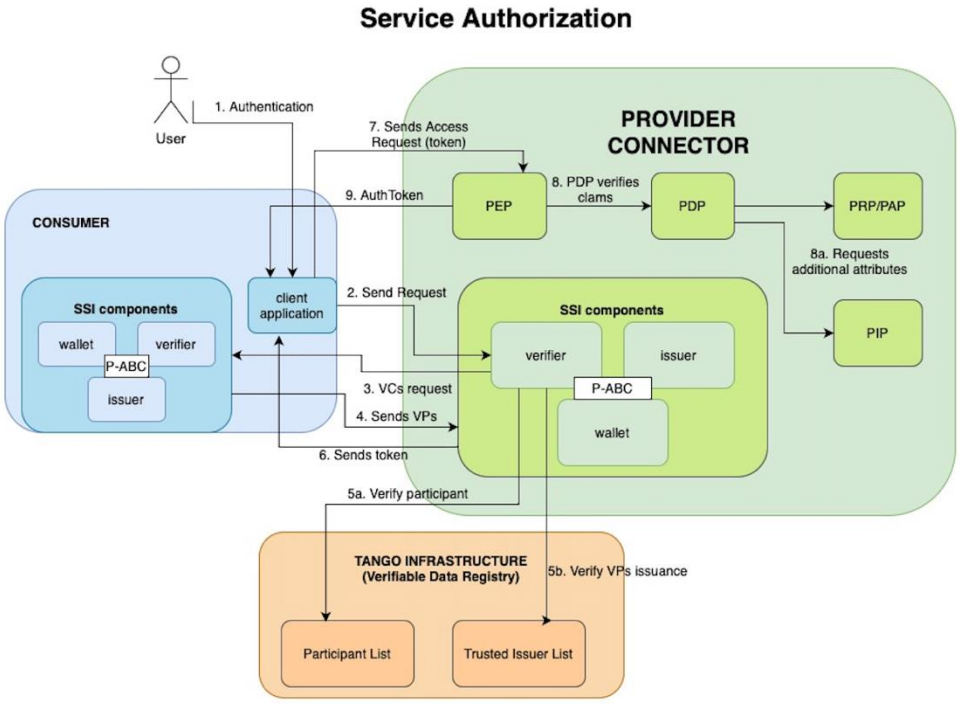

Figure 19. Service Authorization process within the TANGO (FIWARE connector) architecture.

To be noted, the above defined flow is maintained to follow what is defined in the FIWARE dataspace connector authorization process[4]. However, it is susceptible to changes in the next releases once the new SSI components defined in task T4.1. as well as other functionalities such as PEP/PDP are integrated. On the other hand, producer encrypts the data according to attribute-based policies (CP-ABE), this way, once the consumer has been given access, data will need to be decrypted. This decryption process relies

[4] https://github.com/FIWARE/data-space-connector#service-interaction

on a key associated to the consumer's identity attributes, ensuring that data can only be decrypted if the attributes align with the sticky policy. CP-ABE client module operates on both ends, with an encryption module located in the data provider connector and its corresponding decryption module situated on the client side. Process workflow is further detailed in D2.3, 3.3.1.1. (Figure 13, 14).

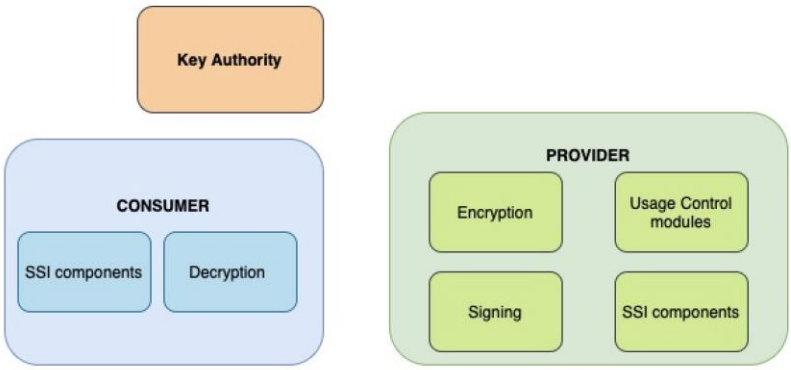

Figure 20 ABE module within FIWARE connector reference architecture

The main steps in the process include:

1. Producer encrypts the data together with the sticky policies (CP-ABE) using the master Public Key provided by the Key Authority (it might include the signature).
2. Consumer requests the asset following authorization process (i.e., see above).
3. Producer sends the encrypted data to the consumer.
4. Consumer decrypts the data using the decryption key obtained from the Key Authority.

These two processes together embody Confidentiality and Privacy by design principles. In particular, the CP-ABE module enhances confidentiality by enabling data owners to define and enforce policies throughout the data lifecycle on the data sharing platform. Simultaneously, the UC module facilitates compliance with GDPR by implementing fine-grained access control mechanisms that enable the enforcement of user consent directives.

3.3.2.2 Internal Architecture

The components that achieve the principles of Confidentiality and Privacy by design are:

- **ABE Toolset**

The ABE Toolset Prototype is designed to facilitate privacy-friendly and secure data exchange between data service providers and consumer organizations. It operates based on a cutting-edge encryption scheme known as Ciphertext-Policy Attribute-Based Encryption (CP-ABE). CP-ABE serves as a secure cryptographic technique for controlling access to encrypted information, ensuring data confidentiality. This toolset plays a vital role in the TANGO pilot scenarios, where precise and granular access to data is necessary. Only users or devices with specific attributes (e.g., managers, administrators) are granted permission to decrypt certain encrypted data.
The ABE Toolset consists of three submodules as described below, through the use of a possible pilot use case.

1. Key Generator (Issuer Authority):

   The Key Generator submodule generates a pair of cryptographic keys (master public and private keys) where the master public key is used for encrypting data and the private key is used to generate CP-ABE decryption keys.
   It provides the following cryptographic key services:

   - Instantiates Master public and private keys, where the public key is returned to the Encryptor upon request to enable the encryption protocol.

- Generates CP-ABE Decryption Keys for users based on their attributes and the master private key.

2. Encryptor:

The Encryptor submodule is employed by the data service provider[5] (e.g. a Map provider to Autonomous Taxi Services). The encryptor provides:

- A CP-ABE data encryption service for the provider to submit (a) data to be encrypted, (b) access policy[6], using the public key previously instantiated for the encryptor, resulting in ciphertext. For instance, for a Map service the access policy could be based on the consumer organization´s name, the role of the end user. Another role could also be for example the End User Taxi customer who is able to access the map data on the taxi service mobile app also.

3. Decryptor:

The Decryptor submodule is used by the consumer end user seeking to decrypt encrypted data based on the end user´s attributes as defined in the encrypted data´s access policy. To facilitate this, the Decryptor provides:

- A CP-ABE data decryption service, where it first obtains the CP-ABE Decryption Key from the Issuer Authority based on the consumer end user attributes. For example, the organization´s name, the role of the end user. If the access policy included in ciphertext is satisfied by the decryption key and end user attributes included within it, then the decrypted text is returned to the end user consumer.

- **Usage Control**

Usage control is approached by means of the XACML (eXtensible Access Control Markup Language) framework, which defines and enforces access policies. Beyond access control, usage control is reached by allowing a more dynamic and context-aware control over the usage of the assets. Policies are defined as a set of rules that consist of a condition and an associated decision (permit, deny, etc).

Additionally, the key components of the architecture that will be included are:

- Policy Decision Point (PDP):

  Module that evaluates and makes decisions based on access control policies. It receives authorization requests from the PEP together with relevant attributes (e.g., VPs) and evaluates the request against the defined policies to permit or deny access to the resource.

- Policy Enforcement Point (PEP):

  Component responsible of intercepting access requests before they reach the resource, the PEP sends the relevant information needed to decide on the request to the PDP (e.g., VPs) for evaluation. Once the PDP makes a decision, the PEP enforces it by allowing or denying access to the resource.

- Policy Information Point (PIP):

  Component that provides additional contextual information needed for the policy evolution process (PDP). It helps the PDP make a decision by providing context-specific data that is not part of the request (e.g., trust scores coming from T3.2.).

- **Signature**

Besides the attribute-based encryption, data can be optionally signed to verify the authenticity and integrity of the shared data. This way, it is ensured that the data provider is the one who claims it and that the data has not been tampered at any point.

[5] The Issuer Authority and data service provider could be the one and the same organization.

[6] The access policy is the responsibility of the data service provider.

The data provider generates a digital signature using their private key (based on SSI attribute), resulting in a unique and verifiable signature which is associated to the data to be shared. The consumer will receive the data together with the provider signature and will use the provider's public key to decrypt and verify the signature. If valid, it is confirmed that the data has not been altered during transmission.

Besides, the main dependency with other components of the architecture relates to privacy preserving identities (associated to T4.1., SSI solution).

- Use of DIDs as identifiers across different domains (compatible with W3C standards).
- Verifiable Credentials for privacy-preserving authentication of attributes, making use of Zero-knowledge proofs for minimal disclosure.

3.3.2.3 Features implemented

The main implemented feature is the ABE encryption/decryption library (at a cryptographic level).
Other features are under development:

- Integration of the encrypt/decrypt flow into connector through ABE component that includes the library plus the signature module based on SSI.
- Integration with SSI for authentication, authorization, and key generation
- Expanding PDP/PEP modules to make attribute-based decisions coming from SSI VPs.

3.3.2.4 Support for pilots

Every feature incorporated into this component is designed to be independent of specific pilots, signifying that no specific considerations are made for a particular pilot. Instead, features are integrated into the generic architecture of the connector.
The pilots the component aims to support, as defined in D2.3., are:

- Pilot 1: Smart Hospitality
- Pilot 2: Autonomous Vehicles
- Pilot 3 Case 1: Smart Manufacturing (FMAKE)
- Pilot 3 Case 2: Smart Manufacturing (RIAS)
- Pilot 5: Public Administration
- Pilot 6: Retail

3.3.2.5 Software artifacts

As of the current writing, there hasn't been any release of a software artifact.
The ABE-toolset, provided by ATOS, will be made available as a docker image and uploaded to the projects Harbor registry. This will enable integration with the signing module provided by UMU and then later the SSI module authentication provided by the Provider Dataspace Connector.
This will be under the umbrella of integration within the TANGO connector.

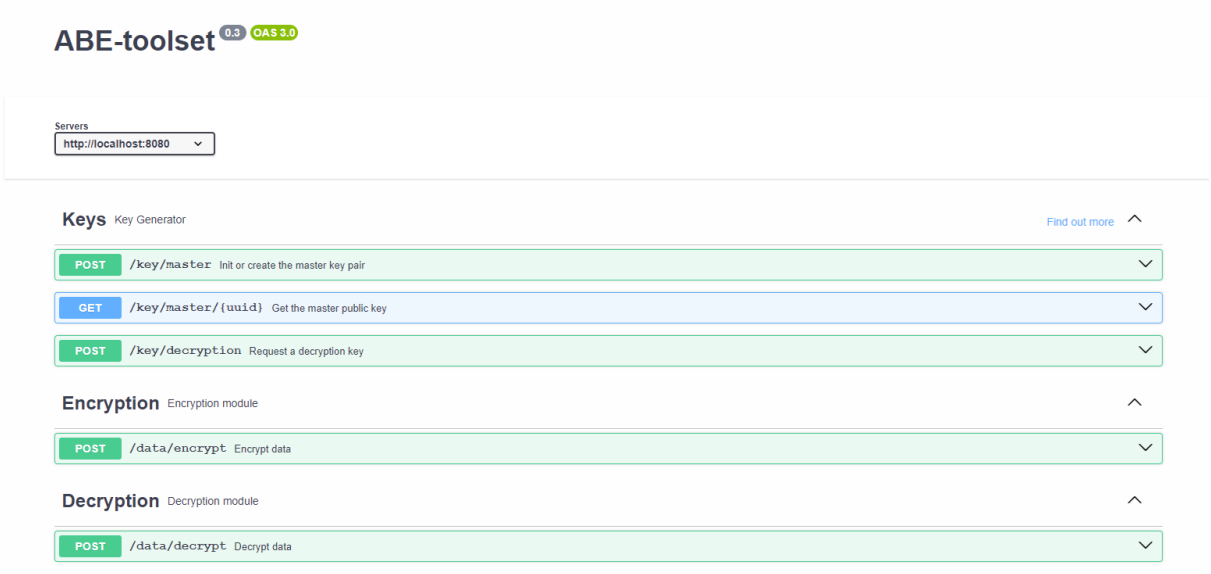

Figure 21 ABE toolset processes

3.3.2.6 Future work

Among the tasks yet to be completed is the integration of the ABE module, including the Key Authority module, and process flow development (e.g., keys request to the KA, derivation of VPs, data signature...) added to the ABE toolset. FIWARE PDP/PEP adaptation so that it makes use of Verifiable Presentations (issued with ZKP) coming from SSI component (T4.1.).

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

3.4.1 Introduction

This section intends to implement a secured information encryption/recovery mechanism to support the confidential information storage and distribution demand for TANGO use cases. Specifically, two components are designed to support the functionality requirement, i.e., self-encryption and decryption module, and multi-factor information recovery mechanism.

- Encryption and decryption: Encryption represent the process through which the contents of a document are made illegible to anyone without the necessary permissions to view them. It usually involves the use of a key, with which the document can be encrypted. The encrypted document is called ciphertext. Decryption is the opposite process, allowing someone to decipher the hidden document. Using again a key, someone is able to reverse the encryption process and view the original document. Self-encryption is a special kind of encryption, which instead of receiving a key from the user, uses the contents of the document to be encrypted as part of the process. This applies for both encryption and decryption.
- Multi-factor information recovery mechanism comes as an additional security measure. Instead of only one person having the key to decrypt the document, the key is split into several shares and divided among multiple people. If someone would want to initiate the decryption process, they would need a certain minimum number of key shares to do it, which increases security. It is less likely that a larger group of people would have malicious intentions.

3.4.2 Self-encryption and decryption

In this section, the structure of the proposed self-encryption and decryption techniques with multi-factor information recovery mechanisms as well as the functionality of each sub-module will be given.

3.4.2.1 Short description of the component

The purpose of designing a self-encryption and decryption (SED) method is to encrypt the plain text with the unique feature of the file rather than user-specified password. Since a minor change cause to a significant (and no-linear) feature discrepancy, the hashed feature (i.e., the key) to encrypt each file is guaranteed to be unique. Moreover, to convenient the information distribution and achieve a higher safety level, the file is split into multiple chunks before encryption, and each chunk is encrypted using the hashed feature of a neighbor chunk rather than its own feature.

Multi-factor information recovery (MFIR) mechanism is used to distribute the key generated by the information encryption/recovery process among a group of participants in such a way that the secret can only be reconstructed when a minimum threshold of these shares is combined. Alternatively, the participants can be replaced by entities of an object, then each piece of key is mapped to an entity, and a minimum threshold of entities lead to a successful secret reconstruction.

The architecture of the proposed self-encryption and decryption with multi-factor information recovery mechanism is demonstrated by three modules, i.e., the encryption module, the decryption module and multi-factor information sharing module. The encryption and decryption module are integrated to the TANGO framework via data connectors. In the data connector, the plain text, required number of participants for multi-factor information sharing and the minimum threshold of participants to recover the share will be sent to the encryption module. The multi-factor information sharing/recovery mechanism serves as an additional plugin. When information sharing process is specified by user, the information sharing process will be called to split the key into multiple pieces (the number of which equals the number of participants who are supposed to hold the key or the number of features for decryption). Similarly, in the decryption process, a key or multiple key piece will be given to retrieve the original key of the encrypted file. In this process, the multi-factor information recovery mechanism will be called when multiple key pieces are given, so that the complete key for decryption can be recovered. Then, the recovered key will be used to execute the decryption process in a data-block manner. Finally, the complete plain text file will be recovered by combining all the decrypted data chunks.

The added value of self-encryption and decryption with multi-factor information recovery can be seen in the following aspects. First off, two modules will be constructed in T3.4, i.e., the self-encryption and decryption module (which is implemented as the encryption and decryption sub-module separately), and the multi-factor information recovery module. These two modules will be served standalone in the form of API. In the future, all these modules can be re-packaged as plugin for new application scenario. Meanwhile, the whole work package is self-included and can be deployed to any online platform without relying on other modules in TANGO. This fact enables our work package to be deployed on any platform, such as software as a service (SaaS), as an open-to-use package for commercial use.

3.4.2.2 Internal architecture

As mentioned above, the whole system includes two workflows, i.e., the encryption workflow and decryption workflow. Besides that, the information sharing module serves as a plugin to split the key to the plain text and distribute key pieces to key holders.

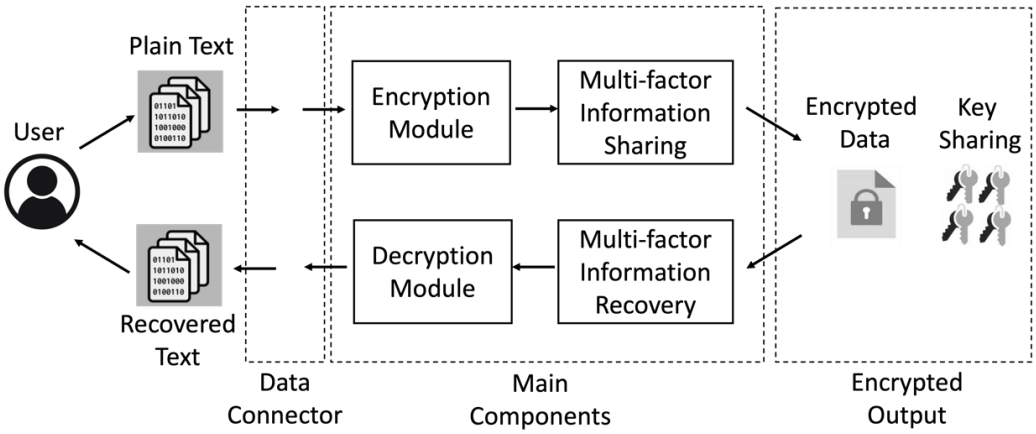

Figure 22 The internal architecture and dataflow of self-encryption and decryption with multi-factor information recovery mechanism

Figure 22 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.

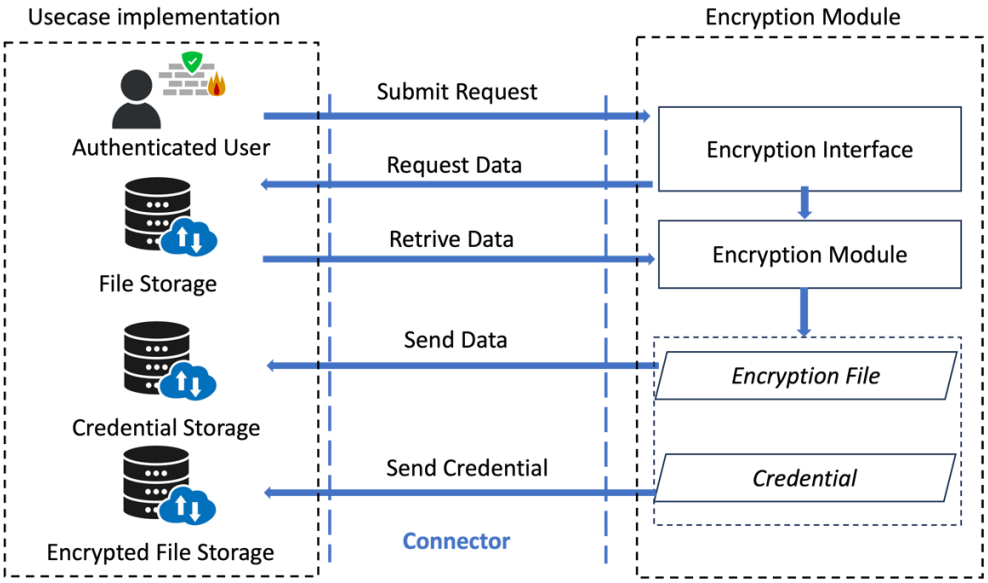

Figure 23 The workflow of encryption module

The workflow of the encryption process is shown in Figure 23. Upon starting the program, the user is asked through the terminal to specify the location of the file they wish to encrypt, the number of chunks they desire for the encryption and the user id under which they wish to encrypt the file. The id has to be of at most 16 bytes, otherwise the user will be constantly prompted to enter an id of the desired length. Additionally, a file that keeps track of users is created (on the first use of the program) or updated. The program will write in the file the user ids of the people who try to encrypt or decrypt the file.

The plain text, that will be encrypted, is split into the number of chunks that has been specified by the user in the initial stage. An additional chunk is created, which contains the information in the user file. The encryption process works in the following manner. All of the chunks' hash codes are calculated using the SHA256 algorithm. Let's consider the chunks C_1, C_2, ... , C_n, with the respective hashes H_1, H_2, ..., H_n. The hash of one chunk is used as the encryption password of the following chunk (H_1 is used to encrypt C_2 and so on). The first chunk will be encrypted using the hash code of the nth chunk. The encryption works in the following way. The text to be encrypted will be referred to as plaintext. The password that has been specified (in this case, the hash of the following chunk) is used to generate a key of 16 bytes. Using the above-mentioned key, the plaintext will be encrypted using AES Cipher Block Chaining mode.
Having encrypted all chunks, each one will be stored in a separate binary file (in encrypted form), marked as chunk ***i*** which representing the number of the chunk. The order will be important for the decryption process. The user will only be given the value of the hash code of the last chunk, to be able to decrypt the file again. The id of the user who encrypts the file is passed as a parameter both when encrypting and decrypting the file.
A separate txt file will be generated, which contains a summary of the encrypted output. This file will present the hashing and encryption algorithms that were used, the number of chunks used in the encryption and the hashes of all the generated binary files.
The workflow of decryption process is demonstrated by Figure 24. When decryption process is called, users are asked to enter the id of the user who encrypted the file, as well as specify their own id. This id (of the user decrypting the file) will be added to the user file.

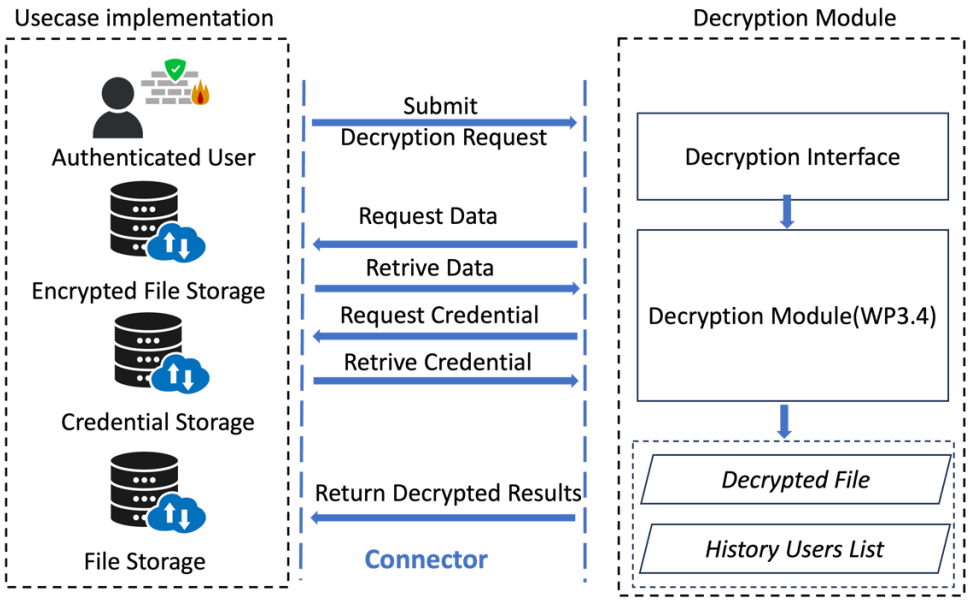

Figure 24 The workflow of decryption module

To decrypt the file, the hash code of the last encrypted chunk is used. A key is generated from this

password, which is used to decrypt the first chunk. Now knowing the value of the first chunk, the program is able to hash this value again, to get the password for decrypting the second chunk. This process is repeated, until all the chunks have been decrypted. The program additionally compares the id of the user who encrypted the file with the id provided by the person trying to decrypt it. If the ids do not match, the file will not be decrypted.

The decrypted chunks will be written in order to another file, thus obtaining the same contents as the original file. To ensure this, the program checks if the hash code of the original document is the same as that of the output file. If the answer is yes, it will tell the user the decryption has been successful, and the user is able to read the output file. If not, something has gone wrong and the program will inform the user of this as well.
The workflow of the multi-factor information sharing mechanism is shown in Figure 25. In the information sharing process, the key of the encrypted chunk (generated by the encryption module) is split into multiple pieces. The purpose of key splitting is to distribute key pieces to multiple participants or map key pieces to multiple features of an object. It worth noting that, when information sharing process is executed, security level should also be specified, which indicates the minimum number of key pieces/features demands to retrieve the original key for encrypted chunk.

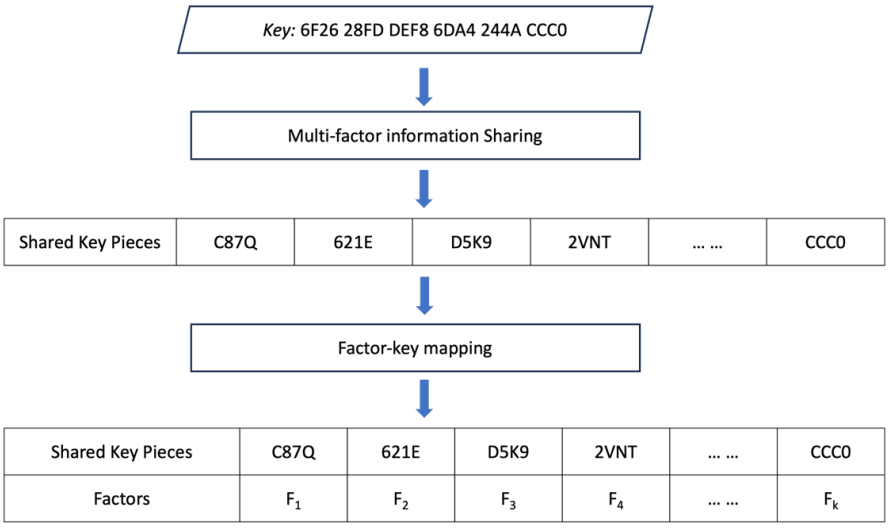

Figure 25 The workflow of the multi-factor information sharing mechanism.

In Figure 26, the mapping between factors and participants/features is briefly presented. The main idea of factor mapping is to build a mapping relationship between key pieces (split by multi-factor sharing mechanism) and participants/features. The number of participants or features is a user-specified parameter, which is given via the encryption/decryption interface in TANGO data connector. By split the key of encrypted data into multiple pieces, the original key can be shared (and in the future, recover) among multiple parties.

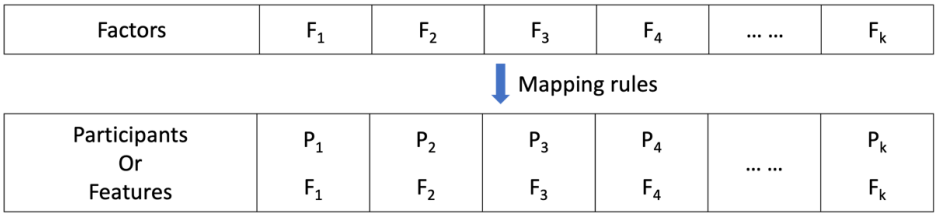

Figure 26 Map factors to participants or features for factor sharing.

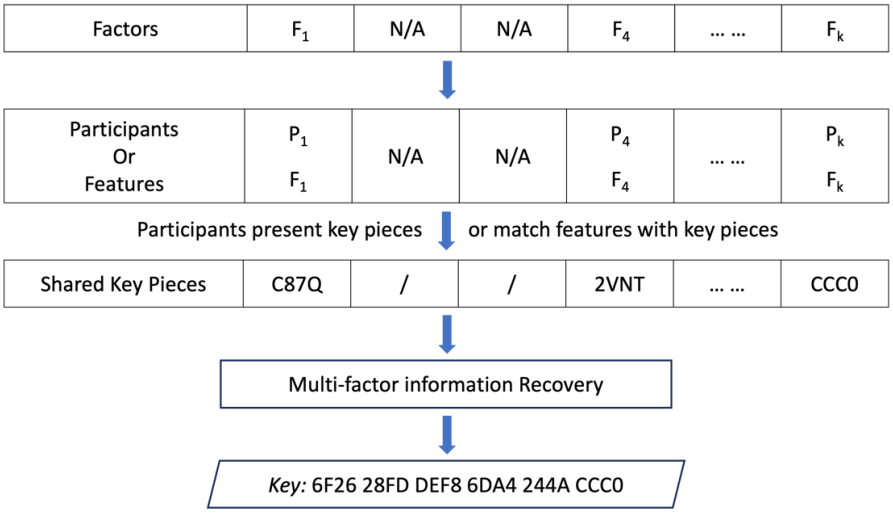

Figure 27 Recover the original key to the encrypted data from given key pieces.

The workflow of multi-factor information recovery mechanism is presented in Figure 27. In this workflow, the factors that participant in the multi-factor information sharing process should present in order to retrieve the original key for encrypted data. N/A indicates that the correspondent key holder (participant) or feature is not presented. Assume the original key is spited into k pieces, the minimum number of factors required to retrieve original key, k_{min} can be less than k, and this way any k_{min} out of k factors lead to a successful information recovery and therefore lead to the successful data decryption. In fact, the threshold can be regarded as a risk level, which can be specified in the factor sharing process. A large threshold (e.g., equals the number of factors) indicates the highest security level, while a minimal threshold of 1 indicates that the original data can be retrieved by any one of the factors (and hereby is in lowest security level).

3.4.2.3 The tools and technologies used within the component.

(1) Software dependencies. This component is implemented with Python programming language version 3.8.0. In principle, the component should be able to deploy in any Python 3 environment with or above Python3.8 (such as Python 3.8.X, 3.9.X and 3.10.X). Besides, several required libraries/plugins have to be configured for this component, includes *math*, OS and etc. All these libraries are implemented by trusted third parties and open source, hereby is safe and legal for Tango project.

(2) Tools and technologies used within the component. Several techniques are used in this component, includes SHA 128/256, AES 128/256 and Shamir Secret Sharing method. Besides, to split plain text into data chunks, the input file (in spite of the file type) will be treated as binary file and separated in terms of binary encoding. To implement the above functions, *hashlib* and *Crypto*, two open source and frequently used libraries are adopted for component development.

(3) The whole component is compatible with docker environment, and will be employed in docker, eventually results in a docker image for distribution. Since docker is a self-inclusive environment which allows the developer to configure most necessary environment dependencies, the docker image of this component should in principle be able to deploy in any virtual environments support docker image.

3.4.2.4 Dependencies with other components.

Dependencies of self-encryption and decryption techniques with multi-factor information recovery mechanisms with other components can be seen in the following aspects:

(1) Dependency with Data connector. This component will be provided as a callable interface for users or other components in Tango Project. To pass necessary parameters and data into this

component and receive the encrypted data and/or secret sharing content, input and output will be delivered to this component or users respectively with data connectors.

(2) Dependency with Fides. This component serves as a functional module, i.e., this component deal with the raw data and output the operated output. However, this component does not provide any storage service. To help those users (usecase) who don't have storage service to store the encrypted data and/or keys, fides is required in such situation.

Table 2 API of Decryption module

Decryption Interface		
Type	**parameter**	**Requirement/Explanation**
Input	Encrypted data chunks with sequence number	Unique identifier (e.g., primary key of user) that indicate the data owner
	Credential(s) for decryption	The key to the last
	User id	The id of user who is trying to access the file
Output	Decrypted file	Original plain text
	History User List	A list of user(s) who have checked this file
	(Updated) encrypted file	The id of the current user will be recorded and appended at the end of the last encrypted chunk.

Table 3 API of Encryption module

Encryption Interface		
Type	**parameter**	**Requirement/Explaination**
Input	User ID	*Unique identifier* (e.g., primary key of user) that indicate the data owner
	Data	The *plain text* for encryption
	# participants for secret sharing	The number of sub-credentials to split in the data sharing process. (ABE rather than secret sharing will be used if #participant = 1)
	Confidential level	The minimum number of key pieces that are able to retrieve the original key for encrypted data
Output	Encrypted file chunks	Encrypted data chunks with sequence id for each chunk
	Credential(s) for Decryption	when #participant > 1: multiple key segments/factors after secret sharing of the credential to the last data chunk. when #participant = 1: The encrypted key of the last data chunk given by ABE.

3.4.2.5 Features implemented

The proposed encryption-decryption module is designed to encrypt and decrypt file(s) with the following features:

- Data splitting mechanism. To benefit the distributed storage and improve the convenience of data distribution, a split mechanism is designed to split the given plain text file before encryption.
- Self-encryption. Traditional encryption methods require a user-specified password as the credential for decryption. However, the password is sometimes vulnerable. As a consequence, the encrypted data can be illegally retrieved. In this component, the key is generated from the property of data. Thanks to the HASH algorithm, the key for encryption is guaranteed to be not only unique for different files but also differ significantly even if the file has a minor change.
- Chain-correlated encryption and decryption. To convenient the data distribution and storage, the plain data will be split into data chunks before encryption.

- An API design of the interface for the encryption and decryption module along with its implementation in this component.

3.4.2.6 Software artifacts

The data splitting method, encryption and decryption methods are implemented. Secret sharing has not been implemented. Two APIs, encryption and decryption interfaces, has been partly designed. The APIs of encryption and decryption can be seen in Table 2 and
Table 3 respectively.

3.4.2.7 Current Development

Currently, the data splitting, hash value generation, chunk encryption, chunk decryption modules are implemented. Specifically, the data splitting module can split a given document into multiple pieces and merge data pieces into a complete file. Hash value generator can generate hash (unique) value for the given data chunk. Encryption and decryption encrypt and decrypt data respectively with the specified hash value or key of the given data chunk.

3.4.2.8 Future work

The following works will be done in the next stage:

(1) Confidential-level adjustable multi-factor information sharing mechanism. In the information sharing process, the minimum number of key pieces to retrieve the original key reflect the confidential level, if the minimum number of key pieces to retrieve the original key equals the number of features/participants, the confidential level reaches the highest; when the minimum number of key pieces c_{min} to retrieve the original key for data decryption is less than the number of features/participants n_{users}, keys hold by any c_{min} users/features is enough to retrieve the original key.
(2) The interface of multi-factor information sharing/recovery module
(3) Multi-factor information sharing/recovery module will be implemented.
(4) The connection between encryption/decryption module and multi-factor information sharing module should be implemented.
(5) Connection between interface of this component and data connector should be implemented.
(6) A data storage solution for users without private storage service should be discussed.
(7) Dataflow between usecase side and this component through data connector should be implemented.
(8) Integration of this component with related usecase should be implemented.
(9) Online deployment (software artifacts) and software automatic verification process should be designed for Tango framework.
(10) Documentation concerning component structure and data input/output requirement should be formulated.

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

3.5.1 Purpose of the recommendations

The recommendations set out in this section aim to provide guidance on how the technologies developed under WP3 (hereinafter, the "WP3 technologies") can be designed in a manner that best enables **security** and **privacy-preservation** for data storage and sharing, in line with the EU legal framework outlined in Annex III (hereinafter, the ''relevant legal framework''). For the purposes of this section, security and privacy preservation are given a specific meaning. The **security** of data processing is defined, for the purposes of this section, as *the protection, by means of technical measures, against unauthorized or unlawful processing of the data, and against accidental loss, destruction or damage*. This concept is concerned with the security of *any data*, whether personal or non-personal. By contrast, **privacy-**

preservation is defined, for the purposes of this section, as the *minimisation of any risks to the protection of personal data that may derive from a processing operation*. Hence, this concept relates only to *personal* data, as defined under Regulation 2016/679 (hereinafter, the "GDPR")[7].
There are a few **methodological** clarifications to be made at the outset on the nature and purpose of the list of recommendations:

- The list outlines both design strategies and requirements that, if implemented in the development of the WP3 technologies, enable or facilitate compliance with the relevant legal framework. There may be multiple alternative options to design a technological solution that enables data storage and sharing in a way that is compliant with a legal requirement. For instance, it may be that there are different technical options to ensure security or data minimisation for data storage. However, different technical options can respond to a single design strategy, i.e. they can be different modalities to implement a design strategy aimed at complying with a legal requirement. In this regard, the recommendations focus on the identification of design strategies that can be implemented to comply with the relevant legal framework, rather than prescribing a specific technical feature to adopted in practice;
- The recommendations focus on the use of WP3 technologies for specific data processing activities, i.e. data sharing and storage, with the exclusion of other legitimate uses that they could be destined to;
- The recommendations focus on compliance with legal requirements that call for a certain technological design, and consequently do not concern other non-technical actions that may be needed in a given case for legal compliance, such as organisational measures to be implemented in a given entity (e.g. preparing the documentation to inform data subjects about the processing of their data) or putting in place certain legal arrangements (e.g. non-disclosure agreements to protect trade secrets). They are thus intended to be high-level and context-independent, and do not provide guidance on aspects that are dependent on a specific context. Therefore, it is essential to bear in mind that following these recommendations does not ensure that the use of WP3 technologies is legally compliant in any given case, as additional actions may be needed depending on the circumstances. However, these recommendations aim to ensure that, as long as the design of the technologies is concerned, all the minimum appropriate steps to comply with legal requirements have been taken;
- There may be technical features that are instrumental in implementing more than one of the recommendations, e.g. encryption can achieve both data obfuscation and security of the processing.

3.5.2 Scope and structure of the recommendations

Annex III describes the EU legal framework that lays down the obligations mandating security and privacy preservation, that would apply when data is stored and shared by means of the WP3 technologies. The overview in the Annex shows that EU legislation prescribes the adoption of measures that protect both personal and non-personal data in the context of data storage and sharing. There are requirements that are specific only to personal data or to trade secrets, and there are requirements that apply irrespective of whether personal data or data containing trade secrets are being processed. When the nature of the data to be processed is unknown, the technologies must be designed in a way that complies with all of these requirements. Moreover, the incorporation of privacy-preserving technical solutions in the WP3 technologies is essential to comply with Article 25 of the GDPR that requires data protection by design and by default[8].
The sections below describe the design strategies that can be used when designing technologies, such as the WP3 technologies, to share datasets of both personal data and non-personal data that may contain

[7] 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, OJ L 119 of 04.05.2016.
[8] See Section 2.5 of the Annex for the provisions of the GDPR on data protection by design and by default.

trade secrets. These design strategies are loosely inspired by the privacy design strategies developed by J. H. Hoepman in his work *Privacy Design Strategies – The Little Bluebook*[9]. Hoepman described eight design strategies for the protection and proper handling of personal data, divided into data-oriented strategies and process-oriented strategies. The work of Hoepman provides valuable guidance for the design of privacy-preserving technologies. For the purposes of this document, some of the design strategies of Hoepman are taken into account but integrated with the additional considerations and additional strategies relevant to the specific case where both personal data and non-personal data that may contain trade secrets must be protected in data storage and sharing.
The design strategies described below are data mapping, data obfuscation, data minimisation, data abstraction, data separation, security of the processing, control over the data shared, record-keeping and demonstrability. This is meant to be a non-exhaustive list of recommended strategies to be taken into account for the design of technologies involved in data-sharing. Some of these strategies coincide with legal requirements (e.g. security of the processing, data minimisation, record-keeping and demonstrability), whereas others may just contribute to good data management and facilitate compliance with a legal requirement (e.g. data mapping, data separation and data abstraction). Alongside the description of the design strategies, some considerations are provided in each of the sections below on how they relate to, or facilitate compliance with, the relevant legal framework.

3.5.3 Mapping of the data

Several relevant legal provisions may have different implications depending on the **nature of the data** being stored or shared by automated means. First, the GDPR distinguishes between personal data and special categories of personal data, and lays down different conditions for the processing of these two categories of personal data[10]. Second, Directive 2002/58 on privacy and electronic communications (hereinafter, the ''ePrivacy Directive'')[11] sets out different conditions for the processing and storage of traffic data and location data other than traffic data[12]. Third, for the purposes of respecting the reasonable steps requirement[13] under Directive 2016/943 (hereinafter, the ''Trade Secrets Directive'' or the ''TSD'')[14], trade secrets may need to be differentiated based on their value. Given that the value of trade secrets contributes to determining whether any steps taken are ''reasonable'', when the data contains trade secrets of different value it would be necessary to distinguish them based on such value in order to apply the protective measures that are deemed more reasonable.
In light of the above, **technologies for data storage and sharing should be designed taking into account, through an initial mapping and subsequent tracking, the different categories of data that could be stored or shared**. This exercise is essential to understand the applicability of the relevant legal rules determining which appropriate protective measures should be put in place. It may be that technologies are designed to store and share datasets whose content is not predetermined and unknown. For instance, this may be the case when a technological solution is used to carry out data-sharing transactions in a systematic manner amongst multiple parties, without there being a single party that has knowledge of, or control over, all of the data that is shared. When the nature of the data to be shared or stored is not predetermined, a safe and risk-minimising strategy would be to assume that there could be

[9] J. H. Hoepman, ''*Privacy design strategies*'' (extended abstract), 2022 accessed on 1 November 2023 at https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf, p. 12.
[10] Article 6 of the GDPR lays down the conditions for the processing of all categories of personal data, whereas Article 9 lays down the conditions for the processing of special categories of personal data.
[11] 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 OJ L 201 of 31.7.2002.
[12] Article 6 of the ePrivacy Directive regulates the processing of traffic data and Article 9 the processing of location data other than traffic data.
[13] See Section 6.3 of Annex III on the reasonable steps requirement.
[14] 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 of 15.06.2016.

storage and sharing of **mixed datasets**[15] containing both personal and non-personal data, and that such data may contain business-sensitive information that qualify as trade secrets. This strategy would ensure that all the potentially applicable requirements are complied with.

Main points:

- **The appropriateness of technological design for data storage and sharing depends on the type of data to be processed;**
- **When the type of data to be processed is unknown beforehand, a safe design strategy requires to consider all the possible scenarios.**

3.5.4 Data obfuscation

3.5.4.1 General considerations

For purposes of this document, **data obfuscation** is defined as the *process of disguising the data with the aim to render it unlinkable or unobservable*[16]. Data obfuscation helps to protect the data from unauthorized access in the case where a third party, or a person within the organization of the controller or trade secrets holder, has access to the data without being so authorized. There are many techniques that can be used to obfuscate data, such as encryption or hashing, and the implementation of any data obfuscation technique when a WP3 technology is used for data storage and sharing can ensure, or facilitate, compliance with the relevant legal framework.

Data obfuscation is of fundamental importance for the protection of both personal data and trade secrets in the context of data storage and sharing. **First**, it can qualify under the TSD as a reasonable step implemented to preserve the secrecy of the trade secret. **Second**, data obfuscation has received formal recognition and explicit legal relevance under the GDPR in the form of the concept of ''pseudonymisation'', and when data obfuscation is made in a way that renders data anonymous, it renders the GDPR and the ePrivacy Directive inapplicable to the anonymised dataset. **Third**, the application of data obfuscation in a data processing operation can constitute a security and cybersecurity risk-management measure whose implementation is required, respectively, by Article 12[17] of Regulation 2022/868 (hereinafter, the "Data Governance Act" or "DGA")[18] and Article 21[19] of Directive 2022/2555 (hereinafter, the "NIS2 Directive")[20].

The qualification of data as **pseudonymous or anonymous** under the **GDPR** is at current surrounded by significant legal uncertainty. To provide for guidance on how to navigate this legal uncertainty, the sub-sections below discuss the qualification of data as anonymous or pseudonymous under the GDPR and the related consequences, followed by a subsection on the importance of data obfuscation for the protection of trade secrets and guidance on how data obfuscation could be applied to protect both trade secrets and personal data.

---

[15] In the case of mixed datasets where personal and non-personal data are inextricably linked, please bear in mind that the GDPR applies fully to the entire dataset, as clarified by the European Commission in ''Communication from the Commission to the European Parliament and the Council - Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union'', COM/2019/250 final, 2019, p. 9.

[16] J. H. Hoepman, ''*Privacy design strategies*'' (extended abstract), 2022 accessed on 1 November 2023 at https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf, p. 12.

[17] See Section 5.2 of Annex III on the security requirements imposed by the DGA for providers of data intermediation services.

[18] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJ L 152 of 03.06.2022.

[19] See Section 4.2 of Annex III on the cybersecurity risk-management obligations of the NIS2 Directive.

[20] Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS2 Directive), OJ L 333 of 27.12.2022.

3.5.4.2 The distinction between pseudonymous data, anonymous data and other personal data under the GDPR

The concept of personal data under the GDPR is introduced in Annex III[21]. This sub-section focuses on drawing the distinction in practice between personal data, pseudonymous personal data, and anonymous non-personal data.

Pseudonymous data is personal data within the meaning of the GDPR that has been subject to pseudonymisation. Pseudonymisation is defined by Article 4, point 5) of 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*". Therefore, pseudonymous data remains personal data due to the fact that the data subject it refers to remains identifiable with the use of additional information. Guidance on pseudonymisation has been provided by the Article 29 Data Protection Working Party (hereinafter, the "Wp29") as part of its Opinion on anonymisation techniques[22].

Anonymous data is not a legal concept defined in the GDPR, but it is referred to in the recitals of the GDPR[23] and guidance on it has been provided by Wp29[24] and the European Data Protection Supervisor (hereinafter, the "EDPS")[25].

There are multiple techniques that can be used to achieve pseudonymisation of a dataset. **Encryption** is a classic example of pseudonymisation technique, as it leads to the transformation of plaintext personal data into a ciphertext that acts as pseudonym. Encrypted data is pseudonymous data because, once the data in plaintext has been encrypted, the link to an identity can be re-established by combining the ciphertext (i.e. the encrypted data) and a decryption key. The ciphertext as such may not enable the identification of the data subjects that it refers to, but with the use of a decryption key it is possible to reverse the pseudonymisation process and turn the ciphertext into the original plaintext again, thus allowing whoever has the encryption key to re-identify the data subject(s).

A condition put forward by the legal definition of pseudonymous data is that **the additional information needed to re-identify the data subject is kept separately and protected by means of technical and organisation measures**. If this information is attached to the ciphertext, the data cannot qualify as pseudonymous because identification of the data subjects is as easy as it would be with the plaintext.

Therefore, it can be said that pseudonymous data differs from personal data for at least **two elements**: i) the **de-identification** of the data subjects that the personal data refers to, ii) the fact that any additional information needed to re-identify the data subjects is **not readily available** to whoever has access to the data.

Anonymous data is, by definition, non-personal data within the meaning of the GDPR. While there is no legal definition of anonymous data, the concept of anonymous data can be inferred *a contrario* from the definition of personal data, as anonymous data is any data that is not personal. In this regard, the (non-binding) Recital 26 of the GDPR describes anonymous data as "*information which does not relate to an identified or identifiable natural person*" or as originally personal data that was "*rendered anonymous in such a manner that the data subject is not or no longer identifiable*"[26]. Therefore, the

[21] See Section 2.2 of Annex III on the definition of personal data.

[22] Wp29, ''*Opinion 5/2014 on Anonymisation Techniques*'', 0829/14/EN WP216, 2014, pp. 20-23.

[23] See Recital 26 of the GDPR.

[24] Wp29, ''Opinion 5/2014 on Anonymisation Techniques'', 0829/14/EN WP216, 2014.

[25] AEPD-EDPS, '' joint paper on 10 misunderstandings related to anonymisation'', 2021.

[26] Recital 26 of the GDPR reads as follows:

''(…)

The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in

criteria on identifiability of natural persons laid down in the GDPR should be used to determine whether a given piece of information enables or not the identification of data subjects, and if it can thus be considered personal or anonymous data. The most important guidance provided by the GDPR on the concept of identifiability of natural persons can be found in Recital 26, where it is stated that "*to determine whether a natural person is identifiable, 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*" (emphasis added). This entails that, as also endorsed by Wp29 in its Opinion on the concept of personal data[27], the mere hypothetical possibility to single out a person is not sufficient to qualify that person as identifiable, but it must be proved that the identification is practically possible in the circumstances of the case, taking into account the means that the controller or another person can be reasonably expected to use.

The **''*means reasonably likely to be used*'' test** is thus central for the determination of whether information is anonymous under the GDPR. The Court of Justice of the European Union (hereinafter, the ''CJEU'') interpreted this test in its famous **''Breyer'' case**[28], clarifying that, in the context of **dynamic IP addresses**, it is not necessary that all the information enabling identification is in the hands of the same person, but it may suffice that a single person has the means to access all the information needed to identify the data subjects. In the facts of the Breyer case, dynamic IP addresses were deemed to be personal data from the perspective of online media service providers, who could rely on legal means to obtain the required additional information held by internet service providers to identify the natural person to whom a dynamic IP address relates[29].

Important clarifications, both conceptual and practical, on anonymous data have been provided by the EDPB and the AEPD in their joint paper on 10 misunderstandings related to anonymization[30].

The first clarification, of fundamental practical importance, is that for data to be anonymous 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[31]. An anonymisation process aims to reduce the re-identification risks below a certain threshold, and **this threshold 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. This criterion, however, presents the challenge of determining, on a case-by-case basis, which threshold of identifiability close to zero is accepted for data to qualify as anonymous.

The second important clarification is that **data qualifying as anonymous at a given time may not be anonymous in the future**[32]. Anonymisation may not be a permanent 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.

In light of the above, it can be noted that the distinction between anonymous and personal data is a very **factual assessment** that needs to take account of the means available to the data controller or another person to identify the natural persons based on the data.

---

such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.''

[27] Article 29 Working Party, Opinion 04/2007 on the concept of personal data, WP 136, 2007, p.15.

[28] CJEU 19 October 2016, C-582/14, *Patrick Breyer v Bundesrepublik Deutschland,* ECLI:EU:C:2016:779 ('Breyer case').

[29] CJEU 19 October 2016, C-582/14, *Patrick Breyer v Bundesrepublik Deutschland,* ECLI:EU:C:2016:779 ('Breyer case'), paras. 47-48.

[30] AEPD-EDPS, '' joint paper on 10 misunderstandings related to anonymisation'', 2021.

[31] Ibid, p. 5.

[32] Ibid, p. 4.

3.5.4.3 The status of pseudonymous data under the GDPR and the benefits of pseudonymising data

When data qualifies as pseudonymous under the GDPR, the consequence is that **the provisions of the GDPR fully apply to any processing operation performed on such data**. However, whilst the GDPR fully applies to both pseudonymous and non-pseudonymous personal 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 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) 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 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) 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) 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 explicit obligation on controllers and processors to pseudonymise data, as there might be alternative measures to comply with the provisions cited above, pseudonymisation is highly recommended as a means to comply with the GDPR, especially for complex and risky processing operations.

3.5.4.4 The relative or absolute approach to the identification of anonymous or pseudonymous data

The qualification of data as personal or anonymous is, as seen above, dependent on the identifiability of a data subject from the data itself, using the ‘’means reasonably likely to be used’’ criterion. The assessment of identifiability is evidently highly factual and its outcome would largely depend on the actor from whose perspective the availability of means reasonably likely to be used is assessed.

Therefore, an important question regarding the concept of personal data, and thus of pseudonymous and anonymous data, concerns the **point of view** that must be taken to qualify certain information as personal data. In particular, the question is whether the criteria to consider data as personal, personal pseudonymous or anonymous must be applied solely from the **perspective of the controller only (relative approach)** or **also from the perspective of third parties (absolute approach)**. The answer to this question has huge consequences for the application of the GDPR in practice. For instance, if only the perspective of the controller is relevant, it follows that data pseudonymized by a controller and shared with another controller in pseudonymized form, without providing the second controller with the information necessary to re-identify the data subjects concerned, may be anonymous data from the perspective of the second controller. As a matter of fact, the second controller may not be in possession of means reasonably likely to be used to re-identify data subjects, if it cannot acquire those from the first controller or in other ways. If, on the other hand, the perspective of other persons is also to be taken into account, taking on an absolute rather than relative approach to the qualification of personal data, the data would be pseudonymous for the second controller in the example made above.

The question of whether a relative or absolute approach should be taken under the GDPR has not yet received a definitive answer, even though the **General Court** of the CJEU has taken a stance on it in

the recent judgement on case T-557/20[33]. While this judgement was appealed before the Court of Justice, and may thus be overturned, it represents the first judicial ruling on this question at the EU level and should regarded as indicative at present of the correct interpretation of the concept of personal data.
The judgement concerned an action brought before the General Court for annulment of a decision issued by the EDPS, where the claimant argued that the interpretation of the concept of personal data made by the EDPS was incorrect as the latter adopted an absolute and not relative approach to the qualification of personal data. Even though the judgement related to Regulation (EU) 2018/1725[34], regulating the processing of personal data by EU institutions and bodies, it still concerned the concept of personal data which is equivalent to the analogous concept under the GDPR. Therefore, this judgement can be considered pertinent also in the context of the GDPR.
In the judgement, **the General Court annulled the impugned decision of the EDPS and endorsed the relative approach to the definition of personal data**, stating that, to determine whether the information transmitted to a recipient constitutes personal data, it is necessary to put oneself in the position of that recipient in order to determine whether the information transmitted to it relates to 'identifiable persons' , and in particular if the recipient can identify data subjects based on the transmitted information [35] . When the transmitted information is personal data that has been pseudonymised, it is necessary to ascertain whether re-identification is reasonably possible for the recipient of the information, taking into account the available means to access additional information that would render re-identification possible.
It follows from the judgement cited above that, since it is necessary to consider the data recipient's perspective in qualifying certain data as personal, if the recipient does not have information enabling it to re-identify the data subjects and has no legal means available to access such information, the transmitted data is anonymous for such recipient, even if it is pseudonymous for the transmitter of the data and said transmitter, or other third persons, have the means to re-identify the data subjects.
The judgement in case T-557/20 has important implications for data sharing and storage, especially for data transactions taking place in complex data-sharing ecosystems with the interaction of multiple parties. The guidance provided by the General Court on the concept of personal data can, for now and subject to the judgement's confirmation on appeal, be relied on to design technologies for data sharing and storage in a manner that data remains pseudonymous only for the parties that need to retrieve the original data at some point in the future, rendering it anonymous for all the other parties that handle the same data but do not need to visualise its content in an intelligible form. Therefore, pseudonymization of personal data may be carried out in a way that the data becomes anonymous for certain recipients or processors along the data flow.

3.5.4.5 Obfuscation for the protection of trade secrets in data storage and sharing

Obfuscation techniques such as pseudonymisation and anonymisation techniques have been largely discussed over time as measures to protect personal data and comply with the GDPR. While there is extensive legal research on obfuscation techniques to protect personal data, the discussion on obfuscation for the protection of **trade secrets** is in comparison less developed. However, obfuscation is almost as important for data, either personal or non-personal, containing trade secrets.
Since data obfuscation adds another layer of security that makes it more challenging to interpret data for unauthorised users, it may qualify as a reasonable step under Article 2(1)(c) of the TSD to be implemented by the trade secret holder to preserve the secrecy of information. As explained in Annex

[33] CJEU (General Court) 26 April 2023, T-557/20, *Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS),* ECLI:EU:T:2023:219.
[34] Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, OJ L 295 of 21.11.2018.
[35] CJEU (General Court) 26 April 2023, T-557/20, Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS), ECLI:EU:T:2023:219, para. 97.

III[36], national case-law in the EU has recognised that security measures can be necessary as reasonable steps to protect trade secrets in certain circumstances. A survey addressed to members of multiple European industries has found that, amongst all the possible protective measures that can be implemented to protect trade secrets in data sharing, data obfuscation is not recognised by respondents as the most important[37]. On the contrary, respondents identified contracts as the most important means to protect confidential and commercially valuable data, including trade secrets[38].

Nonetheless, there are reasons to argue that data obfuscation is of fundamental importance in data sharing and storage and may be a decisive element that makes the difference between the preservation or not of trade secrets. **First**, in case of unauthorised access to the shared data, e.g. during transmission, data obfuscation can render data unintelligible to the unauthorised user and protect its secrecy. **Second**, the data may have to be accessed for legitimate purposes by third persons during data sharing and storage, for instance when a third person is managing the technical solution over which the data is being shared and for this reason has access to the data. In this case, data obfuscation, for instance by means of end-to-end encryption, is essential to keep the information secret towards third parties that act only as processors. In complex data-sharing ecosystems, the trade secret holder who sends the data may not be fully aware of all the parties that can lawfully access the data being shared, which would render difficult and inconvenient to stipulate non-disclosure agreements with all of them. **Data obfuscation removes this problem by enabling the trade secret holder to have more control over who can lawfully visualise the data**.

3.5.4.6 Application of data obfuscation techniques to both personal data and trade secrets

In light of the picture drawn in the sections above, it can be noted that data obfuscation has an essential role to play when both personal and non-personal data is stored and shared in order to achieve security and privacy preservation, and comply with the requirements imposed by the relevant legal framework. When obfuscation leads to **anonymisation**, it makes the **GDPR and the ePrivacy Directive inapplicable to the data**, whereas when it leads to **pseudonymisation** it **facilitates compliance with both legal acts**. Pseudonymisation can be essential in data storage and sharing to protect the data against unauthorised access, but also to limit the number of parties that have lawful access to the personal data by rendering the data anonymous for the highest number of parties possible. For instance, the data may be pseudonymised through encryption, and encryption may be carried out in a way that re-identification of the data subjects concerned is possible only for the sender and the recipient of the data sharing. Obfuscation can also help to protect the secrecy of information and be a reasonable step in line with Article 2(1)(c) of the TSD.

It is important, however, to point out that the legal framework on personal data and the TSD may require **different types of data obfuscation**. In particular, obfuscating the link between certain information and a data subject is different from obfuscating the information necessary to extrapolate a trade secret. On the one hand, there may be pseudonymisation limited to the obfuscation of links to identifiable data subjects, while leaving other information concerning trade secrets unaltered and still intelligible for third parties. On the other hand, certain techniques may lead to both pseudonymisation of personal data and the obfuscation of trade secrets, as is the case with encryption and hashing where the plaintext is transformed into a cipher text that is expected to be unintelligible to humans.

When technologies are used to share and store datasets that may contain both personal and non-personal data, only data obfuscation techniques that ensure the unintelligibility of both personal information and trade secrets can ensure compliance with the GDPR and the ePrivacy Directive, on the one hand, and the reasonable steps requirement under the TSD, on the other. As concerns the GDPR and the ePrivacy Directive, obfuscation would lead to the most desired outcome if it renders the data anonymous for all the parties except those for which re-identification is necessary at some point.

[36] See Section 6.3 of Annex III on the reasonable steps requirement.

[37] T. Aplin and others, "The Role of EU Trade Secrets Law in the Data Economy: An Empirical Analysis", International Review of Intellectual Property and Competition Law 54, 826–858, 2023, p. 832.

[38] Ibid.

There are multiple techniques that can lead to obfuscation of both personal data and trade secrets. **Encryption and hashing**, for example, can lead to the replacement of intelligible text with unintelligible text. It must be noted, however, that there is the risk that that the hashing process is reversed revealing the original data. The assessment of this risk must be carried out on a case-by-case basis to conclude whether hashed data is pseudonymous or anonymous data, taking into account a variety of factors, including the availability of technologies that enable to reverse the hashing process. Due the risks of reversibility, hashing is generally regarded as a pseudonymisation technique[39], even though it is not to be excluded that it could lead to anonymisation.
The level of obfuscation of a set of data depends on the **re-identification technologies** that can be available at a given point in time. As technological developments occur, it may happen that data which qualifies as anonymous at a given time becomes personal data again at a later time, or that a given pseudonymisation technique becomes less secure over time. For this reason, technological developments should be monitored in order to understand which is the interplay between an obfuscation technique and the state of the art in technology.

Main points:

- **Data obfuscation enables compliance with multiple legal requirements for both personal and non-personal data storage and sharing;**
- **The point of view (relative v. absolute) to qualify data as pseudonymous or anonymous under the GDPR matters;**
- **In case of mixed datasets, only data obfuscation techniques that ensure the unintelligibility of both personal data and trade secrets can meet the requirements of the relevant legal framework;**
- **Developments in state-of-the-art technologies should be monitored over time to assess the effectiveness of data obfuscation techniques.**

3.5.5 Data minimisation

3.5.5.1 General considerations

For the purposes of this document, **data minimisation** is defined as a *data processing strategy that consists of limiting, to the largest extent possible, the processing of the data*. This concept is based on the definition provided by the GDPR[40] and by J.H. Hoepman[41], who classifies it as a data-oriented strategy, and recognizes that there are multiple tactics that can be followed to achieve data minimisation[42].
Minimisation can be achieved **in many ways and at different levels**, for instance by minimizing the collection of data at the source, by limiting its processing and the number of parties to whom it is exposed after collection, or by deleting, in part or in total, the data when it is no longer needed (storage limitation). Data minimization is an essential requirement under the GDPR. Since **data minimisation and storage limitation are two of the fundamental principles for personal data processing** that must be respected

[39] Wp29, ''Opinion 5/2014 on Anonymisation Techniques'', 0829/14/EN WP216, 2014; AEPD, EDPS, ''Introduction to the hash function as a personal data pseudonymisation technique'', 2019.
[40] Article 5(1)(c) of the GDPR reads as follows:
''*Personal data shall be:*
(…)
adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
(…)''.
[41] J. H. Hoepman, ''Privacy design strategies'' (extended abstract), 2022 accessed on 1 November 2023 at https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf, p. 5.
[42] Ibid.

under Article 5 of the **GDPR**, the processing of personal data must always be limited to what is necessary in relation to the purposes of the processing. Moreover, storage limitation is a requirement under the **ePrivacy Directive** in relation to traffic data, and location data other than traffic data, as Articles 6 and 9 of the Directive prescribe that such data is processed only to the extent and for the duration that is necessary. As concerns trade secrets, data minimisation may be an important strategy to protect them against unintended disclosure, and would likely qualify as a reasonable step that must be adopted under Article 2(1)(c) of the **TSD** when sharing and storing data across. Due to the large number of processors or parties to whom the data may be exposed when shared and stored, it is essential that the trade secret holder minimises its processing in a way that reduces, to the largest extent possible, that the secrecy of the data is compromised. For instance, trade secrets could be protected with data minimization techniques consisting of access limitation, data avoidance and deletion of business-sensitive copies when no longer needed.

It must be noted that the nature of data minimization techniques to be implemented in practice can differ greatly depending on whether the aim is to protect personal data or trade secrets. On the one hand, the GDPR aims to limit personal data processing due to the fact that the processing of personal data is, *per se*, an interference with the right to data protection of the data subject and must thus be limited to what is strictly necessary. On the other hand, data minimization is relevant for trade secrets to the extent that it preserves their secrecy, by mitigating the risk that trade secrets are exposed to parties that should not have access to them. Therefore, it may happen that an unnecessary processing of the data by a person who has lawfully access to it would not pose a problem for trade secrets protection, whereas it could be a violation of the data minimisation principle under the GDPR when it goes beyond what is necessary in relation to the data processing purposes. In many cases the GDPR may require a more pervasive level of data minimization, because its rationale includes, but goes beyond, the prevention of unlawful acquisition, use and disclosure of the data.

Nevertheless, technologies could be designed in a way that, by achieving minimisation of data processing, would contribute in any case to the preservation of trade secrets and to the protection of personal data. This does not exclude that the baseline level of protection ensured through design strategies may need to be complemented with additional measures that are needed on a case-by-case basis, given that compliance with the GDPR, ePrivacy Directive and the reasonable steps requirement relies on a **case-specific contextual assessment**. An overview of three important data minimisation techniques is provided below. The list below is not meant to be exhaustive, but aims to set out some of the most discussed minimisation techniques that can be implemented by design and by default to protect both personal data and trade secrets.

3.5.5.2 Data avoidance and limitation

Data avoidance and limitation are data minimisation techniques that the EDPB recognises as key design and default data minimisation elements in the context of the GDPR[43]. They entail design and default features where the processing of data is avoided altogether or limited in light of what is necessary. For trade secrets protection, data avoidance and limitation has not yet received recognition as a reasonable step in the national case-law of EU Member States. However, these elements can play a crucial role by avoiding or limiting a multiplication of the processing of data containing trade secrets. Information on trade secrets may be copied and stored in multiple locations for a given purpose, for instance to ensure data integrity. However, the multiplication of copies containing trade secrets, or the distribution of the entire piece of information displaying the trade secrets in multiple locations, increases the risks of unauthorised acquisition, use and disclosure of the data. Data minimisation by design and default reduces this risk.

3.5.5.3 Access limitation

Access limitation is a data minimisation technique that has received recognition for both personal data and trade secrets protection.

[43] EDPB Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0, 2020, p. 21.

As concerns personal data, the EDPB has listed it as one of the key design and default data minimisation elements[44]. The EDPB has defined access limitation as the practice to ‘*’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*’’. By minimising the number of persons that can access the data, the intrusiveness into the informational privacy of data subjects is also minimised.
Access limitation has also been recognised as an important measure to protect trade secrets. Access limitation measures have been recognised as reasonable steps to protect trade secrets in the case-law of Austria[45] and Spain[46], and their importance has also been affirmed in legal doctrine[47]. It can be said that access limitation is a core objective that must always be pursued to protect trade secrets, as limiting access to secret information is essential to keep such information secret.
Access limitation can be implemented at different levels. For instance, it can be put in place either at the **organisational level** by defining an access policy and access rules, at the **physical level** with physical restriction of access and surveillance measures, at the **legal level** by imposing non-disclosure obligations. With regards to the technical level, which is relevant for the purposes of the document, examples of access limitation measures include identity verification and private use restriction measures.

3.5.5.4 Partial or total deletion

The **deletion**, in part or in total, of data that is no longer needed allows tailoring the duration and extent of a processing operation to what is strictly necessary.
Under the GDPR, data deletion enables compliance with two data processing principles: data minimisation and storage limitation. For this reason, the EDPB has recognised data deletion as a key design and default element for both data minimisation[48] and storage limitation[49]. In particular, the EDPB has recommended to delete or anonymise personal data that is not necessary for the purpose for which they are being collected or processed, putting in place specific procedures and functionalities to ensure that deletion or anonymisation are carried out as soon as needed. A functionality could consist of automation of personal data deletion[50] at a fixed date or upon other conditions.
With regard to trade secrets, the trade secret holder may be interested in retaining commercially valuable information even when it appears that it is no longer useful for a specific purpose, due to its commercial value that could render it useful in the future. Therefore, data deletion may not need to happen under the same conditions as for personal data. Nonetheless, data deletion can contribute to trade secrets protection when it is applied to unnecessary copies of trade secrets. By limiting the existence of trade secrets’ copies to what is strictly necessary, the risks of disclosure due to unlawful acquisition are also minimised. Different data deletion policies and functionalities may need to be implemented for personal data and data containing trade secrets. Thus, a potential challenge may arise when the data minimization and storage limitation principles under the GDPR would impose the deletion of trade secrets that the holder has commercial interests in retaining. In this case, the holder would be obliged to delete the trade secret despite its economic interest in the data since the TSD does not grant a right to preserve the integrity of the trade secret[51], but merely protects against the unlawful acquisition, use and disclosure of the trade secret[52].

[44] Ibid.
[45] Austrian Supreme Court, Austria, Decision n◦ 4 Ob 165/16t of 2016.
[46] Provincial Court of Madrid, Spain, Decision nr. 441 of 2016.
[47] M. De Vroey, M. Allaerts, ‘’Trade secrets protection: an interim update of Belgian and EU case law’’, Journal of Intellectual Property Law & Practice, Vol. 16, No. 12, 2021, p. 1394.
[48] EDPB Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0, 2020, p. 21.
[49] Ibid, p. 25.
[50] Ibid.
[51] In this case there would not be a conflict between the EU legislation on personal data and on trade secrets.
[52] J. Drexl, ‘’Data access and control in the era of connected devices’’, Report for the European Bureau of Consumers’ Union (BEUC), 2018, p. 101.

Main points:

- **Minimisation is both a legal requirement under the GDPR and a useful design strategy to limit exposure of trade secrets;**
- **Minimisation ensures also storage limitation, which is a legal requirement under the GDPR and a useful design strategy for excessive copies of trade secrets;**
- **Minimisation is to be assessed based on the purposes of the processing and technical feasibility;**
- **Besides data obfuscation (discussed above), key data minimisation techniques are: data avoidance and limitation, access limitation, partial or total deletion.**

3.5.6 Data abstraction

Data abstraction is a strategy that consists of *limiting as much as possible the detail in which personal data is processed*[53]. The distinction between data abstraction and data minimisation lies in the fact that data abstraction is not about avoiding the unnecessary processing of data, but it is about **limiting the level of detail in which data is processed**.

Data abstraction contributes to the protection of both trade secrets and personal data. On the one hand, data abstraction limits the availability of information on the secrets and thus mitigates the risks of damage to the commercial interests of the trade secrets holder in case of unauthorised access or disclosure. On the other hand, limiting the level of detail in which personal data is processed lowers the privacy risks associated with the processing. Furthermore, by ensuring that only the most relevant personal data is processed, data abstraction contributes also to compliance with the data minimisation principle of the GDPR.

Data abstraction can be achieved in different ways, for instance by summarising the information to be processed in a manner that all unnecessary information is left out. When the WP3 technologies are used for data storage and sharing, data abstraction can be used in multiple contexts. For instance, the record-keeping of processing activities could be structured in a manner that there is as abstract information as possible on the personal data and trade secrets that have been subject to processing.

Main points:

- **Data abstraction is not a legal requirement per se, but it is recommended to protect personal data and sensitive information.**

3.5.7 Data separation

Data separation is a strategy that consists of *separating, logically or physically, the processing of data*[54]. Separation could be implemented in different ways, but the ultimate objective would be always to **avoid a centralized processing of the data** where a single entity has control over all the processing operations. An example of separation in the context of cryptography is secret sharing, where a secret is distributed across a group in a way that there is no single participant of the group that holds intelligible information about the secret. The secret can be subsequently reconstructed when a sufficient number of participants in secret sharing combine their shares.

The benefit of separation is to avoid that a single person or entity has control over the whole processing operation(s), which would allow to have a full view of a dataset or of multiple datasets, and to understand correlations between datasets. When unintended, the establishment of correlations between data could be detrimental both for personal data protection and trade secrets protection. When personal data is split

[53] J. H. Hoepman, ''Privacy design strategies'' (extended abstract), 2022 accessed on 1 November 2023 at https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf, p. 10.

[54] J. H. Hoepman, ''Privacy design strategies'' (extended abstract), 2022 accessed on 1 November 2023 at https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf, p. 8.

into different chunks in a way that a single chunk does not allow for the identification of the data subject(s), separation would lead to the creation of pseudonymous data, as the data would not be attributable to a specific data subject without the use of the other data chunks that are kept separately. Separation can contribute to the protection of personal data by **preventing the concentration of control over personal data in the hands of a single entity**, as well as by **pseudonymising personal data** when separation is carried out in a way that meets the requirements of the definition in Article 4(5) of the GDPR. Separation can also constitute a **reasonable step for the protection of trade secrets** under Article 2(1)(c) of the TSD. By separating information on a trade secret in different data chunks, in a way that each chunk alone does not enable to acquire knowledge of the trade secret, the trade secret is protected as there is not a single entity that has sufficient information to have access to the secret. For instance, if secret sharing is structured in a way that none of the participants has access to the trade secret through their shares, the full content of the secret can be reconstructed only with the combination of a sufficient number of shares from the participants.

Main points:

- **Data separation is not a legal requirement per se, but it is recommended to protect personal data and sensitive information.**

3.5.8 Security of data processing

3.5.8.1 General considerations

The **security** of data processing is defined, for the purposes of this document, as the *protection against unauthorized or unlawful processing, or against accidental loss, destruction or damage, of the data*. This definition corresponds to the integrity and confidentiality requirement defined in Article 5(1)(f) of the GDPR[55], and is intended to be a broad concept encompassing all the organizational and technical measures put in place to ensure the security of the data. When the processing of data takes place by automated electronic means, the concept of security partially overlaps with that of information security, a domain of cybersecurity defined by the European Union Agency For Network and Information Security (''ENISA'') as the *''protection against the threat of theft, deletion or alteration of stored or transmitted data within a cyber system"*[56].

Security of the processing is essential for each of the WP3 technologies when used for data storage and sharing. It is necessary to protect personal data in line with legal requirements, and it constitutes a reasonable step to protect information containing trade secrets when it is shared and stored by automated means. Moreover, security requirements are also imposed by the DGA and NIS2 Directive, that apply irrespective of whether the processing of personal data or trade secrets is involved[57].

Since security is a common obligation in different relevant legal instruments, the sections below discuss the elements of convergence that can be identified between the security requirements of different legislative acts. This exercise helps for the design of secure technologies, with the aim to set out security features that, when implemented, enable or facilitate compliance with all the obligations of the relevant legal frameworks.

[55] Article 5(1)(f) reads as follows:
''Personal data shall be:
(…)
(f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').''

[56] ENISA, ''Definition of cybersecurity, gaps and overlaps in standardisation'', 2015, p. 11.

[57] Article 12 of the DGA, however, requires a higher level of security of competitively sensitive information.

3.5.8.2 Common elements in the security requirements of the relevant legal instruments

The relevant security obligations in the GDPR, the ePrivacy and NIS2 Directives, and the DGA (see below, Annex, for more details) contain a series of common elements that can be identified and that translate in common compliance obligations to ensure the (cyber)security of data storage and sharing. Moreover, it must be noted that while the TSD does not impose security obligations in the manner the other relevant legislative acts do, trade secrets protection calls in practice for the adoption of security measures. Ensuring the security of data processing through protection measures represents one of the reasonable steps that must be adopted to preserve the secrecy of the data in accordance with Article 2(1)(c) of the TSD, especially when trade secrets are processed by automated means in complex processing environments[58]. Whenever data containing trade secrets is shared and stored by automated means, information security of the technological solutions used to share and store data containing trade secrets is necessary to preserve the secrecy of the data.

The most evident common element lies in the fact that all security obligations call for a **case-specific proportionality assessment** that takes account of the risks posed, the costs of implementation and the state-of-the-art available technology. In principle, there is not a single solution, or list of solutions, that would ensure compliance in any case with all the relevant security obligations imposed by EU legislation, as the most appropriate measures must be determined on a case-by-case basis, even though the ePrivacy and NIS2 Directives provide for a minimum standard of security to be ensured in any case.

A second common feature of all the security obligations under consideration concerns the relevance of **technological developments** and the consequent duty of controllers, trade secret holders and other obliged entities to **monitor** such developments, in order to **adapt** their security measures where appropriate and ensure that **state-of-the-art** solutions are in place. The monitoring of technological developments is necessary to ensure that any security measures in place remain appropriate to the risks to address, should existing measures become obsolete over time, especially where there are new technologies that empower attackers to more easily breach security safeguards. Moreover, all of the legal acts under consideration require, either explicitly or implicitly, to take into account the state of the art and the latest technological developments[59].

A third common feature of all the security obligations under consideration is that they are **obligations of means**, and not of results. As a consequence, the sole adoption of appropriate measures that meet the relevant requirements should result in compliance with the legal framework, even where a security breach occurs. The qualification as obligations of means is evident for Article 32 of the GDPR[60] and the reasonable steps requirement of the TSD.[61] For the security obligations in the ePrivacy Directive, NIS2 Directive and the DGA, it can be argued that they are also obligations of means, as they are formulated in a similar manner to Article 32 of the GDPR. In particular, the relevant articles all require the adoption of measures that meet certain characteristics, namely provide for a certain level of security taking into account the circumstances of the case, but they do not prescribe a specific result of ensuring that no incidents occur. On the contrary, the potential occurrence of incidents is recognized as an event that may not be always preventable, and for which obliged entities must have incident management measures in place[62].

Finally, there are **consolidated security standards** that are commonly regarded as best practices for (cyber)security, which can be a common starting point to achieve compliance with all the pieces of legislation under consideration. While adaptations may be needed based on the circumstances of the case and the specificities of each legal act, standards like those of ETSI and ISO/IEC constitute a useful starting point that might either be sufficient to comply with the relevant security obligations, or might just need some additional adaptations.

---

[58] See Section 6.3 of Annex III.

[59] See, in particular, Article 32 of the GDPR, Article 21 of the NIS2 Directive, Article 4 of the ePrivacy Directive and Article 12(l) of the DGA.

[60] See Section 2.3.6 of Annex III.

[61] See Section 6.3 of Annex III.

[62] See, in particular, Article 32 of the GDPR, Article 21 of the NIS2 Directive and Article 12(l) of the DGA.

Despite the common elements set out above, the **requisite level of security may differ** under the relevant legal frameworks. These pieces of legislation pursue different objectives and regulate different subject matter, even though they can all come to play in the context of data sharing. Nonetheless, the common elements are relevant from a compliance perspective, as two of them refer to the steps that must be taken when putting in place security measures, namely a context-specific proportionality assessment and the monitoring of state of the art technologies, and one of them delineates the scope of the obligations, i.e. their nature as obligations of means.
It can be concluded that these common elements can be translated in steps to be taken in order to comply with multiple legal frameworks that are relevant to the security of data sharing and storage with WP3 technologies, both when protective measures must be adopted for the first time and subsequently when their appropriateness to the risks presented is assessed on a continuous basis.

Main points:

- **Security is a legal requirement under all the relevant pieces of legislation;**
- **Despite the differences in personal data and trade secrets legislation, they might call for similar security safeguards;**
- **The security obligations under the relevant legal instruments present similar characteristics, i.e.:**
  - **Relevance of circumstances of the processing;**
  - **Duty to monitor developments in state-of-the-art technologies to measure effectiveness of security safeguards;**
  - **Adoption and update of security measures is an obligation of means;**
  - **Importance of taking into account standardised cybersecurity standards.**

3.5.9 Control over the data

Control of the data shared during the lifecycle of the data sharing and storage process is necessary to comply with legal requirements from the GDPR, the TSD and the ePrivacy Directive. By way of example, access limitation policies must be continuously enforced to protect personal data and trade secrets, and the GDPR and the ePrivacy Directive both require that data is deleted or anonymized when no longer needed.
Control over the data is needed to enforce previously defined conditions for data processing, but also to modify or introduce new conditions for data processing. Conditions surrounding the processing of data may change for both personal data and trade secrets. **First**, the GDPR confers specific rights on data subjects which may be exercised at any time while personal data is being processed[63]. When these rights are exercised, it may be that either there is a change in the conditions under which the processing can take place, or that the processing cannot take place anymore (e.g. following a request to erasure under Article 17 of the GDPR). **Second**, when the processing is based on consent under the GDPR and the ePrivacy Directive[64], the withdrawal of consent entails that the processing may not take place anymore. **Third**, policies on the handling of trade secrets may change over time, for instance when new threats to the secrecy of information arise, or the information becomes commercially more valuable and stricter protection measures are warranted. When this happens, the trade secret holder may need control over the data during the data sharing and storage lifecycle in order to change the policies governing access to trade secrets and subsequently enforce them.

Main points:

- **Control over the technical conditions that govern data processing is needed over time;**

[63] See Section 2.4 of Annex III.
[64] See Sections 2.3.2 and 3.2 of Annex III.

- **It must be technically feasible to timely adjust how the data processing takes place, or interrupt it, when the legal conditions for such processing change.**

3.5.10 Record-keeping and demonstrability

When one or more of the precautions outlined above have been implemented, it is essential that this is **recorded in a way that proof of their implementation can be provided at any time** for as long as needed **to demonstrate compliance** with the relevant legal instruments.

Specific record-keeping and demonstrability obligations are explicitly fleshed out in some of the legal acts that compose the relevant legal framework. Article 30 of the GDPR sets out record-keeping obligations for processors and controllers, and controllers are subject to the principle of accountability which is enshrined in Article 5(2) of the GDPR as a general data processing principle[65]. Moreover, the ePrivacy Directive, the NIS2 Directive and the DGA all state that **supervisory authorities should be able to audit obliged entities to verify compliance** with, among others, security obligations. This entails that obliged entities should be able to present **records** of the relevant activities at any time when requested[66]. As concerns trade secrets, record-keeping and demonstrability are also necessary to be able to prove, *ex post*, that the reasonable steps requirement was met, i.e. to demonstrate which reasonable steps were adopted in practice. Record-keeping is thus essential to enforce trade secrets protection in case of unlawful acquisition, use and disclosure of the secrets, as having taken reasonable steps is one of the requirements for information to qualify as a trade secret in the first place.

Due to the partial differences in the security obligations laid down in the different legal instruments, the relevant elements to keep record of for evidentiary purposes may also be different. Moreover, given the more elaborate set of obligations imposed by the GDPR, the consequent record-keeping activities to be carried out may be significantly more extensive compared to those needed to prove that reasonable steps were adopted under the TSD. However, the minimum needed activities to keep record of such elements may be the same, even though the elements relevant to prove compliance *ex post* may be different from time to time. In particular, it may be in any case necessary to keep record of data-sharing transaction details, e.g. by means of an audit log, which can be used for auditing purposes at a later stage. For instance, the data transaction details to be recorded may relate to: i) the processed **data**, ii) the processing **activities** carried out, iii) the **persons** that had access to the data and the access limitation measures in place, iv) the **technologies** used for the processing, v) the technical and organizational **measures** adopted to ensure security of the processing.

Main points:

- **Records must be kept of all the measures implemented to comply with the legal requirements, including the implementation of the design strategies listed above;**
- **Record-keeping should be automated when possible, e.g. with an audit log.**

[65] See Section 2.3.7 of Annex III.

[66] See Article 4(1a) of the ePrivacy Directive, Article 32 of the NIS2 Directive, Article 14 of the DGA. Additionally, Article 12(o) of the DGA requires providers of data intermediation services to maintain a log record of the data intermediation activity.

4 Conclusions

This deliverable outlines the preliminary outcomes through an interim iteration 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 account 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 ongoing release and forthcoming work for each task and component is provided in the subsequent table.

T3.1 norbloc is a Regtech company with HQ in Sweden, offering solutions for data sharing and onboarding. Our solution Fides, which is a decentralized data-sharing platform, will be utilized in Tango in Pilot 5 (T7.6 – Public organizations) to share VISA data between VISARIGHT and one or more German authorities. Functionalities exploitable in TANGO include explicit consent management, private data sharing in regulated environments, and immutable audit logs. Fides nodes will be installed at Pilot participants' premises, for peer-to-peer data sharing, and will provide the software for building distributed storage (the underlying storage/ cloud environment must be provided within Tango). The Fides internal components that will be utilized in TANGO are: 1) Proprietary norbloc data backend: DRILL – fully utilized in TANGO, 2) Proprietary norbloc access management system: DnAMS – partially utilized in TANGO, 3) Blockchain (any private blockchain can be used, by default, Fides offers Hyperledger Fabric blockchain) – fully utilized in TANGO. The exact format of the data stored is not limited by Fides and can be made compatible with IDS/GAIA-X/DID specifications.

T3.2 Trust in data sharing has multiple facets an highly depends on regulations and company policies. In T3.2 three components support privacy and further policies by scoring and profiling capabilities. The stage of development has been outlined and a demonstration tool giving insight into the core features of Trustworthiness Scoring has been described. In further steps the components will be aligned closely with the pilots and integrated into the TANGO architecture contributing to policy drive access and sharing of data.

T3.3. To be aligned with the recommendations defined in T3.5. to enable security and privacy in the data sharing process in TANGO, task 3.3. offers two interconnected mechanisms: sticky policies and user consent management. The first one provides an extra level of confidentiality by empowering data owners to set and enforce policies throughout the data lifecycle on the data sharing process. The latter ensures GDPR compliance by implementing fine-grained access control mechanisms that enable the application of user consent policies. The processes within the architecture and the FIWARE connector, taken as a reference for the project, have been defined. In the next steps it is expected to introduce the outlined functionalities to achieve the privacy and confidentiality by design objective.

T3.4. To offer a multi-factor information sharing based data encryption and decryption mechanism for confidential data storage and distribution, this task intends to provide an open-to-use toolkit for data encryption/decryption and key sharing/recovery toolkit. In particular, the encryption and decryption module guarantee the safety of confidential data in the storage period or data transmission process. Moreover, the key sharing/recovery mechanism ensures the key for data retrieval can be recovered by multiple secret sharing participants. Currently, the former part, i.e., data encryption and decryption module has been implemented, while information sharing mechanism will be implemented soon. In the future, this task will result in a user-friendly API, and user can encrypt/decrypt data by treating this module as black box toolkit.

T3.5 After mapping the legal requirements applicable to the use of the technologies of WP3, and the related design recommendations outlined in Section 3.5, preliminary conclusions have been provided in Annex II with regard to the compliance of each technological solution with the outlined set of recommendations. Trustworthy data sharing of Section 3.2 has not been assessed in light of the recommendations due to the fact that the relevant legal framework does not apply to its operation. All these preliminary legal conclusions will be further assessed, and a definitive version will be provided in the second iteration of the document.

Generally, the intermediate release marks the delivery of an initial functional prototype intended for integration into the TANGO platform and subsequent testing by users. The majority of components operate independently, implying that immediate future efforts will be directed towards achieving the nearest technical milestone. Consequently, partners contributing to the current release of WP3 software artifacts will closely adhere to guidelines from WP6, facilitating the integration and testing of components within the TANGO platform architecture. Each task delineates, within their respective sections of the document, an initial projection of future work leading up to the final release of WP3.

5 Bibliography

A. de Streel, C. Hocepied, ‘’The Regulation of electronic communications networks and services’’, book chapter in Garzaniti et al, ‘’Electronic Communications, Audiovisual Services and the Internet – EU Competition Law & Regulation’’, 4th edition, Sweet and Maxwell, 2019

B. Van Alsenoy, ‘’Liability under EU Data Protection Law From Directive 95/46 to the General Data Protection Regulation’’, JIPITEC 271, 2017

BEREC, ‘’Report of 12 February 2016 on enabling the Internet of Things’’, BoR(16)39, 2016

ENISA, ‘’Definition of cybersecurity, gaps and overlaps in standardisation’’, 2015

European Union Intellectual Property Office (‘’EUIPO’’), ‘’Trade secrets litigation trends in the EU’’, IPR Enforcement Case-Law Collection, 2023

G. Carovano, M. Finck, ‘’Regulating data intermediaries: The impact of the Data Governance Act on the EU’s data economy’’, 50 Computer Science & Law Review 7, 2023

H. Richter, ‘’Looking at the Data Governance Act and Beyond: How to Better Integrate Data Intermediaries in the Market Order for Data Sharing’’, 72 GRUR International 462, 2023

J. H. Hoepman, ‘’Privacy design strategies’’ (extended abstract), 2022 accessed on 1 November 2023 at https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf

J. Drexl, ‘’Data access and control in the era of connected devices’’, Report for the European Bureau of Consumers’ Union (BEUC), 2018

L. von Ditfurth, G. Lienemann, ‘’The Data Governance Act: – Promoting or Restricting Data Intermediaries?’’, Competition and Regulation in Network Industries, 2022

M. De Vroey, M. Allaerts, ‘’Trade secrets protection: an interim update of Belgian and EU case law’’, Journal of Intellectual Property Law & Practice, Vol. 16, No. 12, 2021

T. Aplin and others, ‘’The Role of EU Trade Secrets Law in the Data Economy: An Empirical Analysis’’, International Review of Intellectual Property and Competition Law 54, 826–858, 2023

T. Bobev, V. K. Dessers, C. Ducuing, M. Fierens, A. Palumbo, B. Peeters, L. Stähler, ‘’CiTiP White Paper on the Definition of Data Intermediation Services’’, 2023 accessed on 7 November 2023 at https://ssrn.com/abstract=4589987

6 Annexes

Annex I – OVERVIEW OF DESIGN RECOMMENDATIONS

Recommendations for technological design	Related provisions in the relevant legal framework	Examples of practical implementation of the recommendations
Data mapping ***(predetermination on the type of data to be processed)***	The nature of the data is a precondition to apply the GDPR, the ePrivacy Directive and the TSD	Predetermination of the type of data to be stored and shared with the technologies, or predetermination that the nature of the data is unknown
Data obfuscation ***(disguising the data with the aim to render it unlinkable or unobservable)***	Article 5(1)(f) GDPR, Article 6(4) GDPR, Article 25 GDPR, Article 32 GDPR, Article 89(1) GDPR, Article 4 ePrivacy Directive, Article 2(1)(c) TSD	Encryption, hashing, removal of links to data subjects and trade secrets
Data minimisation ***(limiting, to the largest extent possible, the processing of the data)***	Article 5(1)(c) GDPR, Article 5(1)(e) GDPR, Article 25 GDPR, Article 6 ePrivacy Directive, Article 9 ePrivacy Directive, Article 2(1)(c) TSD	Partial or total deletion, anonymization, data avoidance, access limitation
Data abstraction ***(limiting as much as possible the detail in which personal data is processed)***	Article 5(1)(c) GDPR, Article 25 GDPR, Article 2(1)(c) TSD	Aggregation, summarising detailed information
Data separation ***(separating, logically or physically, the processing of data)***	Article 5(1)(c) GDPR, Article 5(1)(e) GDPR, Article 25 GDPR, Article 32 GDPR, Article 4 ePrivacy Directive, Article 2(1)(c) TSD	Storage and sharing of data in separate data chunks, secret sharing scheme
Security ***(protection against unauthorised or unlawful processing, or against accidental loss, destruction or damage, of the data)***	Article 5(1)(f) GDPR, Article 25 GDPR, Article 32 GDPR, Article 4 ePrivacy Directive, Article 2(1)(c) TSD, Article 21 NIS2 Directive, Article 12(j) and (l) DGA	Encryption, hashing, incident management, backups/logs, firewalls, any solution for network security
Control over the data ***(ability to change the processing conditions of the data at any time during the processing)***	Articles 15-22 GDPR, Articles 6 and 9 ePrivacy Directive, Article 2(1)(c) TSD	Sticky policies that can be updated over time
Record-keeping and demonstrability ***(recording of how processing of the data takes place in order to demonstrate compliance with the relevant legal framework)***	Article 5(2) GDPR, Article 30 GDPR, Article 12(o) DGA, and all the other legislative acts for the purposes of proving compliance	Establishment and update of a secure audit log for data processing activities

Annex II - ASSESSMENT OF DESIGN RECOMMENDATIONS IMPLEMENTATION IN WP3 TECHNOLOGIES

1. Blockchain-based data storage and sharing

Compliant features	Compliant features	Non-compliant features
Data mapping	Predetermined that all types of data may be stored and shared. No pre-assumptions about the type of data, the most stringent requirements are assumed.	None
Data obfuscation	Encryption and hashing	None
Data minimisation	Every participant node has as a plaintext data only the data of their clients for which an explicit consent was given.	Decentralised storage and use of blockchain inherently in conflict with data minimization – assess that any processing is strictly necessary
Data abstraction	Logs contain as little information as possible	None
Data separation	When combined with self-encryption and decryption there is separation in data chunks	None
Security	Security by design	None
Control over the data	By design the data can be accessed only having explicit data subject consent - enforced by DnAMS	None
Record-keeping and demonstrability	Automated Audit Log in the blockchain	None

2. Trustworthy data sharing

Due to the fact that this solution does not involve the processing of personal data, and taking into account its role in the data sharing process, the relevant legal framework is mainly not relevant to its operation. It may only be relevant for the security requirements described in Section 3.5.8. as it contributed to the overall level of security of data sharing and is a preventive measure against unlawful access to the data.

3. Self-encryption and decryption techniques

Recommendations for technological design	Compliant features	Non-compliant features
Data mapping	No assumptions on the type of data to be processed, the most stringent requirements are assumed	None
Data obfuscation	State-of-the-art encryption and hashing	None
Data minimisation	No unnecessary processing of the data takes place	None
Data abstraction	Logs contain as little and abstract information as possible	None

Recommendations for technological design	Compliant features	Non-compliant features
Data separation	Separation of the data in chunks takes place	None
Security	High level of security by design implemented through state-of-the-art encryption and hashing	Users can adjust the security level based on their own reflection on specific scenario.
Control over the data	Use encryption and decryption key as data access control	None
Record-keeping and demonstrability	Available audit log	None

4. Confidentiality and privacy by design

Recommendations for technological design	Compliant features	Non-compliant features
Data mapping	CP-ABE module does not directly perform data mapping, but it can contribute to compliance with data mapping practices.	None
Data obfuscation	CP-ABE module ensures that even if there is non-legitimate access to the data, it will remain unintelligible without the appropriate attributes or access policies.	None
Data minimisation	CP-ABE module enables granular access control- access is only granted to the entities with the necessary attributes, instead of providing broad access. It allows selective disclosure, meaning that the user who requests access to data will only reveal his specific attributes relevant to the access policy.	None
Data abstraction	Selective disclosure and granular access control enable focusing on the essential elements required, limiting as much as possible the personal data that is being processed.	None
Data separation	Granular access control based on attributes facilitate data separation by ensuring that users can only access the data that is relevant for their attributes/roles.	None
Security	CP-ABE module encrypts data ensuring access and usage control, which protect data from unauthorized processing.	None
Control over the data	Control usage provided by sticky policies.	None
Record-keeping and demonstrability	Available audit log	None

Annex III – Legal framework relevant to T3.5

Contents of Annex.3

1. INTRODUCTION AND PURPOSE OF THE ANNEX 65
2. THE GENERAL DATA PROTECTION REGULATION (GDPR) 66
2.1. Scope of application and applicability to TANGO 66
2.2. Definition of personal data 67
2.3. Data protection principles 68
2.3.1. Introduction 68
2.3.2. Lawfulness, fairness and transparency 68
2.3.3. Purpose limitation 69
2.3.4. Data minimisation and storage limitation 69
2.3.5. Accuracy 69
2.3.6. Integrity and confidentiality 70
2.3.7. Data protection principles: accountability 70
2.4. Exercise of data subjects' rights 70
2.5. Data protection by design and by default 72
3. THE ePRIVACY DIRECTIVE 73
3.1. Scope of application and applicability to TANGO 73
3.2. Relevant provisions 76
3.3. Interplay between the ePrivacy Directive and the GDPR 77
4. THE NIS2 DIRECTIVE 78
4.1. Scope of application and applicability to TANGO 78
4.2. Relevant provisions 79
5. DATA GOVERNANCE ACT 80
5.1. Scope of application and applicability to TANGO 80
5.2. Relevant provisions 81
6. TRADE SECRETS DIRECTIVE 83
6.1. Scope of application 83
6.2. Relevance of trade secrets legislation for the design of WP3 technologies 84
6.3. The 'reasonable steps' requirement in the TSD 84

1. INTRODUCTION AND PURPOSE OF THE ANNEX

This Annex identifies the legislative acts in the legal order of the European Union that are relevant to the design of WP3 technologies to ensure secure and privacy-preserving data sharing. They form the basis on which the recommendations set out in Section 3.5 above have been drafted. The overview provided herein is limited to the pieces of legislation, and their provisions, prescribing specific requirements that may influence the design of technologies to be used for secure and privacy preserving data sharing, with the exclusion of the legislative acts and provisions that are not relevant in this respect. Therefore, pieces of legislation that are relevant for data sharing, but that do not set out privacy and security requirements, are not included in the analysis.
Therefore, this Annex considers only the legislative acts that contain the following requirements, as applicable in the context of automated data sharing and storage:

- Requirements on security of data processing, or on the (cyber)security of technical solutions to be used for data processing;
- Requirements on the protection of trade secrets;
- Requirements on access to, and storage of, the processed data;
- Requirements on the rights that can be exercised on personal data while it is being processed;
- Requirements on record-keeping of data processing operations, as necessary to demonstrate with the requirements above.

The relevant legislation is identified starting from the assumption that WP3 technologies are to be used for sharing and storing mixed datasets that may contain both personal and non-personal data, as well as commercially sensitive data that meets the requirements set by EU legislation to qualify as trade secrets.
The sections of the Annex provide a description of the relevant legislative requirements, as identified based on the requirements set out above, and briefly discusses the applicability of the relevant EU legislative acts to the operation of the WP3 technologies in the context of the TANGO architecture. This Annex only considers EU legislation that is currently into force and that starts to apply before the fixed end date of the TANGO project, i.e. before 1 September 2025. Therefore, EU legislative proposals are excluded from the scope of this Annex.

2. THE GENERAL DATA PROTECTION REGULATION (GDPR)

2.1. Scope of application and applicability to TANGO

Articles 2 and 3 of the GDPR outline, respectively, the material and territorial scope of application of the GDPR.

As concerns its **material scope of application**, the GDPR applies when two conditions are 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[67]. Under Article 4(2) of the GDPR, processing means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction. Therefore, the GDPR applies to the collection, storage and transmission of personal data. If personal data is processed for the purposes of sharing and storing it by using the WP3 technologies, this processing operation would clearly fall in the material scope of the GDPR as it is carried out wholly by automated means.

[67] Article 2(1) of the GDPR reads as follows:
''This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system.''

The **territorial scope of application** of the GDPR is described by its Article 3[68]. The GDPR is characterized by its broad extraterritorial reach, beyond EU borders, which calls for a careful analysis of processing operations with a link to the EU. There are three cases where the GDPR applies. First, it applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union. 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. Second, it applies even where the controller or the processor are not established in the Union, if there is processing of personal data relating to data subjects who are in the Union, and the two following conditions are met: (i) the processing activities relate to the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or (ii) the processing activities relates to the monitoring of behaviour as far as this behaviour takes place within the Union. Third, it applies to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law.
The territorial applicability of the GDPR to processing taking place through the WP3 technologies would thus depend on where and how these technologies are to be deployed. In any case, when personal data is processed in the context of a data-sharing ecosystem, there is a high chance of applicability of the GDPR even where the controller and the processor are not established in the EU.
Given that WP3 technologies are designed to be used for the sharing of data that might be personal, and that might take place in the EU, it is assumed that the intended use of TANGO technologies falls under the scope of application of the GDPR. In cases where a dataset containing both personal and non-personal data is shared, and the non-personal and the personal data parts are 'inextricably linked', the data protection rights and obligations stemming from the **GDPR fully apply to the whole mixed dataset**, also when personal data represent only a small part of the dataset[69].

2.2. Definition of personal data

As outlined above, the material scope of the GDPR is limited to the processing of personal data. Therefore, the identification of personal data is a fundamental step to be carried out to assess the applicability of the GDPR for a given processing operation. The identification of personal data can be a difficult exercise in practice, due to the uncertainties that can arise for the distinction between personal and non-personal data. The different categories of data, and the challenges pertaining to the distinction between personal and non-personal data, are addressed in the final Chapter of this document. This Section only aims to set out the main features of the definition of personal data under the GDPR.

Article 4(1) GDPR defines **personal data** as follows: **''***any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*".
The definition of the GDPR must be complemented with the guidance provided by Wp29 in its Opinion 04/2007 on the concept of personal data[70]. The Wp29 provided a breakdown and description for the four elements that compose the definition of personal data: 'any information'; 'relating to'; 'an identified or identifiable'; 'natural person'. The blocks are described as follows:

[68] Article 3 of the GDPR reads as follows:
''1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not.
2. This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:
(a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or
(b) the monitoring of their behaviour as far as their behaviour takes place within the Union.
3. This Regulation applies to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law.''
[69] European Commission, ''Communication from the Commission to the European Parliament and the Council - Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union'', COM/2019/250 final, 2019, p. 9.
[70] Article 29 Working Party, Opinion 04/2007 on the concept of personal data, WP 136, 2007.

- **''Any information''** is a concept that must be considered having regard to the nature, the content and the format of the information.
  Regarding the nature of the information, the concept of personal data includes any sort of statements about a person. It covers "objective" information, such as the presence of a certain substance in one's blood. It also includes "subjective" information, opinions or assessments.
  Regarding the content of the information, the concept of personal data includes data providing any sort of information. It covers sensitive data but also information touching the individual's private and family life *stricto sensu*, but also information regarding whatever types of activity is undertaken by the individual.
  Regarding the format of the information, personal data can take any form, be it alphabetical or numerical data, as well as information stored on videos and pictures';
- As the data must related to data subjects, it is important to precisely find out which are the relations and/or links that matter and how to distinguish them. Personal data is information that is, by reason of its content, purpose of effect, is **linked to a particular person**:
  - **Content**: Information "relates" to a person when it is "about" that person;
  - **Purpose**: the data are used or are likely to be used, taking into account all the circumstances surrounding the precise case, with the purpose to evaluate, treat in a certain way or influence the status or behaviour of an individual;
  - **Result (effect)**: data can be considered to "relate" to an individual because their use is likely to have an impact on a certain person's rights and interests, taking into account all the circumstances surrounding the precise case;
- A natural person must be **identified or identifiable**. A natural person can be considered as "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group. A natural person is, on the other hand, "identifiable" when, although the person has not been identified yet, it is possible to do it.
- The data subject must be a **natural person**, and thus the GDPR does not apply to legal persons or the deceased.

2.3. Data protection principles

2.3.1. Introduction

Article 5 of the GDPR sets out the principles governing the processing of personal data. These are:

- Data processing must be lawful, fair and transparent (lawfulness, fairness and transparency);
- Data must be collected for specified, explicit and legitimate purposes and not further processed for purposes other than specified (purpose limitation);
- Data must be adequate, relevant and limited to what is necessary in relation to the specified purposes for processing (data minimization);
- Personal data must be "accurate and, where necessary, kept up to date" (accuracy);
- Personal data must be stored only as long as it is necessary for the purpose of data processing (storage limitation);
- The security of personal data must be ensured "against unauthorised or unlawful processing and against accidental loss, destruction or damage" (integrity and confidentiality).
- The controller shall be responsible for, and be able to demonstrate compliance with, all the data processing principles.

2.3.2. Lawfulness, fairness and transparency

According to Articles 5(1)(a) and 6 of the GDPR, the processing of personal data is **lawful** only if there is a **legal ground** that allows such processing. Article 6 of the GDPR exhaustively lists the legal grounds that can be relied on for the processing of personal data, and data controllers must be able to demonstrate that any processing of personal data takes place in accordance with one of these grounds. These grounds are: i) **consent** from the data subject, ii) processing is necessary for the performance of a **contract** to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract, iii) processing is necessary for compliance with a **legal obligation** to which the controller is subject, iv) processing is necessary in order to protect the **vital interests of the data subject** or of another natural person, v) 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, iv)

processing is necessary for the purposes of the **legitimate interests** pursued by the controller or by a third party, subject to some exceptions.
Stricter conditions apply, however, for the processing of **special categories of personal data** (so-called sensitive data) in accordance with Article 9 of the GDPR[71].
Article 5(1)(a) of the GDPR requires that personal data be processed lawfully and **fairly**. This principle, often examined in conjunction with transparency, requires that personal data be **processed in a manner that would be expected by data subjects**. For instance, manipulative practices that aim at 'tricking' the data subject into providing consent for the processing of their personal data, e.g. by employing so-called 'dark patterns', would be contrary to the principle of fairness.
Article 5(1)(a) of the GDPR requires that personal data be processed in a **transparent** manner. Transparency is an important building block of European data protection law. It grounds data controllers' transparency duties under Articles 13 and 14 GDPR, and is a necessary precondition for data subjects to exercise their rights under the GDPR.
The EDPB has provided guidance on how to comply with lawfulness, fairness and transparency by design and by default[72].

2.3.3. Purpose limitation

The purpose limitation principle of Article 5(1)(b) 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** and **compatible use**. Purpose specification requires that personal data is processed only for specified, explicit and legitimate purposes, whereas compatible use prevents to further process personal data in a manner that is incompatible with the original purpose(s) for which it was collected.

The EDPB has provided guidance on how to comply with purpose limitation by design and by default[73].

2.3.4. Data minimisation and storage limitation

Data minimisation and storage limitation are principles that can described jointly, as they both impose a *limitation on the nature, extent and duration of personal data processing based on the purposes for which it was collected*.
The principle of data minimization 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.
The principle of storage limitation of Article 5(1)(e) 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. Also in this case, a **necessity test** is required in relation to the duration of the processing operation.
The EDPB has provided guidance on key design and default elements that enable compliance with data minimisation and storage limitation[74].

2.3.5. 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 EDPB has provided guidance on how to comply with the accuracy principle by design and by default[75].

---

[71] According to Article 9 of the GDPR, the special categories of personal data are the following: ''*personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation*''.
[72] EDPB, ''Guidelines 4/2019 on Article 25 Data Protection by Design and by Default'', 2019, pp. 15-19.
[73] EDPB, ''Guidelines 4/2019 on Article 25 Data Protection by Design and by Default'', 2019, pp. 19-20.
[74] EDPB, ''Guidelines 4/2019 on Article 25 Data Protection by Design and by Default'', 2019, pp. 21-23 and 25-26.
[75] EDPB, ''Guidelines 4/2019 on Article 25 Data Protection by Design and by Default'', 2019, pp. 23-25.

2.3.6. Integrity and confidentiality

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, including against unauthorised or unlawful processing and against accidental loss, destruction or damage*.
This data processing principle is connected with Article 32, which lays down **security** requirements for the processing of personal data and therefore better clarifies how compliance with integrity and confidentiality can be ensured in practice. 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[76].
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*’’. The EDPB has provided guidance on how to comply with the security requirements in its Guidelines on Data Protection by Design and by Default[77].

2.3.7. Data protection principles: accountability

According to Article 5(2) of the GDPR, the *controller shall be responsible for, and be able to demonstrate compliance with, the data processing principles of Article 5(1) of the GDPR*. Article 5(2) introduces the accountability principle, which is a cornerstone principle of the GDPR that vests in the **controller** the **accountability for all the processing operations under its control**. This entails that, even when the processing is in practice carried out by a processor, the controller remains accountable for such processing insofar as it takes place under its instructions.
Moreover, another corollary of the accountability principle lies in the fact that the controller must be **able to demonstrate compliance** with the GDPR in relation to its processing operations, and is thus obliged to keep record of the evidence needed to this end. The **record-keeping obligation** is formalised in Article 30 of the GDPR, according to which each controller shall maintain a record of processing activities under its responsibility.
The EDPB has provided guidance on how to comply with the accountability principle by design and by default. [78]

2.4. Exercise of data subjects’ rights

Pursuant to Article 12(2) of the GDPR, the controller must facilitate the exercise by data subjects of their **rights of access, to rectification, to erasure, to restriction of processing, to data portability, to object and not to be**

[76] B. Van Alsenoy, ‘’Liability under EU Data Protection Law From Directive 95/46 to the General Data Protection Regulation’’, JIPITEC 271, p. 284.
[77] EDPB, ‘’Guidelines 4/2019 on Article 25 Data Protection by Design and by Default’’, 2019, pp. 26-28.
[78] EDPB, ‘’Guidelines 4/2019 on Article 25 Data Protection by Design and by Default’’, 2019, p. 28.

subject to automated individual decision-making[79]. 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 the controller is obliged to make sure that the technologies employed for the processing of personal data 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.

The GDPR confers the following rights on data subjects:

- **Right of access**: the right of access of Article 15 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[80];
- **Right to rectification**: the right to rectification of Article 16 confers data subjects the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her[81];
- **Right to erasure**: according to Article 17, 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[82];

---

[79] Article 12(2) of the GDPR reads as follows:
"The controller shall facilitate the exercise of data subject rights under Articles 15 to 22. In the cases referred to in Article 11(2), the controller shall not refuse to act on the request of the data subject for exercising his or her rights under Articles 15 to 22, unless the controller demonstrates that it is not in a position to identify the data subject".

[80] Article 15(1) of the GDPR reads as follows:
"The data subject shall have the 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 the following information:
(a) the purposes of the processing;
(b) the categories of personal data concerned;
(c) the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;
(d) where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period;
(e) the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing;
(f) the right to lodge a complaint with a supervisory authority;
(g) where the personal data are not collected from the data subject, any available information as to their source;
(h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."

[81] Article 16 of the GDPR reads as follows:
"The data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her. Taking into account the purposes of the processing, the data subject shall have the right to have incomplete personal data completed, including by means of providing a supplementary statement."

[82] Article 17(1) of the GDPR reads as follows:
"The data subject shall have 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 following grounds applies:
(a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;
(b) the data subject withdraws consent on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is no other legal ground for the processing;

- **Right to restriction of processing and to object**: Articles 18[83] and 21[84] provide, respectively, for the right of the data subject to demand the restriction of processing and to object to the processing at the conditions specified in the Articles;
- **Right to data portability**: According to Article 20, the data subject has the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, at the conditions specified in the Article[85];
- **Right not to be subject to automated individual decision-making:** According to Article 22, the data subject has the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her[86].

2.5. Data protection by design and by default

Article 25 of the GDPR requires the controller to implement data protection principles by design and by default for its processing operations. This is a central principle of the GDPR for the design of data processing technologies and procedures, as it mandates the implementation of data protection principles by design and by default in such technologies and procedures.

(c) the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2);
(d) the personal data have been unlawfully processed;
(e) the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject;
(f) the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).''

[83] Article 18(1) of the GDPR reads as follows:
''The data subject shall have the right to obtain from the controller restriction of processing where one of the following applies:
(a) the accuracy of the personal data is contested by the data subject, for a period enabling the controller to verify the accuracy of the personal data;
(b) the processing is unlawful and the data subject opposes the erasure of the personal data and requests the restriction of their use instead;
(c) the controller no longer needs the personal data for the purposes of the processing, but they are required by the data subject for the establishment, exercise or defence of legal claims;
(d) the data subject has objected to processing pursuant to Article 21(1) pending the verification whether the legitimate grounds of the controller override those of the data subject.

[84] Article 21(1) of the GDPR reads as follows:
''The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on point (e) or (f) of Article 6(1), including profiling based on those provisions. The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims.''

[85] Article 20(1) of the GDPR reads as follows:
''The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where:
(a) the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a contract pursuant to point (b) of Article 6(1); and
(b) the processing is carried out by automated means.''

[86] Article 22(1) of the GDPR reads as follows:
''The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.''

Data protection by design requires controller to adopt technical and organisational measures designed in a way that implements data protection principles in an effective manner and that integrates the necessary safeguards into the processing in order to meet the requirements of the GDPR and protect the rights of data subjects[87]. Data protection by design must be implemented by controllers having regard to the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing. In essence, data protection by design requires controllers to embed in their processing technologies and procedures the measures that, in light of the circumstances of the processing, are most appropriate and proportionate to implement data protection principles.
Data protection by default requires controllers to put in place technical and organisational measures which ensure that, by default, only the personal data necessary for each specific purpose of the processing are processed[88]. In substance, data protection by default requires controllers to ensure that the technologies and procedures used for data processing implement data protection principles by default, without the need for a specific ''opt-in'' action.

3. THE ePRIVACY DIRECTIVE

3.1. Scope of application and applicability to TANGO

The ePrivacy Directive is an important legal act in EU law, as it complements and particularises the data protection requirements laid down by the GDPR (and previously Data Protection Directive 1995/46/EC) for the **electronic communications sector**.

Contrary to the GDPR, the ePrivacy Directive does not have a general scope of application to all the processing activities involving personal data. In particular, the ePrivacy Directive applies only to the processing of personal data in the electronic communication sector[89], i.e. when personal data is processed by providers of publicly available electronic communications services in public communications networks in the EU[90]. It states explicitly that it aims to "particularise and complement" the provisions of the GDPR, concerning the processing of personal data in the electronic communication sector[91].

[87] Article 25(1) of the GDPR reads as follows:
''*Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.*''
[88] Article 25(2) of the GDPR reads as follows:
'' The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons''.
[89] Article 1(1) of the Directive reads as follows:
''This Directive harmonises the provisions of the Member States required to ensure an equivalent level of protection of fundamental rights and freedoms, and in particular the right to privacy, with respect to the processing of personal data in the electronic communication sector and to ensure the free movement of such data and of electronic communication equipment and services in the Community''.
[90] Article 3(1) of the Directive reads as follows:
''This Directive shall apply to the processing of personal data in connection with the provision of publicly available electronic communications services in public communications networks in the Community''.
[91] Article 1(1) and (2) of the e-Privacy Directive, to be read in light of article 94(2) GDPR.

In order to determine whether the ePrivacy Directive is applicable to the data-sharing activities carried out with TANGO technologies, there are two aspects to consider:

a) Are any of the services provided through the TANGO technologies, or the combination thereof in the TANGO reference architecture, machine-to-machine (hereinafter, ''M2M'') services within the meaning of Article 2(4)(c)[92] of Directive 2018/1972[93] (known as the ''European Electronic Communications Code'' or ''EECC'')? This will be the case when they consist wholly or mainly in the conveyance of signals such as transmission services used for the provision of machine-to-machine services and for broadcasting.

b) Are any of the services provided through the TANGO technologies, or the combination thereof in the TANGO reference architecture, interpersonal communications services within the meaning of Article 2(4)(b) of the EECC?[94]

If the answer to questions a) and b) is in the affirmative, a third question arises:

c) Do any of these services qualify as 'public' within the meaning of the EECC?

In the TANGO reference architecture, the exchange of data is enabled through the ''TANGO connector''. The TANGO connector hosts containerised services and each connector represents a participant to the TANGO ecosystem. Given the central role of connectors for data sharing in the ecosystem, the two questions posed above will be answered from the perspective of the TANGO connectors, looking at whether the interaction between connectors enabled in the TANGO ecosystem constitute the provision of an electronic communications service.

As concerns the **first question** on the qualification as M2M service, it must be noted at the outset that the EECC does not define in detail the category of M2M services. **M2M services are part of the third sub-category ("services consisting wholly or mainly in the conveyance of signals") of the wider category of "electronic communications services"** defined by Article 2(4) of the EECC, but they do not have their own specific definition in the EECC. Given that the TANGO connectors enable communication between devices, it must be examined whether their operation leads to the provision of a M2M service. To this end, three clarifications must be made on the definition of M2M services.

First, as argued by A. de Streel and C. Hocepied[95], only the transmission element of M2M communications should be considered to fall within the scope of the EECC, with the consequent exclusion from its scope of M2M services at the application layer. This interpretation would be best in line with the legal description provided by the EECC to the category that M2M services belong to, i.e. services consisting in the conveyance of signals. In particular,

---

[92] Article 2(4)(c) of the EECC reads as follows:
'' 'electronic communications service' means a service normally provided for remuneration via electronic communications networks, which encompasses, with the exception of services providing, or exercising editorial control over, content transmitted using electronic communications networks and services, the following types of services:
(…)
(c) services consisting wholly or mainly in the conveyance of signals such as transmission services used for the provision of machine-to-machine services and for broadcasting''.

[93] 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 of 17.12.2018.

[94] Article 2(4)(b) of the EECC reads as follows:
''' electronic communications service' means a service normally provided for remuneration via electronic communications networks, which encompasses, with the exception of services providing, or exercising editorial control over, content transmitted using electronic communications networks and services, the following types of services:
(…)
(b) Interpersonal communications service ; and
(…)''

[95] A. de Streel and C. Hocepied, ''The Regulation of electronic communications networks and services'', book chapter in Garzaniti et al, *Electronic Communications, Audiovisual Services and the Internet – EU Competition Law & Regulation*, 4th edition, Sweet and Maxwell, 2019, pp. 30-31.

services that act at the application layer are not directly responsible for the conveyance of signals, which takes place at the transmission level.

Second, it is of no relevance for the finding of an electronic communications service whether the transmission of signals is by means of an infrastructure that does, or does not, belong to the respective service provider. It is only relevant to establish whether the service provider is responsible vis-à-vis the end-users for transmission of the signal[96]. Therefore, it may be the case that a provider of M2M services at the application layer who also resells the connectivity of another provider qualifies as an electronic communications service provider, depending on whether the overall service consists ''wholly or mainly'' in the conveyance of signals. In such cases, the assessment would thus come down to the importance of the transmission element within the overall service. In carrying out this assessment, there should be a weighting of the respective value of the elements of the service that are conveyance and that are not, taking into account technical and functional characteristics, and the perspective of the end-user[97].

Third, it is important to bear in mind that BEREC concluded, in relation to Internet of Things (''IoT'') products, that an IoT user (e.g. car manufacturer, provider of energy including smart meter) who includes connectivity as an input product into his products or services does not seem to provide an electronic communications service when selling a connected device or "smart" service, whereas there may be an electronic communications service where the IoT user is contractually liable vis-à-vis the end-user for the provision of connectivity, and this constitutes a whole or main part of what is sold[98].

In light of the observations made above, it can be concluded that, where a service consists only of enabling data sharing through the TANGO connectors, and the transmission or connectivity element (e.g. internet access) is provided by another entity who is contractually liable towards the end-user, then the service as a whole would not qualify as an electronic communications service. In that case you would have two separate services, and only the service enabling the transmission or connectivity may qualify as an electronic communications service. By contrast, where the **connectivity element is integrated in the service provided through the TANGO connectors, and the same provider of this service is also contractually liable towards the end-user, then it must be examined whether the connectivity element is a main portion of the overall service**. This is unlikely to be the case for the TANGO connectors, where the main service should be the functionalities of the connectors themselves rather than the connectivity which is merely ancillary to enable the transmission of data. Therefore, it cannot be excluded that the operation of the TANGO connectors could, in specific circumstances, qualify as an electronic communications service.

The **second question** on the qualification as an interpersonal communications service should be answered on the basis of the legal definition provided for this service in the EECC. An interpersonal communications service is a service ''*normally provided for remuneration that enables direct interpersonal and interactive exchange of information via electronic communications networks between a finite number of persons, whereby the persons initiating or participating in the communication determine its recipient(s) and does not include services which enable interpersonal and interactive communication merely as a minor ancillary feature that is intrinsically linked to another service*''[99]. It is essential to note that the service consists of enabling an interactive exchange of information, which could in theory be carried out through the TANGO connectors. However, such communication must take place between a finite number of natural persons. While the TANGO connectors can be used for exchanging data with a finite and predetermined number of persons, it might not necessarily be used solely for exchanges between natural persons, but also for applications where the participants behind the connectors are simple devices or organisations. Where there is not a natural person participating in the exchange, the service cannot qualify as an interpersonal communications service. Therefore, the operation of a TANGO connector could be an interpersonal communications service, depending on the use case to which it is destined, and on the condition that the communications aspect is not a minor ancillary feature that is intrinsically linked to another service. As illustrated by Recital 17 of the EECC, this would be the case where ''*its objective utility for an end-user is very*

[96] CJEU 5 June 2019, C-142/18, Skype Communications Sarl, ECLI:EU:C:2019:460, para. 33; CJEU 13 June 2019, C-193/18, Google LLC, para. 38; CJEU 30 April 2014, C-475/12, UPC DTH Sarl, ECLI:EU:C:2014:285, paras. 43 - 44.
[97] A. de Streel and C. Hocepied, ''The Regulation of electronic communications networks and services'', book chapter in Garzaniti et al, ''Electronic Communications, Audiovisual Services and the Internet – EU Competition Law & Regulation'', 4th edition, Sweet and Maxwell, 2019, p. 31.
[98] BEREC, ''Report of 12 February 2016 on enabling the Internet of Things'', BoR(16)39, p. 22.
[99] Article 2(5) of the EECC.

limited and where it is in reality barely used by end-users''. As an example of this situation, Recital 17 of the EECC refers to a communication channel used in online games[100].

In light of the above, the answer to the first and second question must be that **the operation of the TANGO connector might qualify as an electronic communications service in certain circumstances, depending on how the service is provided and for which use cases**.

As to the **third question** on the qualification of the service as ''public'', it must be noted that the EECC does not provide clarification of the circumstances under which a service is considered to be publicly available. However, according to the case law, a service must be considered publicly available when **any part of the public may choose to make use of the service offered**[101]. 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 answer to the third question must thus also be that it would depend on how the service is provided and for which use cases. **In general, it cannot be excluded that it could consist of the provision of a publicly available electronic communications service.**

In light of the answers to the three questions provided above, it must be **concluded that the TANGO technologies could be used for the provision of an electronic communications service in certain circumstances. Therefore, the applicability of the ePrivacy Directive must be assumed**.

3.2. Relevant provisions

The ePrivacy Directive contains high-level provisions on several aspects, from security and confidentiality of communications, to the processing of location data and other traffic data, to the storage of information in the terminal equipment of a subscriber or user and unsolicited communications. In the context of this report, only certain provisions of the ePrivacy Directive are to be considered as relevant.

First, the **security obligation** in Article 4 of the Directive is to be considered, due to its relevance for technologies that enable data sharing and storage. Article 4 of the ePrivacy Directive requires providers of publicly available electronic communications services to take appropriate technical and organisational measures to safeguard the security of their services, if necessary in conjunction with the provider of the public communications network with respect to network security[102]. In prescribing the implementation of security measures, Article 4 sets out a list of minimum measures that should be adopted in any case. [103]

Second, Articles 6 and 9 of the ePrivacy Directive are relevant because they lay down conditions for the processing of certain categories of personal data. According to Article 6, **traffic data** relating to subscribers and users processed and stored by the provider of a public communications network or publicly available electronic

[100] Recital 17 of the EECC, last sentence, reads as follows: ''*an example of a feature that could be considered to fall outside the scope of the definition of interpersonal communications services might be, in principle, a communication channel in online games, depending on the features of the communication facility of the service*''.

[101] EFTA 18 May 2016, E-6/16, Fjarskipti and Icelandic Post and Telecom Administration, para. 56.

[102] Article 4(1) of the ePrivacy Directive reads as follows:

''The provider of a publicly available electronic communications service must take appropriate technical and organisational measures to safeguard security of its services, if necessary in conjunction with the provider of the public communications network with respect to network security. Having regard to the state of the art and the cost of their implementation, these measures shall ensure a level of security appropriate to the risk presented.''

[103] Article 4(1a) of the ePrivacy Directive reads as follows:

''*Without prejudice to Directive 95/46/EC, the measures referred to in paragraph 1 shall at least:*

- *Ensure that personal data can be accessed only by authorised personnel for legally authorised purposes,*
- *Protect personal data stored or transmitted against accidental or unlawful destruction, accidental loss or alteration, and unauthorised or unlawful storage, processing, access or disclosure, and,*
- *Ensure the implementation of a security policy with respect to the processing of personal data,*

Relevant national authorities shall be able to audit the measures taken by providers of publicly available electronic communication services and to issue recommendations about best practices concerning the level of security which those measures should achieve.''

communications service must be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication[104]. In relation to **location data** other than traffic data, Article 9 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[105].

3.3. Interplay between the ePrivacy Directive and the GDPR

The GDPR and the ePrivacy Directive are legislative texts with an evident overlap in scope of application. On the one hand, the GDPR lays down rules for the protection of data subjects' fundamental rights in relation to the processing of their personal data, with a broad scope of application that covers almost every processing operation of personal data. On the other hand, the ePrivacy Directive particularises personal data protection rules for the specific context of the processing of personal data in publicly available electronic communications networks.

In principle, both the GDPR and the ePrivacy Directive apply to the processing of personal data in publicly available electronic networks. This leads to an overlap in the material scope of the two legislations, as explicitly recognised by the EDPB[106]. 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 derogat legi generali* (which essentially means that more specific rules will prevail over more general rules).

For the purposes of this report, it must be noted that processing operations in publicly available electronic networks are subject to the security requirements of both the GDPR and the ePrivacy Directive, and that these requirements substantially coincide under the two legislative texts. Article 32 of the GDPR and Article 4 of the ePrivacy Directive formulate the security obligation in a similar manner, as they both require to take ''appropriate technical and organizational measures'' that must be determined based on the state of the art and the cost of their implementation, and to ensure a level of security appropriate to the risks presented. There is a difference lying in the fact that Article 4 of the ePrivacy Directive specifies a minimum standard of security that must in any case be ensured, by listing a series of measures that shall always be adopted. However, this difference likely does not exist in practice, as the level of protection reflected in these measures is very basic and should certainly be respected by controllers and processors under the GDPR as well, irrespective of the circumstances of the case.

Overall, it can said that **the GDPR and the ePrivacy Directive prescribe security requirements that are either equivalent or coincide to some extent**.

[104] Article 6(1) of the ePrivacy Directive reads as follows:
'' 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 without prejudice to paragraphs 2, 3 and 5 of this Article and Article 15(1).''

[105] Article 9(1) of the ePrivacy Directive reads as follows:
''Where location data other than traffic data, relating to users or subscribers of public communications networks or publicly available electronic communications services, can be processed, 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. The service provider must inform the users or subscribers, prior to obtaining their consent, of the type of location data other than traffic data which will be processed, of the purposes and duration of the processing and whether the data will be transmitted to a third party for the purpose of providing the value added service. Users or subscribers shall be given the possibility to withdraw their consent for the processing of location data other than traffic data at any time.''

[106] EDPB, Opinion 5/2019 on the interplay between the ePrivacy Directive and the GDPR, in particular regarding the competence, tasks and powers of data protection authorities, 2019, pp. 13-15.

4. THE NIS2 DIRECTIVE

4.1. Scope of application and applicability to TANGO

The NIS2 Directive is a legal act currently in force that will start to apply from 18 October 2024. Member States are obliged to transpose the provisions of the Directive by 17 October 2024 and start to apply them by 18 October 2024[107], with the consequence that the (first) NIS Directive[108] is repealed with effect from 18 October 2024. As October 2024 precedes the date of end of TANGO, the provisions of the NIS2 Directive are taken into account as relevant for the purposes of this document.
The NIS2 Directive is a sectorial piece of legislation that applies to specific categories of entities identified in the Annexes I and II of the Directive that qualify as medium-sized enterprises[109], and which provide their services or carry out their activities within the Union[110]. The Directive also applies to the entities of its Annexes I and II that, regardless of their size, are in the situations listed in Article 2(2) of the Directive.
The entities that fall under the scope of application of the Directive are in turn divided into two categories: **essential and important entities**[111]. The Directive lays down partially differentiated obligations depending on whether an entity falls into one of the other category.
Providers of public electronic communications networks and publicly available electronic communications services are listed in Annex I, as well as in the categories of Article 2(2) of the Directive. Therefore, they can fall within the scope of application of the Directive either if they qualify as small and medium enterprises, or if they meet the conditions in Article 2(2) of the Directive. **Since WP3 technologies may be used to provide publicly available electronic communications services or networks, their operation may be subject to the provisions of the NIS2 Directive.**

---

[107] Article 41(1) of the NIS2 Directive reads as follows:
"By 17 October 2024, Member States shall adopt and publish the measures necessary to comply with this Directive. They shall immediately inform the Commission thereof.
They shall apply those measures from 18 October 2024."

[108] Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union, OJ L 194/1 of 19.07.2016.

[109] The enterprises must qualify as medium-sized enterprises within the meaning of Article 2 of the Annex to Recommendation 2003/361/EC.

[110] Article 2(1) of the NIS2 Directive reads as follows:
" This Directive applies to public or private entities of a type referred to in Annex I or II which qualify as medium-sized enterprises under Article 2 of the Annex to Recommendation 2003/361/EC, or exceed the ceilings for medium-sized enterprises provided for in paragraph 1 of that Article, and which provide their services or carry out their activities within the Union.
Article 3(4) of the Annex to that Recommendation shall not apply for the purposes of this Directive."

[111] Article 3(1) and (2) of the Directive reads as follows:
"1. For the purposes of this Directive, the following entities shall be considered to be essential entities:
(a) entities of a type referred to in Annex I which exceed the ceilings for medium-sized enterprises provided for in Article 2(1) of the Annex to Recommendation 2003/361/EC;
(b) qualified trust service providers and top-level domain name registries as well as DNS service providers, regardless of their size;
(c) providers of public electronic communications networks or of publicly available electronic communications services which qualify as medium-sized enterprises under Article 2 of the Annex to Recommendation 2003/361/EC;
(d) public administration entities referred to in Article 2(2), point (f)(i);
(e) any other entities of a type referred to in Annex I or II that are identified by a Member State as essential entities pursuant to Article 2(2), points (b) to (e);
(f) entities identified as critical entities under Directive (EU) 2022/2557, referred to in Article 2(3) of this Directive;
(g) if the Member State so provides, entities which that Member State identified before 16 January 2023 as operators of essential services in accordance with Directive (EU) 2016/1148 or national law.
2. For the purposes of this Directive, entities of a type referred to in Annex I or II which do not qualify as essential entities pursuant to paragraph 1 of this Article shall be considered to be important entities. This includes entities identified by Member States as important entities pursuant to Article 2(2), points (b) to (e)."

4.2. Relevant provisions

The NIS2 Directive lays down obligations that act on various fronts in order to achieve a high common level of cybersecurity across the Union. Amongst these, there are cybersecurity risk-management obligations for the entities that fall under the scope of application of the Directive. For the purposes of this report, it is important to consider the obligations in Article 21 regarding **cybersecurity risk-management measures**. Article 21(1) requires obliged entities ‘*’to take appropriate and proportionate technical, operational and organisational measures to manage the risks posed to the security of network and information systems*’’, as well as ‘*’to prevent or minimise the impact of incidents on recipients of their services and on other services*’’[112]. Article 21(1) provides guidance on how these measures should be determined, clarifying that they: i) must ensure a level of security **appropriate to the risks posed**, ii) must be defined taking into account the **state of the art and their costs** of implementation, iii) be **proportionate** based on the circumstances of the case. The proportionality assessment shall take account of ‘’*the degree of the entity’s exposure to risks, the entity’s size and the likelihood of occurrence of incidents and their severity, including their societal and economic impact’’.* Therefore, the appropriate measures to put in place must be determined on a **case-by-case basis**, in light of all the factors outlined above.

Article 21(2) also provides a list of cybersecurity measures that must ‘’at least’’ be implemented[113], as a **minimum standard of cybersecurity** that shall be respected in any circumstance. This list in not meant to be exhaustive and the adoption of these measures does not automatically lead to a presumption of compliance with the obligations in Article 21, even though it is an essential starting point therefor.

Article 21 must be taken into account as the only provision that may influence the design of technologies to be used for the sharing of data by obliged entities. Alongside organisational and operational measures, obliged entities may need to adopt technical measures to manage risks, to be implemented by means of specific technological solutions. For instance, Article 21(2) mentions the use of cryptography, and where appropriate encryption, as one of the measures that must in any case be put in place. Additionally, Article 21(2) requires the implementation of

[112] Article 21(1) reads as follows:
’’1. Member States shall ensure that essential and important entities take appropriate and proportionate technical, operational and organisational measures to manage the risks posed to the security of network and information systems which those entities use for their operations or for the provision of their services, and to prevent or minimise the impact of incidents on recipients of their services and on other services.
Taking into account the state-of-the-art and, where applicable, relevant European and international standards, as well as the cost of implementation, the measures referred to in the first subparagraph shall ensure a level of security of network and information systems appropriate to the risks posed. When assessing the proportionality of those measures, due account shall be taken of the degree of the entity’s exposure to risks, the entity’s size and the likelihood of occurrence of incidents and their severity, including their societal and economic impact.’’

[113] Article 21(2) reads as follows:
’’The measures referred to in paragraph 1 shall be based on an all-hazards approach that aims to protect network and information systems and the physical environment of those systems from incidents, and shall include at least the following:
(a) policies on risk analysis and information system security;
(b) incident handling;
(c) business continuity, such as backup management and disaster recovery, and crisis management;
(d) supply chain security, including security-related aspects concerning the relationships between each entity and its direct suppliers or service providers;
(e) security in network and information systems acquisition, development and maintenance, including vulnerability handling and disclosure;
(f) policies and procedures to assess the effectiveness of cybersecurity risk-management measures;
(g) basic cyber hygiene practices and cybersecurity training;
(h) policies and procedures regarding the use of cryptography and, where appropriate, encryption;
(i) human resources security, access control policies and asset management;
(j) the use of multi-factor authentication or continuous authentication solutions, secured voice, video and text communications and secured emergency communication systems within the entity, where appropriate.’’

access control policies and the use of authentication solutions, which presumably influence the design of technologies used by obliged entities.

5. DATA GOVERNANCE ACT

5.1. Scope of application and applicability to TANGO

Regulation (EU) 2022/868 (commonly known as ''Data Governance Act'', hereinafter the ''DGA'') is a composite piece of legislation with provisions dedicated to four different areas: a) conditions for the **re-use**, within the Union, of certain categories of data held by **public sector bodies**; b) a notification and supervisory framework for the provision of **data intermediation services**; c) a framework for voluntary registration of entities which collect and process data made available for **altruistic purposes**; and d) a framework for the establishment of a **European Data Innovation Board**[114].

The provisions on the data intermediation services may be of significant relevance to the WP3 technologies. The DGA introduces a new notion of **''data intermediation services''** (hereinafter, ''DIS'') in Article 2(11) of the DGA. Alongside the general definition, Article 10 of the DGA provides a list of three categories of DIS, namely:

a) Intermediation services between data holders and potential data users, including making available the technical or other means to enable such services; those services may include bilateral or multilateral exchanges of data or the creation of platforms or databases enabling the exchange or joint use of data, as well as the establishment of other specific infrastructure for the interconnection of data holders with data users[115];
b) Intermediation services between data subjects that seek to make their personal data available or natural persons that seek to make non-personal data available, and potential data users, including making available the technical or other means to enable such services, and in particular enabling the exercise of the data subjects' rights provided in Regulation (EU) 2016/679[116];
c) Services of data cooperatives[117] within the meaning given to this notion by the DGA.

A data intermediation service is defined by Article 2(11) of the DGA as ''*a service which aims to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means, including for the purpose of exercising the rights of data subjects in relation to personal data*''. Recital 28 of the DGA provides **examples** of data intermediation services, which include '*'data marketplaces on which undertakings could make data available to others, orchestrators of data sharing ecosystems that are open to all interested parties, for instance in the context of common European data spaces, as well as data pools established jointly by several legal or natural persons with the intention to license the use of such data pools to all interested parties in a manner that all participants that contribute to the data pools would receive a reward for their contribution*''.

It can be noted that Article 2(11) sets out a series of general criteria that must be assessed to determine whether a given activity qualifies as a DIS. The **criteria** are: a) service, b) aim to establish commercial relationships for the purpose of data sharing, c) between an undetermined number of data subjects, data holders and data users, d) through technical, legal or other means[118]. It is not clear how the general definition of DIS relates to the three categories listed in Article 10 of the DGA, and in particular whether the three categories in Article 10 are a subset of DIS that alone is subject to the provisions of Articles 11 and 12[119]. While some authors argue that the list in

[114] Article 1(1) of the DGA.
[115] Article 10(a) of the DGA.
[116] Article 10(b) of the DGA.
[117] Article 10(c) of the DGA.
[118] For a detailed analysis of how these criteria may be interpreted and applied, see: T. Bobev, V. K. Dessers, C. Ducuing, M. Fierens, A. Palumbo, B. Peeters, L. Stähler, ''CiTiP White Paper on the Definition of Data Intermediation Services'', 2023 accessed on 7 November 2023 at https://ssrn.com/abstract=4589987.
[119] Carovano and Finck suggest that the listing of Article 10 invites speculation as to whether it really creates a subset of DIS (compared to the general definition in Article 2(11)) that are alone subject to Articles 11 and 12, see: G. Carovano, M. Finck, ''Regulating data intermediaries: The impact of the Data Governance Act on the EU's data economy'', 50 Computer Science & Law Review 7, 2023.

Article 10 should not be understood as creating a subset of services that only are subject to Articles 11 and 12[120], other authors do not seem to hold the same view[121].

Starting from the assumption that the three categories of Article 10 are in any case subject to Articles 11 and 12 in Chapter III of the DGA, for the purposes of this section it is relevant to assess whether the use of the WP3 technologies could fall under any of these categories. If this is the case, it is not necessary to also conduct an assessment under the general criteria in Article 2(11). In this regard, it must be noted that Article 10(a) refers to services consisting of ''*making available the technical or other means*'' enabling intermediation services, including ''*the creation of platforms or databases enabling the exchange or joint use of data*'', or ''*the establishment of other specific infrastructure for the interconnection of data holders with data users''*. Therefore, it appears that the provision of technical infrastructure that enables data intermediation within the meaning of the DGA falls under the scope of Chapter III, even though it is not clear what level of technical contribution to data intermediation can be considered sufficient to fall under the category in Article 10(a). Many technical components can interact to carry out data intermediation, with some having a substantial contribution and others a more ancillary role. It can be argued that the WP3 technologies, if combined with the other technologies of the entire TANGO architecture, can constitute an infrastructure for the interconnection of data holders with data users and to enable the exchange of data.

In light of the above, since Article 12 might apply to the operation of the WP3 technologies, it is considered as a relevant provision below.

5.2. Relevant provisions

Article 12 of the DGA sets out the conditions that must be complied with for the provision of data intermediation services[122]. Amongst the listed conditions, there are some that can influence the design of the technical infrastructure to be used for the exchange of data. These conditions are described below.

[120] G. Carovano, M. Finck, ''Regulating data intermediaries: The impact of the Data Governance Act on the EU's data economy'', 50 Computer Science & Law Review 7, 2023.

[121] H. Richter, ''Looking at the Data Governance Act and Beyond: How to Better Integrate Data Intermediaries in the Market Order for Data Sharing'', 72 GRUR International 462, 2023; L, von Ditfurth, G. Lienemann, ''The Data Governance Act: – Promoting or Restricting Data Intermediaries?'', Competition and Regulation in Network Industries, 2022.

[122] Article 12 of the DGA reads as follows:
''The provision of data intermediation services referred in Article 10 shall be subject to the following conditions:
(a) the data intermediation services provider shall not use the data for which it provides data intermediation services for purposes other than to put them at the disposal of data users and shall provide data intermediation services through a separate legal person;
(b) the commercial terms, including pricing, for the provision of data intermediation services to a data holder or data user shall not be dependent upon whether the data holder or data user uses other services provided by the same data intermediation services provider or by a related entity, and if so to what degree the data holder or data user uses such other services;
(c) the data collected with respect to any activity of a natural or legal person for the purpose of the provision of the data intermediation service, including the date, time and geolocation data, duration of activity and connections to other natural or legal persons established by the person who uses the data intermediation service, shall be used only for the development of that data intermediation service, which may entail the use of data for the detection of fraud or cybersecurity, and shall be made available to the data holders upon request;
(d) the data intermediation services provider shall facilitate the exchange of the data in the format in which it receives it from a data subject or a data holder, shall convert the data into specific formats only to enhance interoperability within and across sectors or if requested by the data user or where mandated by Union law or to ensure harmonisation with international or European data standards and shall offer an opt-out possibility regarding those conversions to data subjects or data holders, unless the conversion is mandated by Union law;
(e) data intermediation services may include offering additional specific tools and services to data holders or data subjects for the specific purpose of facilitating the exchange of data, such as temporary storage, curation,

First, letter j) of Article 12 states that the provider of data intermediation services shall '*'put in place adequate technical, legal and organisational measures in order to* ***prevent the transfer of or access to non-personal data that is unlawful*** *under Union law or the national law of the relevant Member State*''. This condition has a broad and open-ended scope of application as it requires to prevent all transfers or access to non-personal data that might unlawful, now or in the future, under the applicable legislation. Irrespective of the specific cases of unlawful transfer or access that might become relevant, this condition call for the implementation of, among others, technical safeguards that prevent any unlawful transfer or access. These safeguards could consist, for example, of access control measures ensuring that data flows do not lead to unlawful transfer or access to non-personal data. Therefore, if WP3 technologies are used to provide data intermediation services, they should incorporate any technical safeguards needed to comply with this condition.

Second, letter l) of Article 12 requires data intermediation services providers to take measures that ensure an **appropriate level of security for the storage, processing and transmission of non-personal data**. This letter introduces, in essence, a risk-based security obligation which could be compared to those laid down under the GDPR, the ePrivacy Directive, the NIS2 Directive, and to the security precautions that may needed to comply with the reasonable steps requirement under the TSD. Interestingly, it is also required in letter l) that the data intermediation service provider ensures '*'the highest level of security for the storage and transmission of competitively sensitive information*''. This appears to call for a highest level of security for the non-personal data that is competitively sensitive, including but not limited to non-personal data containing trade secrets. The more pervasive security obligation imposed for this type of information is in line with the reasonable steps requirement of the TSD, which would in any case require that commercially valuable non-personal data is subject to protective measures that would not be put in place for commercially not valuable non-personal data. Therefore, if WP3

---

conversion, anonymisation and pseudonymisation, such tools being used only at the explicit request or approval of the data holder or data subject and third-party tools offered in that context not being used for other purposes;
(f) the data intermediation services provider shall ensure that the procedure for access to its service is fair, transparent and non-discriminatory for both data subjects and data holders, as well as for data users, including with regard to prices and terms of service;
(g) the data intermediation services provider shall have procedures in place to prevent fraudulent or abusive practices in relation to parties seeking access through its data intermediation services;
(h) the data intermediation services provider shall, in the event of its insolvency, ensure a reasonable continuity of the provision of its data intermediation services and, where such data intermediation services ensure the storage of data, shall have mechanisms in place to allow data holders and data users to obtain access to, to transfer or to retrieve their data and, where such data intermediation services are provided between data subjects and data users, to allow data subjects to exercise their rights;
(i) the data intermediation services provider shall take appropriate measures to ensure interoperability with other data intermediation services, inter alia, by means of commonly used open standards in the sector in which the data intermediation services provider operates;
(j) the data intermediation services provider shall put in place adequate technical, legal and organisational measures in order to prevent the transfer of or access to non-personal data that is unlawful under Union law or the national law of the relevant Member State;
(k) the data intermediation services provider shall without delay inform data holders in the event of an unauthorised transfer, access or use of the non-personal data that it has shared;
(l) the data intermediation services provider shall take necessary measures to ensure an appropriate level of security for the storage, processing and transmission of non-personal data, and the data intermediation services provider shall further ensure the highest level of security for the storage and transmission of competitively sensitive information;
(m) the data intermediation services provider offering services to data subjects shall act in the data subjects' best interest where it facilitates the exercise of their rights, in particular by informing and, where appropriate, advising data subjects in a concise, transparent, intelligible and easily accessible manner about intended data uses by data users and standard terms and conditions attached to such uses before data subjects give consent;
(n) where a data intermediation services provider provides tools for obtaining consent from data subjects or permissions to process data made available by data holders, it shall, where relevant, specify the third-country jurisdiction in which the data use is intended to take place and provide data subjects with tools to both give and withdraw consent and data holders with tools to both give and withdraw permissions to process data;
(o) the data intermediation services provider shall maintain a log record of the data intermediation activity.''

technologies are used to provide data intermediation services, they should incorporate any technical safeguards that deliver the requisite level of security.

Finally, letter o) of Article 12 sets the condition that providers of data intermediation services shall maintain a log record of the data intermediation activity. This obligation may require WP3 technologies to enable the record-keeping of certain activities, such as records of data sharing transactions.

6. TRADE SECRETS DIRECTIVE

6.1. Scope of application

The Trade Secrets Directive introduced an harmonization of national substantive trade secret law, in relation to central aspects such as the definition of trade secret, the conditions to qualify the acquisition, use and disclosure of trade secrets as lawful or unlawful, and the enforcement of trade secrets protection. As such, the Directive has harmonized trade secrets protection in terms of both substantive and procedural law. The TSD lays down minimum standards of harmonization that the Member States were obliged to transpose in national law.
For the purposes of this report, the most important provision of the TSD is its Article 2(1), which provides a definition of trade secret by setting out three requirements that must be met by a piece of information to qualify as a trade secret[123]. In particular, **a trade secret is information that meets all of the following requirements**:

(a) it is secret in the sense that it is **not**, as a body or in the precise configuration and assembly of its components, **generally known among or readily accessible to persons** within the circles that normally deal with the kind of information in question;
(b) it has **commercial value** because it is secret;
(c) it has been subject to **reasonable steps** under the circumstances, by the person lawfully in control of the information, to keep it secret.

As Article 2 refers to '**information**', it can be argued that trade secrets **protect the semantic meaning of data**, i.e. the meaning understandable in natural language of the information encoded in the data, and it does **not** protect data on a **syntactic level**, which concerns the bits and bytes that compose the data[124]. As a consequence, every type of information can, in principle, be protected as trade secrets, with no limitation that applies *prima facie*. Based on this definition, it can be noted that many types of data can be trade secrets, including subject matter that could be protected by intellectual property rights such as a patent or copyright. Recital 14 of the TSD offers a glimpse of the large scope of the definition, that includes '*know-how, business information and technological information where there is both a legitimate interest in keeping them confidential and a legitimate expectation that such confidentiality will be preserved*'. Therefore, **much of the data to be shared in the data economy may fairly easily qualify as containing trade secrets, if they have commercial value**. There are many types of data, including non-personal data, that can have commercial value, as evidenced by the fact that there are well-developed markets for non-personal data[125]. Where a market already exists for a specific type of data, it could be straightforward to prove the existence of commercial value and of trade secrets, but potential commercial value could also be proved for markets that do not yet exist.

[123] Article 2(1) of the TSD reads as follows:
''trade secret' means information which meets all of the following requirements:
(a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;
(b) it has commercial value because it is secret;
(c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret;''
[124] J. Drexl, ''Data access and control in the era of connected devices'', Report for the European Bureau of Consumers' Union (BEUC), 2018, p. 101.
[125] European Commission, ''Commission staff working document on the free flow of data and emerging issues of the European data economy'', SWD(2017) 2 final of 2017.

6.2. Relevance of trade secrets legislation for the design of WP3 technologies

The TSD was not conceived having the data economy in mind, but with the increasing importance of data and data-sharing, the role of trade secrets for the data economy has gained more attention. **Trade secrets legislation is relevant for any activity entailing the sharing of data that may contain trade secrets.** The relevance of the TSD derives from the fact that trade secret owners may have an interest in ensuring that data containing trade secrets is shared in a way that preserves its secrecy, and thus its protectability under trade secrets legislation, in line with the requirements laid down in Article 2 of the TSD.

When the WP3 technologies are used to share and store data whose content is not pre-defined and may contain trade secrets to be protected, they would need to enable compliance, at least insofar as their role in the handling of the data is concerned, with the reasonable steps requirement of Article 2(1)(c) of the TSD. Due to its importance in the context of data sharing and storage, a description of this requirement, and of its applicative repercussions, is provided in the section below.

6.3. The 'reasonable steps' requirement in the TSD

Of the three requirements listed in Article 2 of the TSD, the most relevant for the discussion around trade secrets protection in data sharing is the ''**reasonable steps**'' that information must be subject to, in order to remain secret. This is the requirement that assumes most importance in the context of data sharing and storage, as it pertains, *inter alia*, to the level of (cyber)security and confidentiality of the data during the transfer. In particular, the data owner who intends to protect trade secrets contained in the data to be shared and stored would need to ensure that the technical cybersecurity measures adopted in practice qualify as reasonable steps under the TSD. Therefore, especially in complex data sharing ecosystems, the protection of trade secrets would be dependent on the implementation of reasonable steps in the form of technical protection measures that enable a certain level of cybersecurity and confidentiality. In this regard, the assessment of whether reasonable steps have been adopted in a given case is highly contextual and dependent on various factors, such as the value of the data, the predictable threats and the available technologies. In addition to the highly contextual nature of this threshold, there is significant legal uncertainty regarding the interpretation of the reasonable steps requirement under the TSD, which does not facilitate the task for the data owner to identify the reasonable steps to take in a given case. The legal uncertainty surrounding the interpretation and application of Article 2 of the TSD derives from the fact that, to date, there have not been any judgements from the CJEU on the interpretation of the TSD[126], and in particular on the meaning of the reasonable steps requirement. However, as is further discussed below, some high-level and relatively context-independent considerations can be made on the technical protection measures that may qualify as reasonable steps for the protection of trade secrets.

The requirement on reasonable steps is of fundamental practical importance for data owners, as it pertains to the actions that must be taken to maintain secrecy of information. However, there is **to date very limited interpretive guidance available on this requirement**, due to the novelty of the TSD and the absence of specific case-law from the CJEU on the meaning of reasonable steps.

A literal interpretation of the text of the Directive offers limited guidance of what could constitute reasonable steps. The only meaningful conclusion that can be drawn from the text relates to the 'under the circumstances' qualification of the requirement. This qualification suggests that the application of the requirement must be subject to a **contextual and case-specific assessment**, possibly encompassing an application of the principle of proportionality. In particular, the reasonableness test could be intended as indicating that the trade secret holder must take the steps that can reasonably expected considering the capacities of the holder, the circumstances of the case and the nature of the trade secret. For instance, factors that could be relevant to the proportionality assessment are the size and resources of the trade secret holder, the expected (cyber)threats to the secrecy of the data, the available technologies that can reasonably be used, and the commercial value of the trade secret. The trade secret holder is thus required to strike the right balance between all such factors and adopt the measures that can reasonably be expected in the circumstances.

Besides the abovementioned limited guidance stemming from the text of Article 2, the Directive as a whole does not offer any other meaningful indications on the meaning of reasonable steps. The recitals of the TSD do not mention the reasonable steps nor provide insights on what they could be in practice. In the absence of case-law by

[126] The CJEU recently received a request for a preliminary ruling that involved the interpretation, amongst others, of Article 2(1) of the TSD in case C-54/21, Antea Polska S.A., of 17 November 2022. However, despite the fact that the request referred also to the interpretation of the TSD, the CJEU did not consider necessary to delve into the interpretation of Article 2 of the TSD in order to deliver a preliminary ruling on the request. Therefore, the CJEU did not provide any clarifications on the interpretation of the TSD.

the CJEU on the meaning of reasonable steps, it may be instructive to consider how this requirement has been interpreted by national courts in the EU Member States, especially in the national case-law that has developed following the transposition of the TSD, and that is thus based on the national transposition of Article 2 of the TSD. Principles developed at the national level can offer particularly insightful interpretive guidance, because they might inform the future case-law of the CJEU on the TSD.

An analysis of the case-law in a portion of EU Member States shows that there is no explicit consensus on the meaning of reasonable steps[127]. Therefore, **while there may be similarities on the interpretation of this requirement across some Member States, an EU-wide understanding of reasonable steps cannot yet be extrapolated from national case-law**.

As a consequence, it is necessary to select the Member States whose case-law can be considered as most instructive to gain an understanding of what steps could be considered reasonable in practice. As a consequence, the case-law of **Germany, Austria, Italy and Spain** are discussed in this section, for the reasons set out below.

The **German** case-law is particularly insightful due to the fact that the reasonable steps requirement is an absolute novelty in German trade secrets law, as it was introduced for the first time by the TSD. Therefore, German case-law on reasonable steps is uniquely based on the transposition of Article 2 of the TSD in German law. German courts have been interpreting reasonable steps as an objective legal standards that is lower than ideal protection or extreme security[128]. This standard is to be set based on the circumstances of the case, and can be met with a combination of different types of measures as deemed appropriate in a given case, including a combination of technical, organizational and contractual means[129]. German case-law has identified a series of factors to be taken into account when assessing the adequacy of protection measures[130], i.e.:

a) the type of trade secret,
b) the specific circumstances of use,
c) the value of the trade secret and its development costs,
d) the nature of the information,
e) the importance for the company,
f) the size of the company,
g) the usual confidentiality measures in the company,
h) the type of labelling of the information, and
i) agreed contractual provisions with employees and business partners.

These factors confirm that the reasonable steps requirement entails a proportionality assessment, based on the resources of the trade secret holder and on the value and type of information to be protected as trade secret.

As concerns **Austrian** case-law, it is relevant for the **conceptual framing as an obligation of means** of the requirement to implement reasonable protective measures. In particular, the Austrian Supreme Court (*Oberster Gerichtshof*) ruled[131] that a logging system with a user name and password, that could be accessed only by a limited number of people, could be deemed as a sufficient protective measure, even if a security breach of this system occurred. Therefore, the trade secret holder was not under an obligation of result to guarantee security, but under an obligation of means to put in place protective measures. Even though this decision was taken before the implementation of the TSD in Austria, the Court referred to the TSD and stated that their interpretation was in line with the reasonable steps requirement of the Directive. **The understanding of reasonable steps as an obligation of means has important conceptual implications, as the trade secret holder is only required to demonstrate that reasonable steps were taken.** It also allows to draw a comparison with the security obligation of means laid down in Article 32 of the GDPR in relation to personal data, as will be further examined below.

---

[127] European Union Intellectual Property Office (''EUIPO''), ''Trade secrets litigation trends in the EU'', IPR Enforcement Case-Law Collection, 2023, p. 64.
In this report, the EUIPO has conducted an analysis of the litigation trends in 10 EU Member States selected on the basis of the volumes of legal proceedings. The Member States under scope in the analysis are: Belgium, Bulgaria, France, Germany, Italy, Netherlands, Poland, Romania, Spain and Sweden.

[128] Higher Regional Court of Düsseldorf, Decision No 15 U 6/20, 2021.

[129] Ibid, as the Higher Regional Court of Düsseldorf concluded that the claimant took reasonable steps through a combination of technical, organizational and contractual means.

[130] European Union Intellectual Property Office (''EUIPO''), ''Trade secrets litigation trends in the EU'', IPR Enforcement Case-Law Collection, 2023, p. 65.

[131] Austrian Supreme Court, Decision No 4 Ob 165/16t, 2016.

Italian case-law is remarkably extensive compared to that of other Member States[132] and offers interesting insights regarding two factors: i) the **conduct of the alleged infringer**, and ii) the **actions taken by the trade secret holder following an incident**[133]. With regard to the **first** factor, reasonable steps have been interpreted as measures that do not make it easy for specialists in the relevant sector to access the secret, or that would require extraordinary efforts. This interpretation implies that accessibility to the secret must not necessarily be rendered impossible by the reasonable steps, insofar as the infringer would face a certain level of difficulty. This is in line with the orientation in German case-law that ideal protection or extreme security is not needed. As concerns the **second** factor, Italian case-law has been taking into account the measures implemented by the trade secret holder after the infringement, in order to assess whether reasonable steps were taken before the infringement occurred. This *ex post* assessment is aimed at concluding whether the trade secret holder could have reasonably implemented certain protective measures even before the infringement, and not only after it occurred. For instance, the fact that a protective measure is implemented after the infringement can indicate that the trade secret holder had the resources to adopt it even before the infringement.

Finally, **Spanish** case-law is instructive with regard to the **holistic** nature of the analysis to conduct to verify if reasonable steps have been adopted in practice. The Court of Madrid affirmed in a 2016 case that the **steps to avoid disclosure should be adequate and reasonable, and that both internal and external measures are required**[134]. The external and internal dimension of reasonable steps offers important guidance to gauge how large is the array of measures that could constitute reasonable steps. Internal steps should be aimed at limiting access by employees and collaborators to trade secrets only to the extent that it is strictly necessary, whereas external steps must prevent unauthorized access to trade secrets by third parties. Both external and internal protective measures may be organizational, physical, legal and IT/technical. However, it may be the case that organizational, physical and legal barriers are more commonly implemented as internal steps, whereas IT barriers (such as cybersecurity safeguards) become more relevant to prevent unauthorized access by third parties.

While the Spanish judgement mentioned above was delivered before the transposition of the TSD in Spain, its guidance on internal and external steps is likely to be relevant also for the TSD. The particularly broad definition of reasonable steps in Article 2 of the Directive, and the reference to the circumstances of the case, suggest that reasonable steps may need to be adopted at any level, when the circumstances so require.

Based on the principles of EU Member States' case-law highlighted above it is possible to identify a set of **common features** that would most likely belong to the meaning of reasonable steps under the TSD. In particular, these features are as follows:

a) A **case-specific contextual assessment** is always warranted, because the qualification as reasonable of the measures adopted by the trade secret holder can be only be judged on the basis of the circumstances. This does not exclude that a low, objective threshold needs to be met irrespective of the circumstances, such as that the storage of data containing trade secrets is accompanied by measures controlling access to that data;

b) The existence of reasonable steps in a given case must be assessed in light of the principle of **proportionality**, in order to ascertain if, under the circumstances and considering the capacities of the trade secret holder, the steps taken are proportionate to the value of the trade secret. There are two implications stemming from this conclusion:
   a. The adoption of reasonable steps should always start from an analysis of the information protected as trade secrets;
   b. Reasonable steps can be designed on the basis of the characteristics and available resources of the trade secret holder, and of the costs that can be sustained by the latter;

b) The requirement to adopt reasonable steps does **not prescribe the attainment of a specific result**, and even less so to achieve optimal and extreme security of trade secrets;

c) Both **internal and external steps** may be needed, combining different types of safeguards. Based on the national case-law developed at the EU level, it can be argued that reasonable steps may be implemented at four levels[135]:
   - **Organisational**, e.g. by limiting access to trade secrets through a strict access policy, marking documents as confidential, splitting confidential information;

[132] European Union Intellectual Property Office (''EUIPO''), ''Trade secrets litigation trends in the EU'', IPR Enforcement Case-Law Collection, 2023, p. 66.

[133] Ibid.

[134] Provincial Court of Madrid, Decision No 441, 2016. This judgement was delivered before the transposition of the TSD in Spain.

[135] M. De Vroey, M. Allaerts, ''Trade secrets protection: an interim update of Belgian and EU case law'', Journal of Intellectual Property Law & Practice, Vol. 16, No. 12, 2021, p. 1394.

- **Physical**, e.g. camera surveillance, physical restriction of access, document destruction and physical authentication and identity verification of personnel;
- **Legal**, e.g. non-disclosure agreements with the relevant employees and collaborators;
- **Technical/IT**, e.g. encryption, obfuscation, remote storage, private use restriction and *ad hoc* cybersecurity measures.

The features identified above may thus be used as high-level guidance for the adoption of reasonable steps, with the disclaimer that any such steps must always be based on the circumstances of the case and that official guidance on the interpretation of Article 2(1)(c) of the TSD is lacking at present. Despite their high-level guidance, the principles set out above can be used to provide practical recommendations on how to facilitate trade secret protection by default, in particular in the design of technological solutions used for data sharing and storage.