tango-project.eu

TANGO

DIGITAL TECHNOLOGIES ACTING
AS A GATEKEEPER TO INFORMATION
AND DATA FLOWS

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

A TANGO Project White Paper on Distributed Infrastructures, Secure Data Exchange & Data Spaces - T8.6

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

Document Information

This section provides overall information about contributors and reviewers to the document along its iterative production.

List of Contributors	
Name	Partner
Renato Santana	EGI
Lauresha Memeti	ECO
Ladan Raeisian	ECO
Sotirios Michagiannis	DBC
Eleni Bakalarou	DBC
Giulia Giussani	IDSA
Ioannis Drivas	KUL
Andrea Palumbo	KUL
Dolores Ordóñez	ANYSOL

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

Table of Contents

Contents

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

List of Acronyms

Acronyms appearing along the document, in alphabetical order.

Abbreviation / acronym	Description
Data Set	A single, often structured, collection of data.
Data Space	A distributed system defined by a governance framework that enables secure and trustworthy data transactions between participants while supporting trust and data sovereignty. A data space is implemented by one or more infrastructures and enables one or more use cases.
EC	European Commission
Mxy	Month xy of the project's timeline
TANGO	Digital Technologies ActiNg as Gatekeepers to information flOws.
T8.6	Task 8.6, part of WP8
WP	Work Package

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

1 Introduction

1.1 Purpose and Structure of the Document

This document, a TANGO Project White Paper, concerns the concepts of Distributed Infrastructures, Secure Data Exchange and Data Spaces. These code concepts are spelt out in detail in three specific chapters, to provide the relevant and structured information on these. It aims to provide what is available and offers recommendations on what could be used by the project.

Following this introduction, the next three chapters dive deep into what is available for Distributed Infrastructure, Secure Data Exchange and Data Spaces, what has been used, and what recommendations are available for the TANGO Project. Finally, the conclusions summarize this White Paper.

2 Introduction to Distributed Infrastructure

In today's digital landscape, Distributed Infrastructure has become essential for organizations seeking to scale operations efficiently while ensuring robust resilience. This approach decentralizes computing resources across multiple locations, offering benefits such as enhanced load management and data redundancy.

However, the Internet currently lacks a critical component necessary for a decentralized and non-monopolized structure, with numerous concurrent cloud services and nodes, as envisioned by projects like TANGO. This missing element is an identity and trust layer essential for consuming services and interacting in a trustworthy manner.

Traditionally, users have been constrained by centralized cloud identity providers, limiting their options despite inherent advantages and disadvantages. Over recent decades, as digital reliance has grown, a shift in mindset has spurred the emergence of decentralized Web3 components, aimed at restoring the Internet's original decentralized design. One such innovation is Self-Sovereign Identity (SSI), rooted in the Web-of-Trust principle, which returns identity and data control to the individual.

What is recognized is the economic potential and stringent data protection requirements, businesses and governments worldwide, increasing integrating SSI, supported by initiatives like the European Union's ESSIF (European Self Sovereign Identity Framework). This convergence of Distributed Infrastructure and SSI promises to establish a foundation of trust and empower organizations to optimize deployment, security, compliance, performance, and sustainability in the evolving digital ecosystem.

2.1 Distributed Infrastructure

Distributed Infrastructure or decentralized infrastructure is a core principle, ensuring that data is not controlled by a single entity. Policies should encourage the development and adoption of decentralized technologies, such as distributed ledgers and peer-to-peer networks, to enhance data sovereignty and maintain a fair and open digital environment.

TANGO aims to drive innovation and bolster users' control over their data by establishing a secure, distributed digital environment. In this environment, data is provided, collected, and shared with utmost trust, allowing data owners to retain complete control over their information. TANGO also supports the creation of self-managed TANGO ecosystems, where various participants collaborate to create value and explore new market opportunities.

In the realm of Distributed Infrastructure, TANGO has identified and integrated existing software components and developed strategies to ensure seamless interoperability. This approach facilitates the creation of interconnected systems that support collaborative innovation. TANGO's Distributed Infrastructure is designed to integrate diverse data and infrastructure ecosystems, enabling efficient and resilient operations across multiple nodes.

TANGO's focus on Distributed Infrastructure ensures that its frameworks are adaptable and scalable, supporting dynamic, evolving needs and fostering continuous improvement. By leveraging a decentralized architecture, TANGO enhances system reliability and performance while providing the flexibility to accommodate new technologies and emerging requirements. This strategic approach positions TANGO as a leader in advancing Distributed Infrastructure solutions and meeting the demands of a rapidly changing technological landscape.

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

2.2 Policies and Recommendations for TANGO on Distributed Infrastructures

This section details the key policies and recommendations essential for the successful implementation and operation of the TANGO Project. Here, we outline the fundamental policies to govern security, data privacy, compliance, and operational resilience, along with best practices to enhance system performance, scalability, and interoperability. These guidelines are designed to support a robust and resilient Distributed Infrastructure, enabling effective and secure data management within the TANGO ecosystem.

For effective management of Distributed Infrastructure several fundamental policies and recommendations are essential.

Security policies should include strict access controls, data encryption for both transit and rest, and a well-defined incident response plan to address potential breaches.

Data privacy policies must clearly establish data ownership and retention guidelines to ensure responsible data management.

Compliance with relevant regulations and regular audits are crucial for adhering to data protection laws.

Operational policies should focus on disaster recovery planning to ensure service continuity during disruptions.

Recommendations for best practices include adopting a modular design for scalability and flexibility, ensuring system interoperability, automating routine tasks to enhance efficiency, and maintaining robust monitoring and regular updates to keep systems secure and performant. These policies and recommendations collectively support a resilient, secure, and efficient Distributed Infrastructure.

2.3 Infrastructure and Data Ecosystems: Data Interoperability, Data Models and Formats

As digital transformation accelerates globally, infrastructure and data ecosystems are becoming fundamental drivers of innovation, economic growth, and the realization of a data-driven society. A key enabler in this transformation is *****data interoperability*****, which facilitates seamless data exchange between different systems, organizations, and sectors.

At the heart of this ecosystem are **data models and formats** that standardize the way data is structured and shared. This section explores the critical components driving data interoperability, focusing on **Data Space Connectors**, their role in current data ecosystems, their usage in the market, and their specific implementation in the TANGO Project.

2.3.1 Data Interoperability

Data interoperability is essential for enabling diverse systems to communicate and share data effectively. It ensures that data from various sources, often in different formats, can be integrated and understood across platforms.

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

Key technologies, such as JSON-LD[1] and Smart Data Models, provide a standard structure for data. JSON-LD, for example, helps annotate web elements, creating structured data that can be indexed and processed by search engines and other applications.

Smart Data Models go a step further by specifying schemas and payloads that standardize technical data types, ensuring consistency across applications and industries.

A robust data interoperability framework is built on **data exchange APIs**, such as NGSI-LD, and advanced provenance and traceability tools. These components form the backbone of data-driven ecosystems, supporting everything from AI-enhanced services to secure data sharing within distributed networks.

2.3.1.1 What is a Data Space Connector?

A Data Space Connector is a software component that provides secure access to a data ecosystem, allowing participants to exchange data in a governed, interoperable manner. Data Space Connectors play a pivotal role in ensuring data sovereignty, meaning that organizations retain control over their data even when sharing it across platforms. They act as gateways between different systems, ensuring that data flows comply with defined security, privacy, and usage policies.

Data Space Connectors are based on the International Data Spaces (IDS)[2] and the IDS Reference Architecture Model[3], which define the technical requirements and governance structures necessary for trusted data sharing across diverse environments. More information is provided in Section 4 (Data Spaces).

Data Space Connectors typically include features such as usage control, encryption, and access management to ensure that only authorized entities can access and use the data. More information on Data Space Connectors is provided in the IDSA RAM in the specific chapter on "Connector Functionalities"[4].

[1] https://json-ld.org/

[2] https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram 4/introduction/1_1_goals_of_the_international_data_spaces

[3] https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/introduction/1_1_goals_of_the_international_data_spaces/1_2_purpose_and_structure_of_the_document

[4] https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer/3_5_2_ids_connector#ids-connector-functionalities

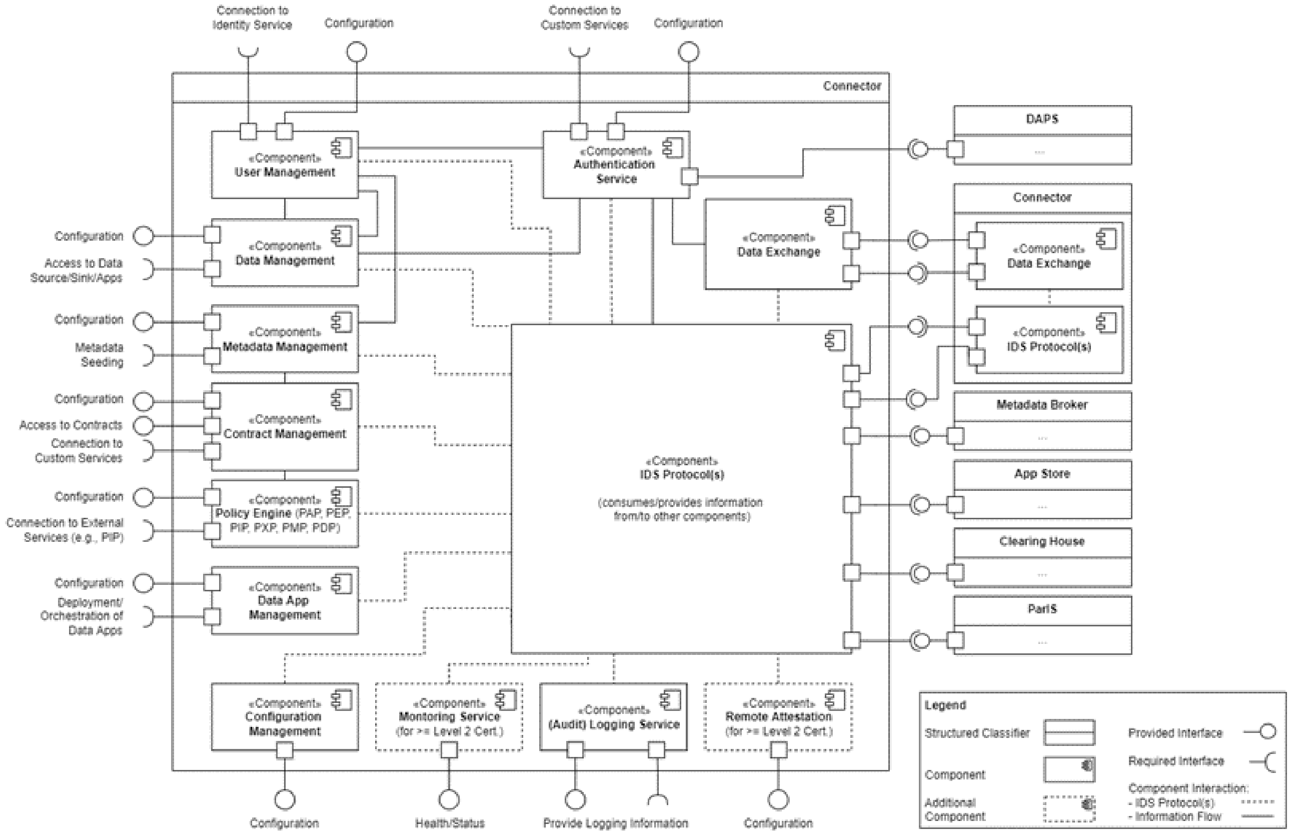

Connection to Identity Service
Configuration
Connection to Custom Services
Configuration
Connector
«Component» User Management
«Component» Authentication Service
DAPS
...
Connector
«Component» Data Exchange
«Component» Data Exchange
«Component» IDS Protocol(s)
Configuration
Access to Data Source/Sink/Apps
«Component» Data Management
Configuration
Metadata Seeding
«Component» Metadata Management
Configuration
Access to Contracts
Connection to Custom Services
«Component» Contract Management
«Component» IDS Protocol(s)
(consumes/provides information from/to other components)
Metadata Broker
...
App Store
...
Configuration
Connection to External Services (e.g., PIP)
«Component» Policy Engine (PAP, PEP, PIP, PXP, PMP, PDP)
Clearing House
...
Configuration
Deployment/ Orchestration of Data Apps
«Component» Data App Management
ParIS
...
«Component» Configuration Management
«Component» Monitoring Service (for >= Level 2 Cert.)
«Component» (Audit) Logging Service
«Component» Remote Attestation (for >= Level 2 Cert.)
Configuration
Health/Status
Provide Logging Information
Configuration
Legend
Structured Classifier
Component
Additional Component
Provided Interface
Required Interface
Component Interaction:
- IDS Protocol(s)
- Information Flow

Figure 1: Connector Functional View [5]

An essential element to make Data Space Connectors interoperable is implementing the Data Space Protocol (DSP). The DSP is a set of specifications designed to facilitate interoperable data sharing between entities, governed by usage control and based on web technologies. It defines the schemas, protocols, and interfaces required for entities to publish data, negotiate usage agreements, and access data within a federation of technical systems known as a data space. This protocol is crucial for ensuring technical interoperability and secure data exchange, enabling organizations to share data seamlessly while maintaining control over its usage. The DSP is due to become an international standard before the end of 2024[6].

2.3.1.2 Data Space Connectors Available on the Market

Several Data Space Connectors have emerged in the market, each designed to address specific challenges in data exchange and sovereignty. These connectors are used in various sectors, including manufacturing, healthcare, and autonomous vehicles, enabling applications like federated learning, smart contracts, and privacy-preserving AI. Their usage is cantered on enhancing data trust and minimizing risks associated with data sharing, especially in cross-border and multi-cloud environments. An extensive mapping of 30+ connectors available in the market is provided in the IDSA Data Space Connector Report. Published regularly by IDSA since November 2022, the Connector Report is a key publication to dig deeper on the topic of connectors. It does not only highlight the importance of Data Space Connectors, explaining

[5] https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3_5_0_system_layer/3_5_2_ids_connector#ids-connector-functionalities

[6] https://internationaldataspaces.org/offers/dataspace-protocol/

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

what they are and how to make them interoperable, but it also gives visibility to existing connector implementations providing insights on their unique value propositions, adoption examples, technical features, following their evolution over time.[7]

Examples include:

1.Eclipse Data Space Components (EDSC)

EDSC is a set of software components used to build connectors compliant with Data Space Protocol (DSP), though not connectors themselves.

While DSP v1 was only released at the end of last year, the level of compliance of these components with different specifications (e.g., RAM 4.0 / DSP 1.0) remains unclear. In theory, all DSP-based connectors should be interoperable, but this is not always the case.

2. Tractus-X EDC (Milestone 0.7)

Tractus-X is built upon Eclipse Dataspace Components. Despite being based on EDSC, many claim Tractus-X EDC is not fully compliant with DSP. This is attributed to its reliance on legacy systems, making it challenging to adopt DSP standards cleanly.

3. Japanese Connectors (CAADE)

Connectors developed in Japan (e.g., CAADE) are not on the list but are relevant players in the data space market.

4. Gaia-X / Prometheus-X / FIWARE Collaboration

Entities like Pierre, Prometheus-X, and FIWARE are jointly working on a connector that is purpose-built for DSP, avoiding legacy complexities. The focus is on building a clean implementation of the DSP from scratch, addressing criticisms of overly complex, legacy-based implementations.

5. Dawex Connector

Proprietary Approach: Dawex has its own proprietary connector.

6. C2D Connector

The C2D connector, developed by Kai, is used in several Gaia-X projects and follows its own connector protocol. The development is geared towards Gaia-X's vision for data spaces.

Overall, there is a wide array of solutions in the market, with many working towards DSP compliance. However, the complexity of legacy systems, trust models, and transport protocols still need to be addressed for full interoperability and clean implementations. The landscape is developing, but cooperation and clearer modular standards are required for seamless data sharing.

Different formats and models like JSON-LD and Smart Data Models contribute to broader data interoperability and exchange efforts, especially within initiatives like Gaia-X and IDSA.

2.3.1.3 Data Space Connectors in TANGO Project

The TANGO Project utilizes the FIWARE Data Space Connector, extended with several custom components to meet specific needs such as energy efficiency and green data

[7] https://internationaldataspaces.org/data-connector-report/

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

operations. This connector supports the project’s focus on data sovereignty and compliance with Gaia-X[8] and IDS frameworks.

One of the key advantages of the FIWARE Data Space Connector[9] is its ability to integrate seamlessly with distributed and federated infrastructures, allowing TANGO to operate across multi-cloud environments while maintaining stringent control over data sharing processes. Additionally, TANGO incorporates advanced security mechanisms like Self-Sovereign Identity (SSI) and “behavioural” authentication to further enhance trust in data exchange.

The project demonstrates the advantages of using robust Data Space Connectors, including:

- **Enhanced Data Sovereignty**: Ensures organizations retain control over their data, even in federated environments.

- **Energy Efficiency**: Implements smart contracts and encryption techniques that minimize energy usage.

- **Interoperability**: Adheres to international standards, ensuring compatibility across different systems and platforms.

- **Trust and Security**: Utilizes SSI and advanced encryption to protect data throughout its life cycle.

[8] https://gitlab.com/gaia-x/lab/lsd-connector/ and Home - Gaia-X: A Federated Secure Data Infrastructure
[9] https://github.com/FIWARE/data-space-connector

This project has received funding from the European Union’s HE research and innovation programme under the grant agreement No. 101070052

3 Policies and Recommendations Concerning Secure Data Transfer

In the digital landscape of the TANGO Project, ensuring the secure exchange of data is crucial for maintaining the integrity and confidentiality of information. This chapter introduces a comprehensive framework designed to safeguard data through robust encryption, secure communication protocols, and stringent access controls. It highlights the implementation of Identity and Access Management (IAM) and Data Loss Prevention (DLP) measures, while also emphasizing the need for ongoing security training and audits. Furthermore, the chapter explores the development and enforcement of Contract and Runtime Policies, supported by tools such as the Open Digital Rights Language (ODRL), to ensure effective management and compliance of data exchange agreements.

3.1 Compliance with Privacy and Security Legal Requirements

Technical solutions and data processing flows can be designed in a manner that enables, or at least facilitates, security and privacy-preservation in data exchanges. European Union (EU) legislation sets out multiple obligations to ensure the security, confidentiality, and privacy of data in different circumstances. Under EU law, security requirements generally aim to ensure the protection, by means of technical measures, against unauthorized or unlawful processing of the data, and against accidental loss, destruction, or damage. Confidentiality and privacy requirements, instead, aim to protect the confidential and private nature of data, as is the case for personal data and trade secrets. All these requirements generally pertain to the concept of security.

These obligations apply depending on a series of contextual elements, such as the circumstances of the processing and the type of data to be processed. However, when designing technologies to be used for data sharing in contexts, and with categories of data that are not clearly predetermined *ex ante*, certain policies can help design technical solutions and organizational measures to comply with the relevant EU legal framework. The subsections describe a non-exhaustive list of such policies, with the caveat that additional policies may become relevant depending on the circumstances of the case.

The policies described below are data obfuscation, data minimization, data abstraction, data separation, security of the processing, record-keeping and demonstrability. Some of these policies coincide with legal requirements (e.g. data minimization, record-keeping and demonstrability), while others may just facilitate compliance with a legal requirement (e.g. data separation and data abstraction).

3.1.1 Compliance Policies

- **<u>Data Obfuscation</u>**

Data obfuscation is the process of disguising data with the aim to render it unlinkable or unobservable. Data obfuscation has an essential role to play when both personal and non-personal data is stored and shared in order to achieve security, confidentiality, and privacy preservation, thus enabling compliance with the requirements imposed by the relevant legal framework.

It is of relevance when personal data and trade secrets are shared. When obfuscation techniques lead to anonymization, they make the GDPR and the ePrivacy Directive

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

inapplicable to the relevant data processing activities. When they result in pseudonymization, they facilitate compliance with both legal acts.

Pseudonymization can be essential in data storage and sharing to protect the data against unauthorized access, but also to limit the number of parties that have lawful access to the personal data by rendering the data anonymous for as many parties as possible. For instance, the data may be pseudonymized 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.

- **Data Minimization**

Data minimization is a data processing strategy that consists of limiting, to the largest extent possible, the processing of the data. Data minimization may be achieved with different techniques, such as minimizing the collection of data at the source, limiting the processing of data and the number of parties to whom it is exposed after collection, or 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, with Article 5 mandating that the processing of personal data must always be limited to what is necessary for the purposes of the processing. Data minimization is also instrumental to achieve compliance with the ePrivacy Directive.

Storage limitation, which can in turn be achieved through data minimization, is a requirement under the ePrivacy Directive in relation to traffic data, and location data other than traffic data. In particular, Articles 6 and 9 of the Directive prescribe that such data is processed only to the extent and for the duration that is necessary. Moreover, data minimization may be an important strategy to protect trade secrets against unintended disclosure. For this reason, it 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 different parties.

Data minimization is a data processing policy that can in turn be implemented using different sub-policies or techniques. Among the most relevant, it is worth mentioning data avoidance and limitation, access limitation, and the partial or total deletion of data.

- **Data Abstraction**

Data abstraction consists of limiting as much as possible the detail in which personal data is processed. It can be distinguished from data minimization since data abstraction is not about avoiding the unnecessary processing of data but rather limiting the level of detail in data processing.

It facilitates compliance with the relevant legal framework, in particular the legislation on trade secrets and data protection, for similar reasons outlined above about data minimization. A limitation on the level of detail in which data is processed facilitates the preservation of privacy and the confidentiality of commercially sensitive information.

- **Data Separation**

Data separation consists of logically or physically separating the processing of data.The aim of separation is to avoid a centralized processing of the data where a single entity has control over all the processing operations. An example, 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 any intelligible information about the secret. The secret can

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

subsequently be reconstructed when enough participants in secret sharing combine their shares.

Data separation offers significant benefits for the security, confidentiality, and privacy of the data being shared. Since no single person has a full view of a dataset or multiple datasets, it may not be possible to extract meaningful information from the separate data chunks that enables the identification of individuals, access trade secrets, or other sensitive information. To this end, separation must be carried out in a way that the possession of a single data chunk does not enable the acquisition of the protected information.

Data 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. Depending on how it is carried out, data separation would likely lead to the pseudonymization of personal data and, in certain cases, also anonymization. Separation can also constitute a reasonable step for the protection of trade secrets under Article 2(1)(c) of the TSD.

- **Security of the Processing**

Security is both a technical concept and a legal requirement. Article 5(1)(f) of the GDPR requires integrity and confidentiality in personal data processing, i.e., that is processed in a manner that ensures appropriate security of the personal data. This includes protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures.

When data is processed through automated 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’’.

- **Record-keeping and Demonstrability**

Certain information about how data processing takes place, especially with regard to the modalities and means of the processing, needs to be recorded to ensure security in data sharing. Record-keeping allows to keep track of information that may be needed to assess what went wrong in case of data breaches, as well as to monitor that data is being shared securely over time.

The EU legal framework sets out record-keeping and demonstrability obligations. Article 30 of the GDPR mandates record-keeping 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. 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, which entails that obliged entities should be able to present records of the relevant activities at any time when requested. As concerns trade secrets, record-keeping and demonstrability are also necessary to be able to prove, *ex post*, that the reasonable steps requirement were taken, i.e. to demonstrate which reasonable steps were adopted in practice.

3.1.2 Policies Enforcement

This chapter also explores the establishment and enforcement of policies essential for secure data exchange. These policies define the terms and conditions for data exchange and usage, focusing on Contract Policies and Runtime Policies.

This project has received funding from the European Union’s HE research and innovation programme under the grant agreement No. 101070052

Contract Policies lay the foundation for agreements between data exchange participants. They are designed to be both interoperable and unambiguous, are both machine- and human-readable. These policies cover access and usage, defined using ODRL (Open Digital Rights Language).

- **Key Elements:**

1. **General Description**: Details about the data asset, parties involved, and general terms.
2. **Access Policies**: Rules for accessing data on the provider's side.
3. **Usage Policies**: Obligations for the data consumer.
4. **Signatures**: Formalizes the agreement.

Contract negotiation uses frameworks like IDSA or GXFS-DE, resulting in a verifiable credential describing the data asset and contract.

- **Runtime Policies**

Derived from Contract Policies, Runtime Policies enforce these agreements within participants' systems. Defined using languages like Rego or XACML, they ensure the terms of the Contract Policies are upheld during data exchange.

- **Usage Control**

Usage control specifies and enforces what must or must not happen to data after access is granted. It is crucial for:

1. Intellectual Property Protection
2. Regulatory Compliance
3. Digital Rights Management

- **Access Control**

Access control restricts access to resources based on attribute-based policies, while usage control enforces restrictions on data usage post-access, binding policies to the exchanged data.

- **Policy Classes**

  Information Models: To express and execute Contract Policies during runtime:

  a. **Generic Ecosystem/Gaia-X Policy Data Models**: For basic discovery and trust negotiation.

  b. **Ecosystem/Industry-Specific Data Models**: Understood by all ecosystem participants.

  c. **Data Contract/Data Asset-Specific Data Models**: For specific usage restrictions.

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

- **ODRL (Open Digital Rights Language)**

ODRL describes content usage, including authorized, prohibited, and mandatory actions, along with resources and additional information like responsibilities and regulations. It supports usage control, ensuring usage restrictions are applied according to licenses and other conditions.

- **Main Concepts:**

1. **Policy**: A group of rules, including permissions, prohibitions, or obligations.
    a. **Set**: Generic collection of rules.
    b. **Offer**: Rules offered by an assigner.
    c. **Agreement**: Rules agreed upon by assignee and assigner.
2. **Rule**: Defines the rules of policies.
    a. **Asset**: The target resource.
    b. **Action**: Operation on the resource.
    c. **Constraint**: Conditions for the rule's validity.
    d. **Party**: Actors involved in the rule.
3. **Asset**: Targeted resources.
4. **Action**: Operations on assets, like "use" or "transfer."
5. **Constraint**: Logical expressions filtering collections.
6. **Party**: Involved actors, such as assignees (recipients) or assigners (issuers).

Additional metadata can be specified using Dublin Core Metadata, and policies can use inheritance for existing rules, with conflict resolution mechanisms.

Overall, this chapter underscores the importance of well-defined and enforceable policies for secure data exchange, using tools like ODRL to control and comply with agreed-upon terms.

3.2 Measures for Secure Data Transfer

To ensure compliance with the General Data Protection Regulation (GDPR), all processing operations on personal data should be identified and assessed pre-emptively, both in the case of any personal data transfers, as examined here, and under the "principle of lawfulness, fairness and transparency". The data controller is accountable for compliance with the GDPR and should be able to demonstrate compliance therewith. This subsection is focused mostly on the measures implemented in the TANGO platform, which ensures the secure transfer and exchange of information and personal data between the parties utilizing it. The implementation of robust encryption standards, utilization of secure communication protocols, and establishment of strong access controls contribute to a reliable and trustworthy data exchange environment.

Secure Data Transfer, or Data Exchange, plays a vital role in ensuring the project's success by safeguarding the confidentiality and integrity of information transmitted within the Distributed

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

Infrastructure. The implementation of robust encryption standards, utilization of secure communication protocols, and establishment of strong access controls contribute to a reliable and trustworthy data exchange environment.

Furthermore, the adoption of Identity and Access Management (IAM) solutions and Data Loss Prevention (DLP) measures enhances the overall security of the TANGO Project. Regular security awareness training and audits ensure that the system remains resilient against potential vulnerabilities and threats. By prioritizing Secure Data Exchange, the TANGO Project can maintain the trust of stakeholders, participants, and end-users while fostering a secure and efficient Distributed Infrastructure.

The following components/functions aim to ensure the project's security, reliability, and adherence to the data protection related legislation and regulations:

- **Data Encryption**: The use of encryption during the data exchange process is a vital part of the TANGO platform. Work package 3, which oversees the development of privacy-preserving data management and storage components, robust encryption standards will be implemented, such as Advanced Encryption Standard (AES) with 256-bit keys, to safeguard sensitive information during transmission and storage. Encryption will be used throughout the course of the data transfers through the use of the TANGO platform. Particularly, as a result of the operations of T3.4, the self-encryption and decryption component is foreseen and will be developed, which will be a useful tool for the users to ensure simultaneously the protection of data confidentiality and traceability and will allow them to recover the information by the use of the key which will be linked to the actual source of the information. The latter ensures data protection and prevention of unlawful access to the data since, in case of loss of the decryption tool/key, the actual "owner" of the information will be the only user capable of retrieving the information.
- **Secure Communication Protocols**: Secure protocols will be utilized, including HTTPS, SFTP, and FTPS, in order to prevent unauthorized access, data tampering, and other cyber threats, during the data exchange. Moreover, the TANGO platform will adopt standardized communication protocols, like the Message Queuing Telemetry Transport (MQTT) protocol and the Open Data Protocol (OData), to enable secure and real-time data transmission among different transport components.
- **Access Controls**: A data "tokenization" solution, combined with and adhering to the privacy and confidentiality by design, will be developed to establish stringent access controls to manage user access and enforce the principle of least privilege, ensuring data confidentiality and integrity. The access control protocols will be based on the user and device continuous "behavioural" authentication through the patterns of both the users and devices, while monitoring for any abnormalities.
- **Identity and Access Management (IAM)**: IAM solutions will be adopted to manage digital identities and access rights, guaranteeing that only authorized parties can access and share data.
- **Data Loss Prevention (DLP)**: DLP measures, both internally and externally, will be implemented to detect, monitor, and prevent unauthorized data transfers and exfiltration.
- **Privacy Related protocols**: For the development of TANGO components, and in order to ensure data security during the data exchange phase, the partners participating in T3.5 will be in charge of developing protocols and policies, related to the applicable legislation (most crucially the GDPR and the ePrivacy Directive). The aim of this process is to develop protocols with personal data and non-personal data containing business-sensitive information will be treated in the same manner, thus ensuring the development of a technological solution that will offer a standardized level of security for all users.
- **Regular Security Audits**: The conduction of periodic security assessments and audits is foreseen to identify and address vulnerabilities, ensuring the TANGO Project's security and reliability. During both the course of the project related activities, and also during the

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

utilization of the final TANGO platform, the partners in charge will conduct regular audits and assessments to ensure that data security and integrity is achieved during the use of the platform. This is especially important in terms of the processing operations on personal data as a result of the data transfers.

- **Compliance with Data Protection Regulations**: In general, the WP3 partners oversee ensuring the adherence and compliance of the developed components of the TANGO framework to/with the relevant data protection regulations, such as the GDPR. Partners involved in T1.4 are also in charge of monitoring the project-related activities in terms of privacy and data protection, to maintain trust, protect the privacy of individuals involved in the project, and to ensure the compliance of the project-related activities with the applicable legislation.

As mentioned in the opening paragraph of this section, the measures implemented to ensure data security and integrity during the data exchange and transfer period are outlined.. With their implementation, the TANGO Project can establish a secure and efficient Distributed Infrastructure, protecting sensitive data and maintaining the confidence of stakeholders, participants, and end-users. Further information regarding the components developed and utilized because of the TANGO Project, from a more technological perspective, will be presented in following publications.

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

4 Data Spaces

As organizations embrace data-driven strategies, understanding and harnessing the potential of data spaces becomes crucial. For this reason, this chapter provides insights on Data Spaces Concepts and Policies on Data Spaces.

The section “Data Spaces Concepts (4.1)” explores the fundamental concepts related to data spaces, providing the definition of data space, highlighting its purpose and providing insights on foundational elements.

The section “Policies on Data Spaces (4.2)” delves deeper into the landscape of organizations and provides a set of recommendations for policies. It sheds light on the existing initiatives, provides some anchors on relevant enablers, and highlights key policies to the set up and maintain of data spaces.

4.1 Data Spaces Concepts

This section first outlines the general definition and purpose of a data space, then dives deeper into the specifics of its foundational concepts

4.1.1 Definition and Purpose of Data Spaces

A data space is a “distributed system defined by a governance framework that enables secure and trustworthy data transactions between participants while supporting trust and data sovereignty. A data space is implemented by one or more infrastructures and enables one or more use cases.”[10]

Data spaces represent a paradigm shift, from isolated data repositories and data silos, toward a federated data ecosystem based on shared policies and rules. In data spaces, diverse actors collaborate to share data in a secure, reliable, and trustworthy manner. These spaces follow common governance, organizational, regulatory, and technical mechanisms. Users within data spaces can access data seamlessly, benefiting from a unified approach that ensures security, transparency, and ease of use.

A key aspect of data spaces is data sovereignty, i.e., as cited by the International Data Spaces Association, regarding “the ability of organizations, governments, and individuals to exclusively and sovereignly decide how their data is collected, stored, shared, and used by others”.[11]

By establishing data sovereignty, organizations create an environment of trust and reliability, fostering innovation and collaboration in data-sharing ecosystems.[12]

Data spaces are based on two main principles: interoperability and transversality. Interoperability relates to the capacity to connect with any system and platform for data sharing, while transversality refers to the possibility that data spaces offer to exchange data across different sectors. This is especially important for TANGO use cases, since this possibility can generate new business models apart from new efficiencies and increase sustainability.

4.1.2 Foundational Concepts

Data and technology – and data spaces – are integral components of complex human systems which reflect the values of the people involved. Data sets are collected by people, who decide what data to collect and how to do so. These choices, in turn, are linked to values; they indicate

[10] https://dssc.eu/space/Glossary/176554052/2.+Core+Concepts

[11] Data Sovereignty as a Key Capability | IDS Knowledge Base (internationaldataspaces.org) and Data Sovereignty as a Key Capability | IDS Knowledge Base (internationaldataspaces.org)

[12] Data Sovereignty - International Data Spaces

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

which data people consider important to measure and collect. They can be used for purposes that support or go against the values of their users and their societies. An example of this is nuclear technology, which enabled both the atomic bomb and radiation therapy to treat cancer.

Solid values and ethics are fundamental to any technical implementation; this is why the use of data is not a mere usage of technology but needs good governance goals. Data spaces are deeply rooted in the European values such as freedom, inviolability, privacy, security, humanity, and respect, and therefore bring forward a holistic approach to the data economy weighing the impact on people and societies.

To identify these guiding principles in data spaces, the PESTLE analysis has been used (ref).[13] The result is a macro picture of the environment of a data space, tackling six core aspects: political, economic, social, technical, legal, and environmental levels:

4.1.2.1 Political Level:

The European Data Strategy aims to maintain data sovereignty and promote digital transformation by 2030 through various legislative acts, emphasizing the need for legal interoperability and standardization.

4.1.2.2 Economic Level:

Data sharing is crucial for economic success for local, national, and international economies. Data spaces can enhance collaboration, resilience, and fairness in value chains.

4.1.2.3 Social Level:

European ideals such as freedom, privacy, and respect are embedded in data spaces. The implementation of these values varies based on the needs of different communities and stakeholders.

Technical level: Data spaces should use widely accepted protocols and standards to ensure interoperability. Organizations like W3C, ISO, and IEC play key roles in developing these standards.

4.1.2.4 Legal Level:

Legislation often follows political decisions, and new regulations must consider political and social sentiments. Key legal areas in data sharing include antitrust, data protection, copyright, and intellectual property, with significant regulations introduced by the European Data Strategy.

4.1.2.5 Environmental Level:

Data usage has a significant environmental impact, with the ICT sector consuming a large portion of global electricity and emissions. The EU Data Strategy aims to use data to address climate challenges, promoting sustainable digital technologies and continuous monitoring to ensure a positive climate impact.

These guiding principles highlight that data spaces are very complex structures, which aim to solve not purely technical challenges like interoperability but encompass also business and legal challenges. For this reason, several organizations have been working to enable data spaces and have provided different policy recommendations.

[13] Guiding Principles | IDS Knowledge Base (internationaldataspaces.org)

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

4.2 Policies in Data Spaces

This section aims to provide insights off policies and recommendations for data spaces. To achieve this, it first briefly outlines the landscape of major initiatives associated with data spaces, introduces the fundamental policy types, and concludes with a selection of policies and recommendations to facilitate navigation through this emerging framework.

4.2.1 Data Space Landscape

The landscape of data space initiatives is diverse, including various organizations working together to offer frameworks, tools, and support for data spaces, as well as several endeavours which have or are in the process of setting up fully operating data spaces or demonstrators. This Section focuses on the first category. For the second category, an extensive mapping is available and kept up to date in the Data Spaces Radar. The Radar is a publicly accessible tool designed to offer a panoramic view of various data space initiatives worldwide, offering insights into their focus sectors, location, and development stages, as well as use cases which are enabled by a data space (ref.).[14]

The **International Data Spaces Association (IDSA)** is one of the pioneering initiatives promoting data spaces since 2016. IDSA plays a crucial role in developing standards and frameworks in achieving interoperability and data sovereignty in data spaces. Their work is fundamental to achieve interoperability and trust within data spaces at a global level.

Another significant initiative is **Gaia-X**, a European Association founded officially in January 2021.[15] The Gaia-X Association focuses on sovereign cloud services and cloud infrastructure by defining guidelines for the soft data infrastructure (ref. internationaldataspaces.org).[16]

The **FIWARE** Association is also noteworthy: it is a not-for-profit association driving the definition – and the Open Source implementation – of key open standards that enable the development of portable and interoperable smart solutions in a faster, easier, and affordable way. FIWARE offers an open-source platform that supports the development of Smart Solutions, Digital Twins, and Data Spaces in several domains (ref. https://www.fiware.org/about-us/).

The **Big Data Value Association, BDVA**, is an industry-driven research and innovation organization with a mission to develop an innovation ecosystem that enables data-driven and AI-enabled digital transformation of the economy and society in Europe.[17]

To speak with one voice and share expertise, IDSA, Gaia-X, FIWARE, and BDVA have created the Data Space Business Alliance (DSBA) in September 2021. The proceedings of the alignment of architectures are published under the name of “Technical Convergence Paper” on the DSBA webpage.[18] Its latest version is 2.0 from 2023.[19]

The initiative **iSHARE** also contributes to the creation of data spaces by providing a framework for data sharing in the logistics sector, simplifying and standardizing data sharing agreements to make the process more efficient and secure.

Lastly, the **Data Spaces Support Centre** is a European project started in October 2022 and Terminating in March 2026. It aims to support and coordinate data spaces across various sectors by providing blueprint architectures and infrastructure requirements, as well as

[14] https://www.dataspaces-radar.org/radar/
[15] https://gaia-x.eu/what-is-gaia-x/about-gaia-x/
[16] IDSA-Position-Paper-Data-Spaces-Landscape-1.pdf
[17] https://bdva.eu/
[18] The Data Spaces Business Alliance - Data Spaces Business Alliance (data-spaces-business-alliance.eu)
[19] Data-Spaces-Business-Alliance-Technical-Convergence-V2.pdf

This project has received funding from the European Union’s HE research and innovation programme under the grant agreement No. 101070052

facilitating cross-sector data reuse. It offers a starter kit with resources to help organizations create and maintain data spaces, and supports the Data Innovation Board in enhancing data interoperability and sharing services.[20]

These examples illustrate the collaborative efforts and diverse approaches being taken to build a robust data space ecosystem. By working together, these organizations are paving the way for a future where data can be shared securely and efficiently, driving innovation and digital transformation across industries.

More details and additional initiatives are described in the IDSA Position Paper.[21]

4.2.2 Types of Policies Needed in Data Spaces

The creation of a data space is a complex process, involving a variety of complex decisions across several domains, from technical to business and governance. A general guideline to the process is provided in the IDSA Rulebook.[22] Based on these steps, four major types of policies can be identified.

Membership Policies (MP): These policies ensure that only qualified participants can join the data space. They make sure that potential participants meet specific criteria before they are allowed to enter the data space, e.g. verifying the nationality of participant or checking if they have the necessary industry certifications.

Access Policies (AP): These policies control who can access data contracts within the data space. They define which attributes must be available to access data contracts. For example: some policies might allow access to all participants within the data space but hide items from non-members or ensure that only participants with specific attributes can see certain data contracts. Each participant can define such policies, whether providing or consuming data. This enables a participant to define a policy to see only data with a distinct proof of origin or to offer data restricting access to members of a certain jurisdiction. This is often referred to as provider policy and consumer policy.

Contract Policies (CP): These policies govern the terms and usage of data within contracts. Contract Policies include Contract Agreement Policies, which define the attributes needed to be able to establish a contract.

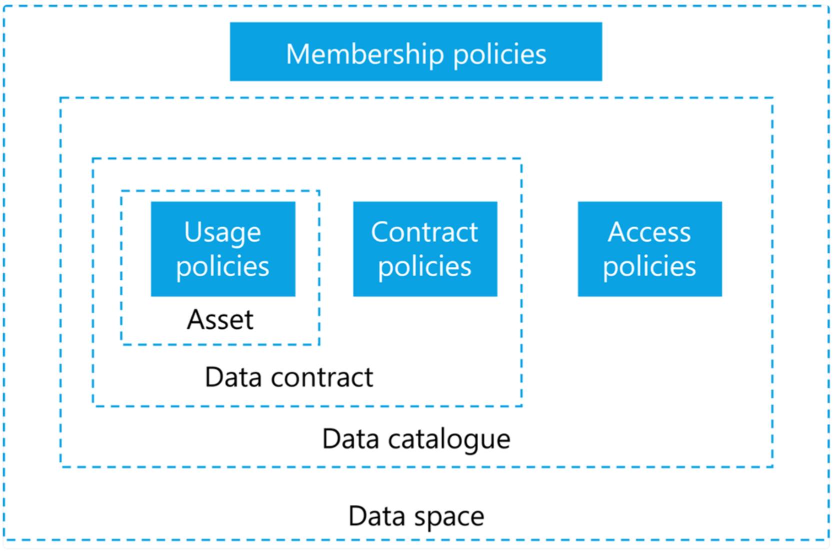
Usage Policies (UP): CP also include Usage Policies (UP), which control how the data can be used by the receiving party after transmission, for example ensuring that the participant uses a specific encryption.

A summary of the different policies is reported (see Figures 3 and 2 bellow).

[21] https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-Data-Spaces-Landscape-1.pdf

[22] https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/idsa-rulebook/3_functional_requirements#creating-a-data-space

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052



Different policies in data spaces

Figure 3 [23]

It is important to note that Membership Policies are defined by the data space authority and may vary depending on the design and requirements of the data space. Data access policies and data contract policies are specified by participants in their data contracts to further restrict access and usage

Additionally, we should not forget that data spaces are rooted in our societies; therefore, additional levels of policies should be considered, for example, policies established by governments to provide a legal and organizational framework for data sharing.

In terms of interaction of policies, the International Data Spaces Association (IDSA) provides a technical framework to support data sharing across different sectors. The legal and organizational framework is established by governments to ensure a trusted environment for data sharing. The data space authority defines data space policies and ensures compliance with global or general policies. Participants connect to the data space using Data Space Connectors which understand, and exchange data based on the defined policies, as spelt out in Section 4.1. An overview of these interactions is provided below:

[23] https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/idsa-rulebook/3_functional_requirements#policies

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052

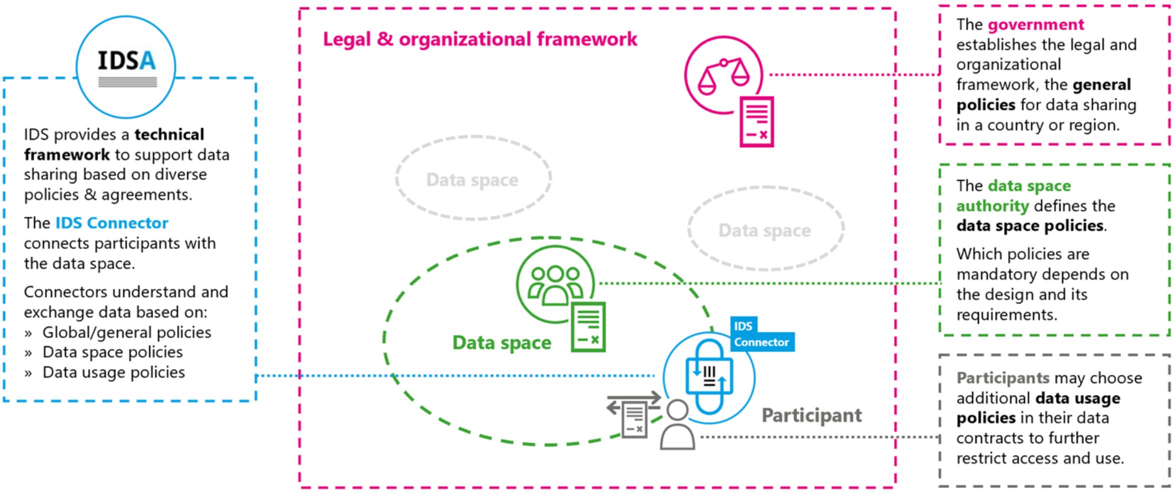

IDSA
IDS provides a technical framework to support data sharing based on diverse policies & agreements.
The IDS Connector connects participants with the data space.
Connectors understand and exchange data based on:
» Global/general policies
» Data space policies
» Data usage policies
Legal & organizational framework
Data space
Data space
Data space
IDS Connector
Participant
The government establishes the legal and organizational framework, the general policies for data sharing in a country or region.
The data space authority defines the data space policies.
Which policies are mandatory depends on the design and its requirements.
Participants may choose additional data usage policies in their data contracts to further restrict access and use.

Figure 4: (Picture from IDSA)

4.2.3 Collection of Policies and Recommendations

Data spaces are a complex and multifaceted topic, involving a diverse landscape of initiatives and organizations working together to create a secure, interoperable, and trustworthy environment for data sharing. The following policies and recommendations have been collected in the context of the TANGO Project:

DSSC Blueprint: The DSSC Blueprint provides a comprehensive set of guidelines to support the data space development cycle. It includes a conceptual model, building blocks, and a recommended selection of standards, specifications, and reference implementations. More information at: [24]

IDSA Reference Architecture Model (IDSA RAM): The IDSA RAM serves as the conceptual basis for IDS-compliant data exchange between organizations. It defines the fundamental concepts, functions, and processes involved in creating a secure network of trusted data. The RAM is essential for ensuring data sovereignty, security, and interoperability within data ecosystems. By adhering to the IDSA RAM, organizations can establish a robust framework for data sharing that aligns with industry standards and promotes trust among participants. Currently, the latest version of the IDSA RAM is version 4, and he IDSA is currently working on a version 5. More information at:[25]

IDSA Rulebook: The IDSA Rulebook provides a comprehensive governance framework for the development and operation of data spaces. It includes guidelines for functional, technical, operational, and legal dimensions, ensuring that all aspects of data space management are covered. This resource is crucial for maintaining data sovereignty, compliance with regulatory requirements, and fostering collaboration among data space participants. By following the

[24] Data Spaces Blueprint v1.0 - Blueprint v1.0 - Data Spaces Support Centre (dssc.eu)
[25] https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4

IDSA Rulebook, organizations can ensure that their data spaces operate smoothly and securely. More information at: [26]

Data Space Protocol (DSP): The DSP is a set of specifications designed to facilitate interoperable data sharing between entities, governed by usage control and based on web technologies. It defines the schemas, protocols, and interfaces required for entities to publish data, negotiate usage agreements, and access data within a federation of technical systems known as a data space. This protocol is crucial for ensuring both technical interoperability and secure data exchange, enabling organizations to share data seamlessly while maintaining control over its usage. The DSP is becoming an international standard before the end of 2024. More information at: [27]

In addition to these essential cross-domain resources, some **domain specific blueprints** are also available. They have been developed in the context of Coordination and Support Action (CSA) projects funded by the European Commission. Note: [28]

Some examples of such CSAs and their domain-specific blueprints relevant for the TANGO use cases are listed below:

- DATES project, led by AnySolution, who is also the lead of the Smart Hospitality use case.[29]
- PrepDSpace4Mobility.[30]
- Manufacturing data space.[31]

[26] https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook

[27] https://internationaldataspaces.org/offers/dataspace-protocol/

[28] CSA were designed to support the coordination and networking of research and innovation projects, programs, and policies. They often involve activities such as mapping existing ecosystems, identifying gaps, and proposing frameworks for future developments.

[29] https://www.tourismdataspace-csa.eu/wp-content/uploads/2024/01/DRAFT-BLUEPRINT-Tourism-Data-Space-v3.3_final.pdf

[30]https://www.mobility-data-space.de/content/dam/ivi/mobility-data-space/documents/Mobility_Data_Space_2022_EN.pdf

[31] https://manufacturingdataspace-csa.eu/#about

5 Conclusions

The following conclusions can be drawn from this TANGO White Paper:

- Concerning Distributed Infrastructures:

Data Space Connectors are vital components of modern data ecosystems, enabling secure and interoperable data exchange across various platforms and sectors. As exemplified by the TANGO Project, leveraging advanced connectors like the FIWARE Data Space Connector helps organizations navigate the complexities of data sovereignty, security, and energy efficiency in a rapidly evolving digital landscape. By embracing standardized data models and formats, organizations can unlock new opportunities for innovation, collaboration, and sustainable development in the data economy.

- Concerning Secure Data Transfer:

By implementing the policies and recommendations, in Chapter , the TANGO Project can establish a secure and efficient Distributed Infrastructure, protecting sensitive data and maintaining the confidence of stakeholders, participants, and end-users.

Chapter 3.1 the implemented measures that will ensure data security and integrity during the data exchange – transfer period

By utilizing the connectors 3, the TANGO Project aims to create an integrated, sustainable, and user-centric transport ecosystem, addressing the challenges of urban mobility and fostering collaboration among public and private stakeholders.

- Concerning Data Spaces:

Data spaces are a complex and multifaceted topic, involving a diverse landscape of initiatives and organizations working together to create a secure, interoperable, and trustworthy environment for data sharing. The following policies and recommendations have been collected in the context of the TANGO Project:

DSSC Blueprint:

IDSA Reference Architecture Model (IDSA RAM)

IDSA Rulebook
Data Space Protocol (DSP)

This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101070052