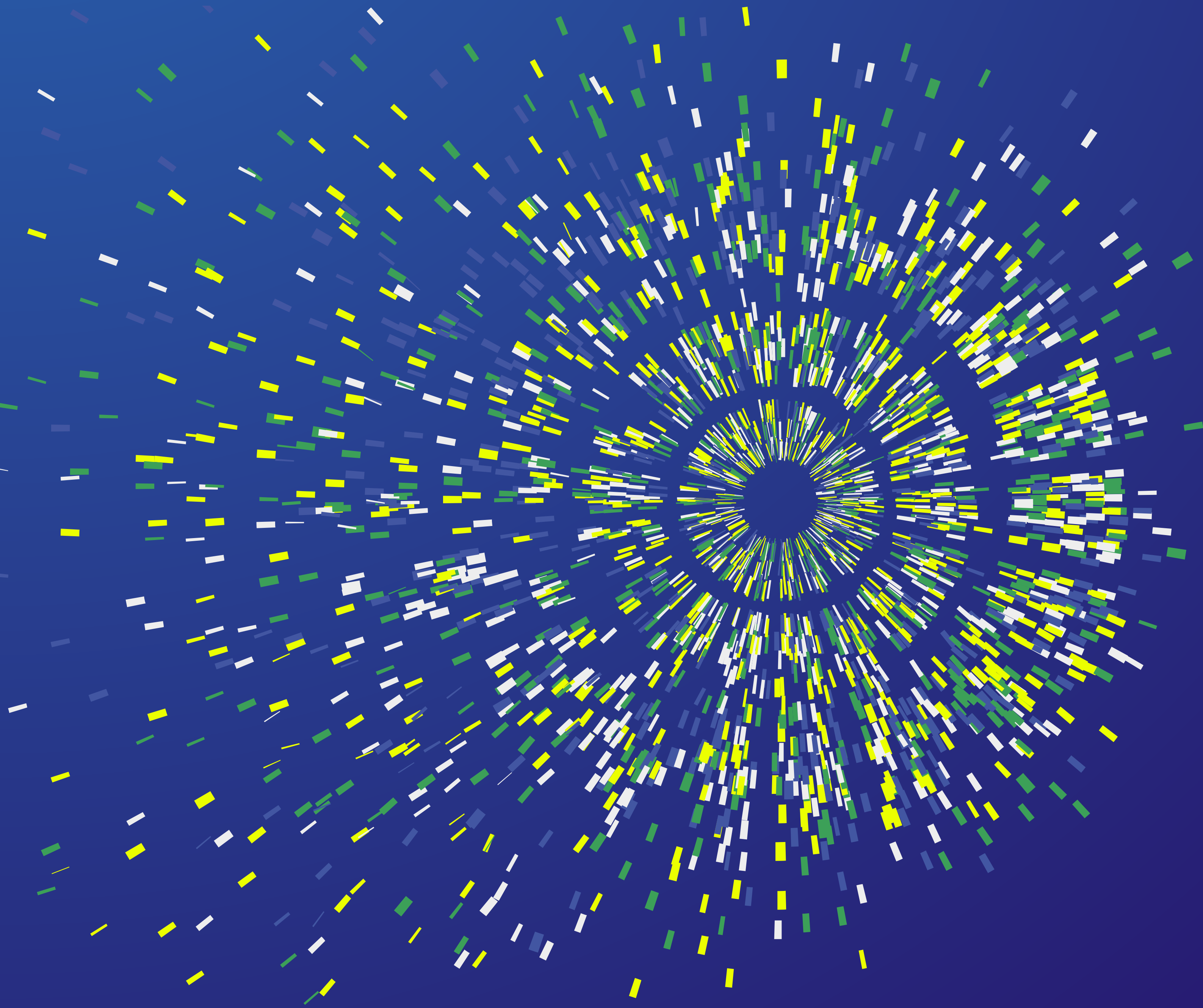
LEVERAGING THE BENEFITS OF COMBINING DATA SPACES AND PRIVACY ENHANCING TECHNOLOGIES

Joint BDVA and CoE DSC White Paper



BDV BIG DATA VALUE ASSOCIATION

CoE DSC

I. MANAGEMENT SUMMARY AND AUTHORS

I.1. Management Summary

Data Spaces form the core of the EU Data Strategy, providing a framework for sharing of data between stakeholders according to the European values on data sovereignty and trust. Data Spaces include important architectural elements on both the governance and the technical level, like findability and accessibility of data sources, data interoperability and semantics, usage control and trust mechanisms.

Privacy Enhancing Technologies (PETs) refer to a (relatively new) set of technologies that no longer requires the sharing of (privacy) sensitive data between stakeholders in a readable way, which still forms an information leakage risk, despite contractual and other trust measures. PETs encompass various technologies that enable specific pre-agreed analytics to be carried out while keeping the sensitive data secret. As such, they provide considerably privacy guarantees at the technical level *"by-design"*.

Data Spaces can benefit from PETs, as the latter enrich the set of data sharing services available in a Data Space, increase the number of use cases supported in a Data Space and made them available to participants (data providers and consumers) through a single-point-of-entry, yielding economy of scope benefits for a Data Space and efficiency in onboarding benefits for its participants.

Vice versa, PETs can benefit from Data Spaces as the latter make PETs deployment easier than in many other environments, paving the way towards scale-up of PET implementations. For instance, many stakeholder roles and operations processes are already defined for Data Spaces and may be re-used in the operations of PETs. Moreover, for data providers, the availability of (integrated) PETs means less implementation risks and lower vendor lock-in. The providers of PET services benefit from the surge in attention and drive for introducing Data Spaces as part of the EU Data Strategy.

However, although potential mutual benefits may be clear, alignment between Data Spaces architectures and PET solutions is currently not straightforward, requiring continued effort. Aligned business models, interoperability of architectures and solutions, and appropriate legal and governance approaches are among the main open points that need to be solved. Privacy patterns can provide a methodology to link a common technical grounding and the operations processes provided in a Data Space to PETs orchestration processes for information protection, enabling technical interoperability and ease of deployment. The definition of a joint Data Space and PET role model will provide guidance and the fundament to an aligned business model and reference architecture.

This white paper is intended for the Research & Innovation communities for both Data Spaces and PETs. It includes recommendations on both organizational and business and on technical aspects for further alignment of the Data Spaces and PETs initiatives as listed in the table below. It proposes to put the initial focus for Data Spaces and PET alignment on the operations processes. Common Privacy Patterns may be supported in Data Spaces for interfacing and abstracting between the Data Space capabilities and responsibilities and the PET capabilities and responsibilities.

I.2. Recommendations for Data Space and PET alignment

Organizational and Business Recommendations	Technical Recommendations
• Define a harmonized (business) role model for Data Space and PET alignment	• Develop a market place for both data sources and data processing algorithms
• Align on common operations processes	• Provide a common technical foundation
• Define a legal and governance structure and framework	• Define the appropriate abstraction layers and interfaces between Data Space and PET responsibilities
• Identify and describe a common scope and terminology	• Identify common Privacy Patterns (archetypes)
• Set up a community on Privacy Patterns & models for PET and Data Space interoperability	• Support Business Process and Workflow Management and Orchestration tooling
	• Develop a tool-set for interoperability and integration

I.3. Authors

This paper is the result of a cooperative work by Big Data Value Association (BDVA) members and the Center of Excellence Data Sharing and Cloud:

Abhishek Mahadevan Raju (TNO)
Aitor Corchero (NTT Data)
Antonio Kung (Trialog)
Daniël Worm (TNO)
Elena Lazovik (TNO)
Enric Staromiejski (NTT Data)
Evangelos Markatos (Forth)
Freek Bomhof (TNO)[1]
Harrie Bastiaansen (TNO)
Jose M. del Alamo (UPM)
Lucas Sanjuan Viñas (ITI)
Monica Florea (SIMAVI)
Patricia Jimenez (NTT Data)
Rizkallah Touma (i2CAT Foundation)
Roberto di Benardo (ENG)
Venkatesh Kannan (Ichec)
Vincenzo Savarino (ENG)
Zoltan Mann (UvA)

Reviewers:
Ana Garcia Robles (BDVA)
Caj Södergård (NextAI)
Gert Kruithof (CoE-DSC)
Tuomo Tuikka (VTT)
Valerio Frascolla (Intel)

Version 1.0, March 2024.

[1] Corresponding author

II. BACKROUND: SETTING THE SCENE

As the European Commission (EC) has clearly recognized the importance of federative data sharing, the development of Data Spaces is a core element of the European Data Strategy [1]. The EC has expressed its ambition on federative data sharing in the EU Data Strategy as the "common European Data Spaces". A major goal of the European Data Strategy is to develop a common (technical) ground that enables data sovereignty and trust through controlled data sharing in a federation of interoperable Data Spaces. A multitude of European Data Spaces is currently emerging, e.g. for individual sectors, application areas or geographical regions. The EC plays an active role in the development and the deployment of Data Spaces by providing the foundational regulations, developing the reference architectures and associated (open-source) building blocks, and supporting the deployment of Data Spaces in multiple sectors considered to be of main interest for the EU [2].

At the same time, Privacy Enhancing Technologies (PETs) are gaining popularity for accessing and processing data in cases in which the various data sources cannot simply be shared between stakeholders. This may specifically be the case due to sensitivity, confidentiality, ethical, privacy or other legal issues, which require that sensitive data remain within the security domain of its provider or administrator and are not transferred to or shared with other organizations. For such cases, PETs allow to process sensitive data locally within the provider's security domain, allowing only non-sensitive (and/or encrypted) processed information to be shared with external stakeholders. Hence, only controlled access is provided to sensitive data within the provider's security domain without sharing any sensitive information.

The Data Space and PET initiatives pursue a common goal in providing concepts, architectures and solutions for making data analysis available to be used across organizations, while emphasizing the importance of European core values with respect to data sovereignty, trust and privacy. With these common goals, the question arises to which extent the Data Space and PET concepts overlap, and to which extent they are complementary and/or reinforcing each other.

This white paper identifies the potential benefits, the challenges and the needs for aligned development of PET and Data Space concepts and (reference) architectures. Chapter III describes Data Space and PET developments. Chapter IV elaborates the potential benefits for alignment in development and deployment of the Data Space and PET concepts and architectures, considering the functional, operations, and business and funding perspectives. The approach for alignment in terms of various commonalities to be jointly exploited is described in Chapter V. The final Chapters VI and VII provide the overarching conclusions and recommendations with a call to action for further alignment of Data Space and PET development efforts.

The audience for this work is the Research & Innovation communities in both Data Spaces and PETs.

III. DATA SPACES AND PRIVACY ENHANCING TECHNOLOGIES: THE DEVELOPMENTS

Both Data Spaces and PET initiatives are currently attracting major attention in several vertical market segments, however to unleash their joint potential some gaps have been identified. This chapter addresses those gaps by describing both the Data Space and PET developments leveraging on an illustrative use case. This is followed by a section on the state-of-the-art on the work in combining the Data Space and PET concepts.

III.1. Data Spaces

The ambition of the European Data Strategy [1] is summarized as providing a "common European Data Spaces". Various EU initiatives are exploring and developing reference architectures for Data Spaces. Currently, the EU Data Spaces Support Centre (DSSC) [3] is the main initiative working towards a blueprint for the emerging (federation of) Data Spaces in Europe. It has defined a Data Space [4] as *"an infrastructure that enables data transactions between different data ecosystem parties based on the governance framework of that Data Space. Data Spaces should be generic enough to support the implementation of multiple use cases"*.

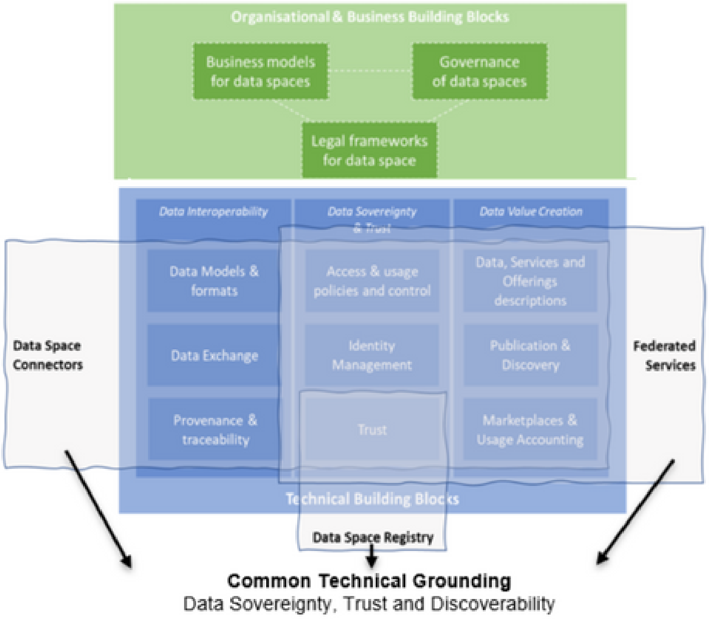

Organisational & Business Building Blocks
Business models for data spaces
Governance of data spaces
Legal frameworks for data space
Data Interoperability
Data Sovereignty & Trust
Data Value Creation
Data Models & formats
Access & usage policies and control
Data, Services and Offerings descriptions
Data Space Connectors
Data Exchange
Identity Management
Publication & Discovery
Federated Services
Provenance & traceability
Trust
Marketplaces & Usage Accounting
Technical Building Blocks
Data Space Registry
Common Technical Grounding
Data Sovereignty, Trust and Discoverability

Figure 1: The DSSC taxonomy of building blocks

The functionality that may be provided by a Data Space has been defined by the DSSC as a set of building blocks, both as part of its taxonomy [5] and its blueprint [6], as depicted in Figure 1.

The DSSC taxonomy distinguishes between two categories of building blocks for a Data Space:

Organizational and business building blocks
These relate to business models, the governance and the legal frameworks for Data Spaces.

Technical building blocks
These relate to the technical aspects and technical agreements that individual Data Space participants and trusted intermediaries need to adhere to.

In addition, the technical building blocks require a “Common Technical Grounding", which involves their implementation in software or services intended for use not only in individual Data Space instances but also for ensuring interoperability between a multitude of emerging Data Spaces. The focus of the Common Technical Grounding is on building blocks for data sovereignty, trust and discoverability. It encompasses three main categories of building blocks:

1 **Data Space connectors** serving as secure gateways, enabling systems and organizations to access a Data Space securely,

2 **Federated services** offering various functionalities, such as validation or cataloguing of services

3 **Data Space registries** registering the participants of a Data Space

The DSSC initiative has the charter to define the architectural blueprint of the Data Spaces and their building blocks [3]. The EU SIMPL procurement project [7] will adopt the DSSC blueprint, develop the associated building blocks, and make these available open-source SW for large scale deployment in the various EU sectoral Data Space deployment initiatives.

The text box below describes how this approach is currently paving the way towards large scale deployment and operations of Data Spaces in a multitude of vertical sectors.

Data Spaces: context

To pursue its strategy towards the "Common European Data Spaces", the EC steers the definition of the regulatory and legislative foundation, developing the reference architectures and (open-source) building blocks, and supporting the deployment of a multitude of sectoral Data Spaces, as depicted in Figure 2.

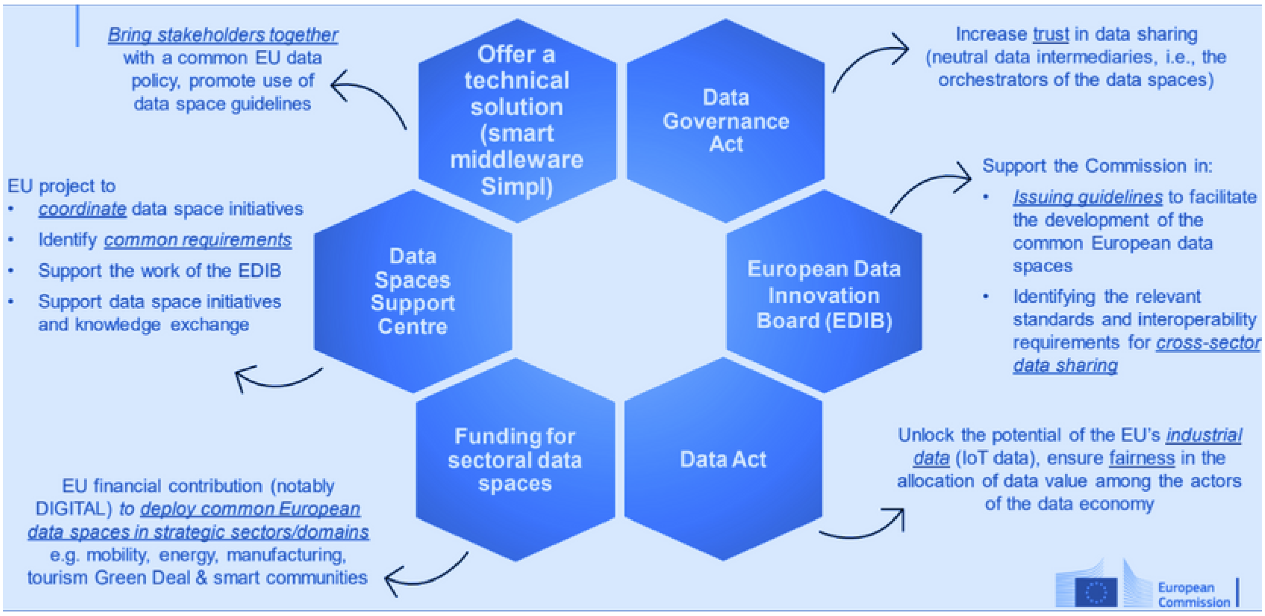

Bring stakeholders together with a common EU data policy, promote use of data space guidelines
Offer a technical solution (smart middleware Simpl)
Data Governance Act
Increase trust in data sharing (neutral data intermediaries, i.e., the orchestrators of the data spaces)
EU project to
coordinate data space initiatives
Identify common requirements
Support the work of the EDIB
Support data space initiatives and knowledge exchange
Data Spaces Support Centre
European Data Innovation Board (EDIB)
Support the Commission in:
Issuing guidelines to facilitate the development of the common European data spaces
Identifying the relevant standards and interoperability requirements for cross-sector data sharing
Funding for sectoral data spaces
Data Act
Unlock the potential of the EU's industrial data (IoT data), ensure fairness in the allocation of data value among the actors of the data economy
EU financial contribution (notably DIGITAL) to deploy common European data spaces in strategic sectors/domains e.g. mobility, energy, manufacturing, tourism Green Deal & smart communities
European Commission

Figure 2: EC role in supporting the creation of Data Spaces [8].

Various regulations have been developed that support the "common European Data Spaces", including the Data Act [9] and the Data Governance Act (DGA) [10]. Additionally, the Digital Services Act [11], the Digital Markets Act [12] and the Artificial Intelligence Act [13] touch on topics related to Data Spaces. Together, these regulations aim at creating a level playing field for sharing data under the core European values of sovereignty, trust and privacy. In addition, various European initiatives are exploring and developing reference architectures for Data Spaces, including the International Data Spaces Association (IDSA) initiative [14], the Gaia-X initiative [15][16], the FIWARE initiative [17], the iSHARE initiative [18][19] and the Data Space Business Alliance (DSBA) initiative [20][21].

Currently, the EU Data Spaces Support Centre (DSSC) initiative [3] is the leading initiative building upon these reference architectures and working towards a blueprint for the emerging (federation of) sectoral Data Spaces in Europe. The EU SIMPL procurement project [7] will adopt the DSSC blueprint and develop the associated building blocks and make these available open-source. This relation between the sectoral Data Space initiatives, the DSSC initiatives and the SIMPL initiative is depicted in Figure 3. (Note that there are more Data Spaces in development than identified in the picture.)

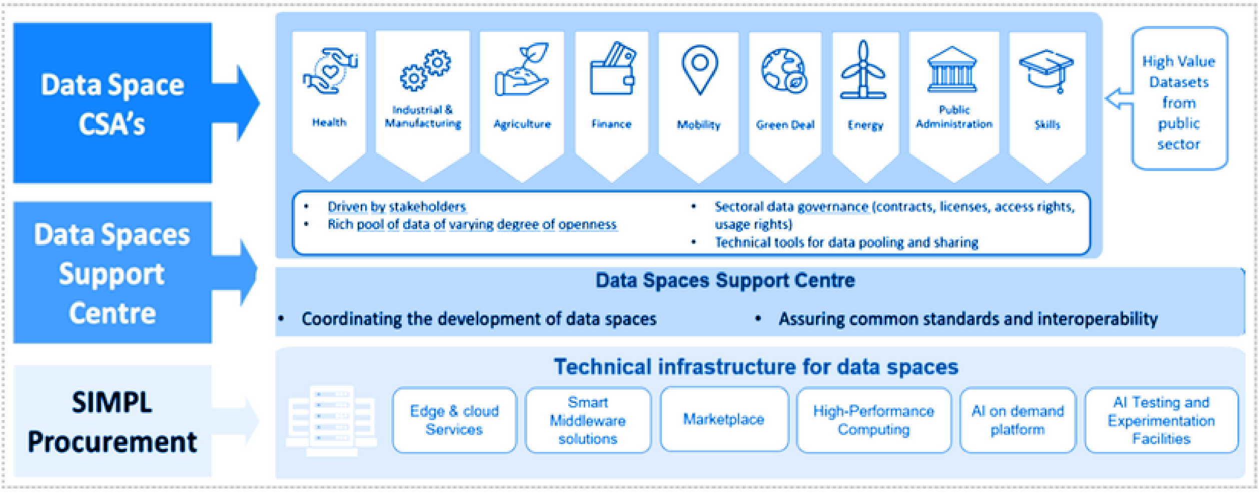

Data Space CSA's
Data Spaces Support Centre
SIMPL Procurement
Health
Industrial & Manufacturing
Agriculture
Finance
Mobility
Green Deal
Energy
Public Administration
Skills
High Value Datasets from public sector
Driven by stakeholders
Rich pool of data of varying degree of openness
Sectoral data governance (contracts, licenses, access rights, usage rights)
Technical tools for data pooling and sharing
Data Spaces Support Centre
Coordinating the development of data spaces
Assuring common standards and interoperability
Technical infrastructure for data spaces
Edge & cloud Services
Smart Middleware solutions
Marketplace
High-Performance Computing
AI on demand platform
AI Testing and Experimentation Facilities

Figure 3: EC approach for realizing the common European Data Spaces: Visualization of the various development initiatives.

A multitude of Data Spaces is currently already emerging, building upon the various reference architecture initiatives as described above and in anticipation of the results of the DSSC blueprint and the associated SIMPL building blocks for Data Spaces. The left side of Figure 4 provides a recent overview (a "radar") on Data Space initiatives, categorized by both sector and maturity level, whereas the right side provides the visualization of the data categories currently supported by the German Mobility Data Space, as an example Data Space.

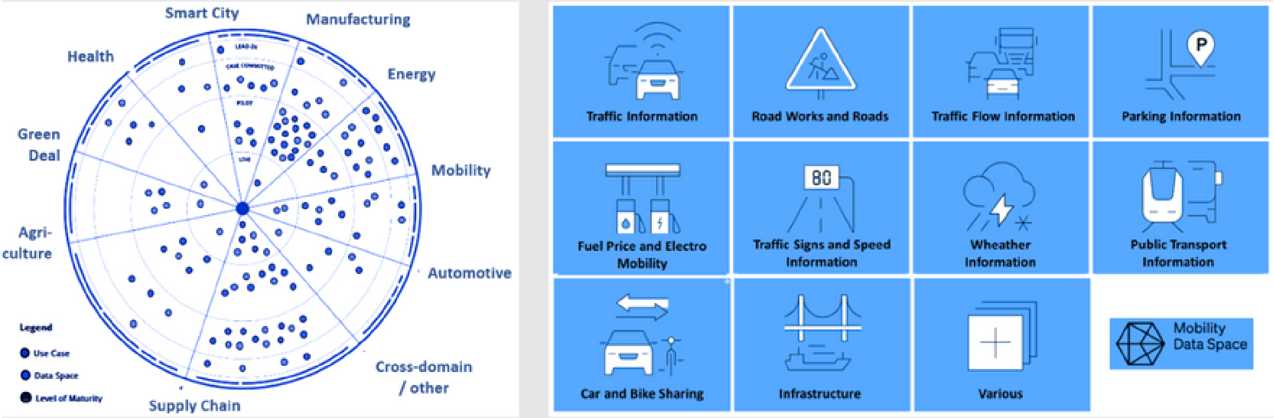


Figure 4: The IDSA Data Space Radar (left, screenshot November 2023, [22]) and the (illustrative) data service offering of the German Mobility Data Space (right, screenshot November 2023, [23]).

Being able to seamlessly share data over the multitude of emerging Data Spaces yields clear advantages. It extends the reach and scope of accessible data and allows new business models and solutions to be developed across sectors and regions. For Data Spaces to seamlessly interconnect in such a federation of interoperable Data Spaces, an interoperability framework is needed. Such a framework provides a basis to manage, coordinate and control data sharing between participants in the federation of Data Spaces, with specific focus on interoperability capabilities with respect to data sovereignty, trust and discoverability. The new European Interoperability Framework (EIF), developed by the European Commission [24], gives a systematic approach for addressing the interoperability challenges. It shows that Data Space interoperability is more than only the interoperability of its technical components. It distinguishes four interoperability levels (technical, semantic, organizational and legal interoperability) under an overarching integrated governance approach.

III.2. Privacy Enhancing Technologies

Privacy Enhancing Technologies (PETs) are “a coherent system of ICT measures that protects privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data, all without losing the functionality of the information system” [25]. Many ‘classical’ PETs, like data access control, separation of data, anonymization and others, are already available in Data Spaces. In this paper, we concentrate on a relatively recent subset of PETs also known as “Privacy-preserving computation”: a collection of digital technologies enabling secure processing, analyzing and sharing of sensitive data (of one or more parties) while protecting the confidentiality of this data [26].

The significance of these more recent PETs has grown over the last few years due to the EU developments and decisions about the protection of personal data, and the establishment of privacy regulations (such as the GDPR) as well as the increasing market need for data sovereignty. Therefore, the attention given to PETs is increasing as well, as indicated by recent reports from the OECD [26], the Royal Society [27], and the United Nations [28] on the role and PETs positioning. As these reports indicate, PETs are rapidly developing in various directions, enhancing their applicability in many different domains. PETs encompass a variety of solutions and approaches. The OECD report [26] provides a functionality-focused classification of PET approaches, distinguishing between the categories of data obfuscation tools, encrypted data processing tools and federated and distributed analytics tools. These factors increasingly play a significant role in the paradigm of "privacy by design" or "data protection by design" [26].

In practice, PETs are a complex umbrella of technologies. Categorizing these diverse techniques has been attempted several times [29][30]. In the functionality-focused categorization by the OECD [26], the following PET-categories are identified:

1 Data obfuscation tools

This type of tools concerns PETs that conceal data from external world by processing the data locally and altering the data by adding "noise", swapping the attributes and values or by simply removing identifying details. Since a long time methods exist for anonymization and pseudonymization, but these have limitations in terms of privacy and utility. The modern even more elaborated solutions represent technologies for guaranteeing that non-authorized persons and organizations do not see the confidential data. Synthetic data generation consists of methods to generate a completely new dataset that resembles the original data but does not contain sensitive information. Differential privacy techniques add noise to the original data to hide sensitive details and randomize the actual data, yet keeping the necessary level of the details. Zero Knowledge Proofs are techniques to prove statements based on the data between prover and verifier organizations without having to reveal data itself.

2 Encrypted data processing tools

New technologies offer possibilities to analyze (multi-organizational distributed) data sets while the data in exchange is encrypted. This includes both encryption methods and computational solutions. Homomorphic Encryption (HE) is used as a form of encryption that enables calculations on encrypted data without decrypting it. Secure Multi-Party Computations (MPC), a variety of technologies based on, for example, secret sharing, HE or garbled circuits, enable joint computations between multiple parties without them seeing each other's data. Trusted Execution Environments offer calculations in a secure area of a processing element, allow code and data to be isolated and protected from the rest of the system.

3 Federated and distributed analytics

Distributed analytics in general allows to arrange the analysis of data and models distributed between different places. Federated learning (FL) is a decentralized and privacy-friendly form of machine learning. Instead of bringing the data to the machine learning model, FL leaves data at the premises where it resides, bring the local personalized model to perform analysis on data premises, and shares periodically only the results, which is a much smaller amount of data than transferring the whole set of data, with multiple parties. The models are locally trained, aggregated, and this process is iterated to train the global model. This is an example of 'data visiting' [31].

A fourth category mentioned in [26] consists of data accountability tools. These tools are not always considered PETs, since their focus is to enhance privacy and data protection by enabling data subjects' control over their data, and to set and enforce rules stating how data can be accessed, by whom, what, why, when and at which condition. Data accountability is typically a requirement for Data Spaces, and this is one of the areas where Data Spaces and PETs may strengthen each other on data sharing scenarios.

For most PET categories there are tradeoffs between privacy/security, utility, computational performance, communication overhead. Choosing the most beneficial combinations of PETs depends on the application. This aspect can be an impediment for the automated deployment of PETs in new situations: an analysis of the information and privacy structure is needed to design a solution that guarantees (or minimizes) the absence of information leakage.

While there is still a lot of academic research on PETs, the last decade has seen more and more companies offering commercial solutions based on PETs. There has also been a significant increase in standards-related activity relevant to PETs in the last few years [28].

Standardization of Application Programming Interfaces (APIs) will make it easier to integrate PETs in Data Spaces, however the time of agreed-upon PETs standards is yet to come. For instance, in the PET community interoperability is not yet a priority and most of the proposed solutions are custom-made, i.e., not generally applicable to other scenarios. The situation is going to get better as the integration of PETS into Data spaces will become a priority in the Research & Innovation ecosystem.

Privacy-enhancing technologies: a use case

Especially for sensitive data, it is worth observing that the objective of the involved parties is not to share data but only to share answers to specific questions: for instance, do not share nationality, citizenship and date of birth, but only the outcome of the algorithm running on that data; for instance, that a specific person is eligible for some service or benefit.

An example for a PET use case comes from the mobility sector. The sharing and correlation of data originating from different actors within the ecosystem, both public and private, allows to improve internal processes within the organizations, as well as at a global level, in order to provide more intelligent and sustainable mobility services. However, guaranteeing the data privacy is a key factor in stimulating data sharing, especially when considering valuable business data or user data protected by the GDPR. The data is sensitive from a personal point of view (privacy) but also from a business point of view (competition).

For instance, a public transport operator wants to cross-reference its public transport usage data with data provisioned by a private telecommunication operator in order to perform advanced data analytics, to better understand mobility patterns around a city. This requires data about the usage of public transport as well as highly-sensitive and protected personal data collected by the telecom operator. The data originating from both sources needs to be forwarded to a third-party data analytics service provider to perform this analysis and return actionable insights to the public transport operator.

Aside from interoperability issues that may arise from the data sharing scenario above, ensuring that the personal data of users is not exposed to third-parties is a crucial aspect for a successful pick up of this scenario in a real-world application with commercial impact. In this case, various data obfuscation technologies are needed to guarantee the privacy of the shared data. For instance, data anonymization or pseudonymization can be used to protect sensitive data, and more complex techniques such as differential privacy can also be applied to provide further protection. However, applying these PETs adds another layer of complexity to the data sharing scenario, considering the fact that these technologies are usually developed by specialized third-parties that also need access to the original data.

A Data Space that is designed to handle mobility data in a privacy-preserving way would be a clear benefit in this situation.

III.3. Combining Data Space and PETs: terminology considerations

Despite these high-level common goals, it is to be noted that the Data Space and PET initiatives, and their related communities, have their own view and terminology on how data sovereignty, trust and protection of sensitive data are to be realized, which in some cases seem not to be aligned. Exemplary is the term "trust". Trust has a positive connotation in the Data Spaces community as a basis for sharing data between organizations. For Data Spaces, establishing and maintaining trust among the participants is considered as important value add of the Data Space concept as it allows to establish a trust framework between participants by means of a combination of technical and non-technical mechanisms [5] as basis for sharing potentially sensitive data on a peer-to-peer basis between the Data Space participants. However, trust has a negative connotation in the PET community. In the development of PETs, the goal is to ensure security properties by means of strong mathematical and technical guarantees, removing the need to trust other participants. In PETs, if trust in other participants is necessary, this is seen as a vulnerability.

In addition, the concept of "data owner" or "data entitled party" is used in the Data Spaces community to denote a formal legal participant that holds rights on deciding whether, with whom and under what conditions its data may be shared with other participants. However, since data can be copied, transformed, aggregated etc., data cannot always be owned in the same way as physical objects can be owned. The GDPR, for example, defines various roles (data subject, controller, processor) and stipulates their rights and obligations, but does not define data ownership. Hence, from a data sovereignty and protection perspective, data ownership or entitlement can be misleading concepts. An important step in integrating Data Spaces with PETs is to address these discrepancies in terminology, so as to ensure consistency and avoid misunderstandings. The Terminology mappings done by DSBA, IDSA, and JRC (the EU Joint Research Centre) need to be taken into account when moving forward.

State-of-the-Art of the alignment between Data Spaces and PETs use case

This white paper builds upon the insights and results gained from multiple previous projects. The EU TRUSTS project has developed a Data Space environment [32] based on the reference architecture of IDSA. It specifically describes high-level architecture requirements related to PETs, being able to run computations on distributed and controlled environments, including various options for distributing the PET algorithms over the infrastructure of the PET service operator and the data providers, applicable to various types of PETs, including MPC, HE and FL. In [33] it is argued that PETs can facilitate trustworthy data sharing in the context of Data Spaces and several topics and challenges related to PETs are discussed, including legislation such as the DGA, and the need to make clearer in the DGA how PETs could increase the level of trust and control of data holders over their personal data. On the technical aspects of PETs, it describes the need to address its current performance bottlenecks to enable more "widespread application of privacy-preserving analytics for data sharing spaces and beyond", as well as the need for quantum-secure systems. The Platoon project [34] has adopted IDSA Data Space connectors to support multi-party data exchange in energy sector, with usage control capabilities for requesting personal data by means of the MyData operator concept [35] as provided by the CaPe platform [36]. Moreover, the need to align the PET and Data Spaces concepts has also been noted by the Spanish Data Protection Authority (AEPD [37]). The EU EnerShare project [38] builds a service to support an FL solution for training models on privacy sensitive personal energy consumption data locally on stakeholders' premises while preserving privacy over collected training data. The proposed model uses anonymization of privacy sensitive data by means of the differential privacy PET concept. The implementation is based on integration with the energy Data Spaces approach by means of an IDS-based Data Space connector from which the user can start FL platform and models. Also the EU SIFIS-Home project [39] has leveraged on FL but its focus has been on preserving the privacy of individuals in smart home environments making use of several PETs techniques, also proposing an open-source interoperable secure stack and introducing the concept of privacy gain and its trade-off with accuracy in an exemplary use case of a GDPR-compliant face recognition system, suitable for several Internet-of-Things applications.

These initiatives have shown how the combination of Data Spaces and PETs involves not only aspects of compliancy from a legislative point of view, but also the need to ensure semantic and technological interoperability: semantic formalization of datasets, interoperability supported by open API, semantic formalization of privacy rules.

IV. BENEFITS FOR ALIGNMENT: THREE PERSPECTIVES

The potential benefits for alignment in development and deployment of the Data Space and PETconcepts and architectures are addressed from three perspectives in the subsequent section of this chapter: functional, business and process perspectives.

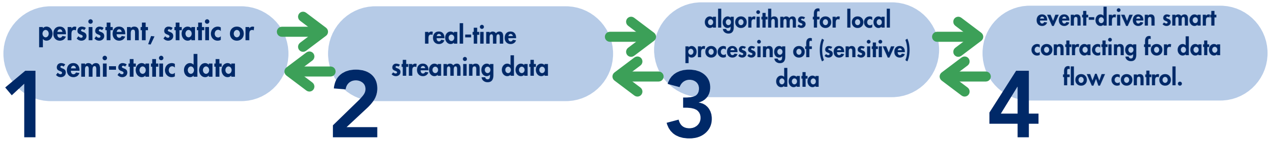
IV.1. Functional perspective

Functionally, there are various benefits in the combination of Data Spaces and PETs. The functional benefits can be addressed both from the Data Spaces and from the PETs view.

01

Functional benefits from the Data Space view

Organizations may have multiple types of data to share in different contexts. As such, privacy sensitive data for which PETs are required is a specific type of data sharing among a multitude of other types of data sharing techniques that may be relevant to an organization. For instance, [2] distinguishes between four types of data sharing that may apply to organizations in the mobility and logistics sector:


1
persistent, static or semi-static data
2
real-time streaming data
3
algorithms for local processing of (sensitive) data
4
event-driven smart contracting for data flow control.

The first two data sharing types are considered "generic and traditional", involving the sharing of potentially sensitive data between Data Space participants. The support of PETs applies to the third type of data sharing in which algorithms are shared with the data provider for local processing of sensitive data. Including PETs as an integrated type of data sharing will enable a considerably improved security level for highly sensitive data. In some cases, such a choice may even be mandatory because regulations require their usage; PETs help both data provider and providers of Data Spaces to comply with the GDPR. Nevertheless, as each type of data sharing may be relevant for a data provider, they need to be supported simultaneously within a specific Data Space.

From the Data Space view, this provides major opportunities to capitalize on both economies of scope and scale, as the value of a Data Space increases with the participation of more stakeholders. By fostering an ecosystem that connects these diverse stakeholders and promotes data sharing across various Data Spaces, there is potential to pave the way for innovative solutions and novel business models.

In addition, the data provider needs a single entry point to simultaneously manage and control the provisioning of these multiple types of data. It can prevent the data provider from the threat of vendor lock-in by service providers or environments supporting only a single type of data sharing. Also, major integration efforts in developing, enforcing and managing data sovereignty, trust and security solutions across multiple data sharing environments can be avoided. A single entry point for the data provider yields clear operational benefits compared to siloed approaches to user-friendliness, complexity, efficiency and costs. Providing PET capabilities as part of the multi-service data sharing options in converged Data Spaces enables these potential benefits.

The joint implementation of Data Space and PET technologies can support automated law and policy enforcement; to ensure that data is only handled according to lawful (for instance, GDPR) and agreed principles. PETs enable specific agreed calculations to be executed, so an additional analysis that would violate the goal limitation[2] of the shared data can be prevented. In this way, PETs can provide technical enforcement of agreements that are made within a Data Space.

Accountability on data sharing or data processing transactions is an important capability that can be supported in Data Spaces. In cases, where PETs have to be used for processing privacy sensitive data, specific additional requirements may apply. For instance, it may be required to have the accountability mathematically proven. Some PETs have accountability already ‘built in’ by-design. This implies that the capability for accountability in Data Spaces has to be re-evaluated in cases where PETs are to be supported. For the Data Spaces, this may imply that the same accountability features can be used for other types of data sharing as well.

[2] Data can only be used for the goal with which it was collected

02

Functional benefits from the PET view

Currently, PET implementations are typically tailor made, case-specific, solutions. The required PET capabilities are not yet commercially available as a re-usable service. They need to be procured from a specific solution supplier. For specific isolated situations this may work. However, when an organization needs to work with data from many other partners, it becomes a challenge when the data sharing or data processing solution for each case and partner is different. This relates to the need for a single entry point for data providers as described above.

When a PET is considered as a standard service, which is supported in Data Spaces, benefits may result both for their adoption (in terms of ease of deployment) and their scalability (in terms of reach of potential participants). For instance, by aligning and reusing the capabilities as provided by the emerging federation of Data Spaces, the adoption and scalability of PETs may benefit from various Data Space capabilities, including:

01 **Discoverability capabilities**
e.g., for the discoverability of PET algorithms by including them in the Data Space app store,

02 **Trust capabilities**
e.g., for identity management and consent management, for defining and enforcing authorization policies, for certification of participants and components, and for defining the legal agreements and conditions.

03 **Data interoperability capabilities**
e.g., for defining, managing and deploying semantic data model mapping and conversion features to make the data sources more accessible.

IV.2. Business perspective

The business opportunities and benefits for an aligned approach to Data Spaces and PETs can be considered from the perspective of various stakeholders.

From the ***perspective of data providers***, a single point-of-contact and ease-of-onboarding to support multiple types of data sharing in various data sharing communities or Data Spaces will provide major advantages in terms of complexity of integration, cost-efficiency and prevention from vendor lock-in. This has previously been identified as a functional benefit as well. Harmonization of governance processes, reference architectures, building blocks to support multiple types of data sharing (including PETs) in federation of Data Spaces provide the means to do so.

From the ***perspective of enabling Data Space and PET infrastructure providers***, major opportunities exist to capitalize on both economies of scope and scale, as the value of both Data Space and PET initiative increases with the participation of more stakeholders. By fostering an infrastructure and an ecosystem that combine these initiatives, connect the diverse stakeholders and promote interoperability, there is potential to pave the way for innovative solutions and novel business models. However, the market will not naturally forge such links, hence the need for proactivity in developing the required alignment. An improved market value proposition as a one-stop and single-point-of-entry for multiple types of data sharing will result.

From the ***perspective of PET service providers***, piggy-backing on the anticipated fast roll-out of Data Spaces as part of the EU data strategy allows them to exploit economies of scale. They can more easily extend the reach of their PET service offering across the emerging federation of Data Spaces. Not only does this allow them to reuse the discoverability, trust and data interoperability capabilities of Data Spaces, it may also give the opportunity to reuse the supporting structures for clearing, transaction logging, billing, reporting, traceability and auditability.

IV.3. Process perspective

Although Data Spaces and PETs seem to have their own architectural concepts for data sovereignty, trust and privacy, when making data available to sharing or processing across organizations, there are also commonalities in required management processes. This can be observed from the five main management processes in the governance framework for PETs [40],[3] which can be categorized into operations processes and orchestration processes:

Operations processes

- the on-boarding process for data providers joining a PET compute group
- the off-boarding process for data providers joining a PET compute group
- the access process for beneficiaries requesting to access insight in the PET results

Orchestration processes

- the query process for data scientists requesting to run PET queries
- the change process for data scientists requesting to add new use cases

These five main PET management processes have analogous counterparts in a Data Space. Moreover, these processes may be composed of activities, which are already provided by associated capabilities and building blocks in the emerging federation of Data Spaces. This specifically applies to the operations processes, which are fundamental capabilities provided as part of the Data Space concept and architecture, but are mainly supportive for the PET environment. On the other hand, the orchestration processes are key for the PET environment but beyond the scope of the Data Space concept.

[3] In [40], these five main management processes have been identified for the specific situation of using MPC for an elderly care-taking care case. As they appear to be generically applicable, they are generalized for the broader use of PETs cases in this white paper.

Hence, complementarity in management processes can be observed with clear separation of concerns between the Data Space and the PET environment. Moreover, the deployment of individual PET implementations by means of its orchestration processes may benefit from operations processes as being provided by the Data Space implementations. The common ground in these operations processes is provided by the basic Data Space capabilities for discoverability, trust and data interoperability. In this manner, alignment with Data Spaces can allow PET implementations to focus on their specific added value and service offerings instead of having to take into account the basic and generic capabilities for data sharing and thereby pave the way for adoption and scalability, which is currently lacking.

The associated capabilities and building blocks for discoverability, trust and data interoperability as foundation for common Data Space and PET operations processes may be implemented by agreed-upon and well-defined standards in the Data Space reference architectures as part of the Common Technical Grounding in the DSSC blueprint, see Figure 1. At the same time, it has to be realized that for exploiting these potential benefits of joint Data Space and PET operational processes, the alignment has to be technically defined and organizationally embedded in a corresponding strategy and governance model on business roles and responsibilities.

V. ALIGNMENT APPROACH: EXPLOITING COMMONALITIES

This chapter addresses aspects for alignment between the Data Spaces and the PET concepts and architectures in terms of various commonalities to be jointly exploited. The following sections address the commonalities in Privacy Patterns, in technical grounding and in strategy.

V.1. Common Privacy Patterns: connectivity and interactions

At its current stage, the alignment and integration between Data Spaces and PETs is already feasible, but it has only happened on a small scale, involving technologies like anonymization, synthetic data generation, differential privacy or simple HE. Extending this scheme to more complex PETs faces the challenges of workflow management and orchestration of interactions. The various types of PETs may require only a limited set of connectivity and interaction patterns between the stakeholders, which may be the various data providers involved in the PET and the PET service provider. Such connectivity and interaction patterns jointly constitute a Privacy Pattern [41][30]. A limited set of such Privacy Patterns need to be supported between the stakeholders. A method for modelling the Privacy Patterns and their associated connectivity and interactions is described in the text box.

Privacy patterns: modelling

Data Spaces involve several situations where interoperability is addressed. As explained in the Open DEI report on reference architectures and interoperability in digital platforms [42], the construction of interoperability involves three artefacts: an interoperability point ,i.e., a location in the overall system where data is exchanged according to an agreed interoperability specification, an interoperability case, i.e., a documented justification and agreement on an interoperability point, and an interoperability profile, i.e., a documentation of requirements allowing implementation of a conformant system.

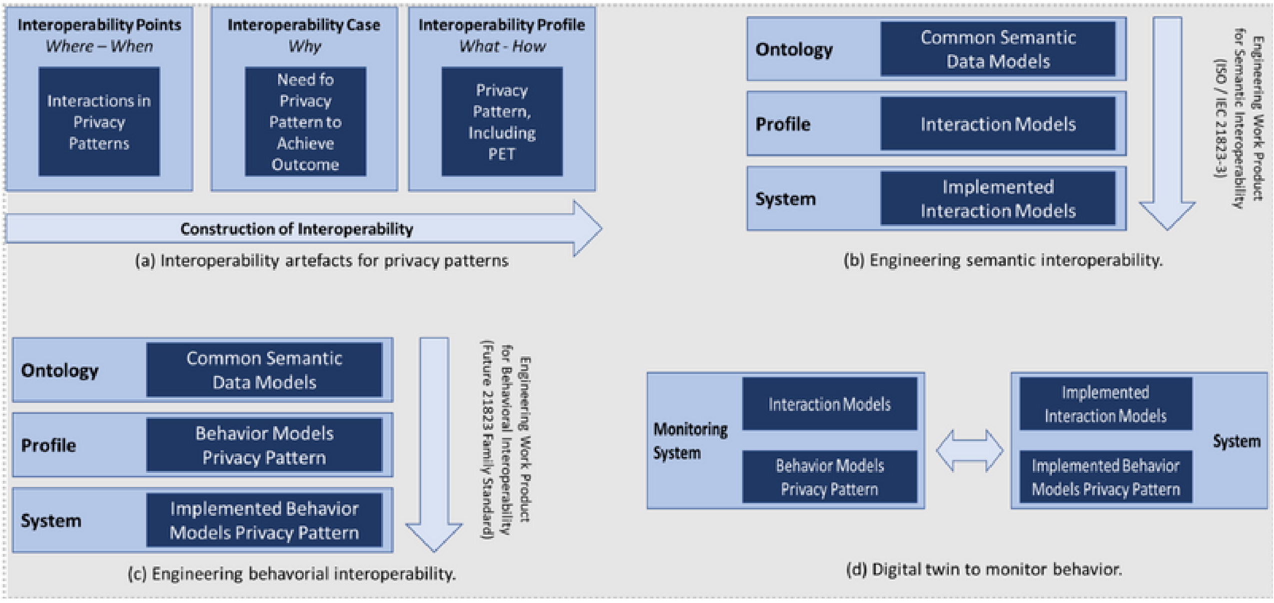

Interoperability Points
Where – When
Interactions in Privacy Patterns
Interoperability Case
Why
Need fo Privacy Pattern to Achieve Outcome
Interoperability Profile
What - How
Privacy Pattern, Including PET
Construction of Interoperability
(a) Interoperability artefacts for privacy patterns
Ontology
Common Semantic Data Models
Profile
Interaction Models
System
Implemented Interaction Models
Engineering Work Product for Semantic Interoperability (ISO / IEC 21823-3)
(b) Engineering semantic interoperability.
Ontology
Common Semantic Data Models
Profile
Behavior Models Privacy Pattern
System
Implemented Behavior Models Privacy Pattern
Engineering Work Product for Behavioral Interoperability (Future 21823 Family Standard)
(c) Engineering behavorial interoperability.
Monitoring System
Interaction Models
Behavior Models Privacy Pattern
Implemented Interaction Models
Implemented Behavior Models Privacy Pattern
System
(d) Digital twin to monitor behavior.

Figure 5: Developing common Privacy Patterns

Applied to Data Spaces and PETs, the construction of interoperability is shown in Figure 5 (a): the interoperability point refers to interactions that take place according to a Privacy Pattern. A Privacy Pattern is a reusable solution to a commonly occurring privacy problem [43][44], the interoperability case is a documentation of the behavior of a Privacy Pattern, and the interoperability profile is a documentation of requirements allowing the implementation of the Privacy Pattern.

Figure 5 (b) shows the engineering work products that enable semantic interoperability as described in ISO/IEC 21823-3 (Part 3: Semantic interoperability)[4]: ontology engineering enables the provision of common semantic data models, profile engineering enables the specification of interaction models, system engineering enables the implementation of interaction models.

Figure 5 (c) the engineering work products that enable behavioral interoperability for privacy. Note that behavioral interoperability is currently studied as a preliminary work item in ISO/IEC JTC 1/SC 41 (Internet of things and digital Twins)[5]: ontology engineering enables the provision of common semantic data models, profile engineering enables the specification of a behavior model described by a Privacy Pattern, and system engineering enables the implementation of a behavior model or Privacy Pattern.

The support of Privacy Pattern and its integration of a behavior interoperability standard can pave the way to interoperable data access and data usage enforcement. Figure 5 (d) shows an example of a digital twin providing privacy assurance. The digital twin consists of the following twins: the monitoring system and the system. The monitoring system has access to interaction models and Privacy Patterns. The system implements the interaction models and Privacy Patterns. The monitoring system can report on the compliant behavior of the system.

[4] https://www.iso.org/standard/83752.html
[5] https://www.iso.org/committee/6483279.html It is expected that this Work Item will produce a new standard.

A good example of pattern for de-identification scheme based on pseudonyms that was proposed for secure vehicular communication [45]. An explanation of this model was provided in the TAM project on Trust Autonomous Mobility [46][47].

For the support Privacy Patterns in an aligned Data Space and PET architecture, two additional capabilities need to be supported:

A workflow management entity

To manage the flow/routes of data between the various modules (data apps) involved in the implementation of the Privacy Pattern at its various stakeholders, including the possible sequencing of data packets exchanges between the data apps and the starting and stopping of data apps. The workflow management capability makes the configuration of the PET Privacy Patterns transparent and easier to adapt as the flow of data processing can be made explicit using a script. A complete workflow or set of workflows of data apps on one or multiple connectors can be seen as ***"Data Analytics"***.

A PET-specific data plane

As part of the emerging DSSCs Data Space blueprint architecture that distinguishes between a control plane and a data plane [46]. The control plane is responsible for deciding how data is managed, routed and processed. The data plane is responsible for the actual moving of data. As such, the data plane handles the actual exchange of data. For recurring PET Privacy Patterns, a PET-specific data plane can be developed to enable ease-of-use in configuration and onboarding of PET use cases within a Data Space environment.

V.2. Common (technical) ground: discoverability, trust and data interoperability

As described in the section on Data Space in Chapter III and depicted in Figure 1, the emerging federation of Data Spaces will be realized by means of a *"Common Technical Ground"*, with building blocks for data sovereignty, trust and discoverability and three main categories of technical services:

01

Data Space connectors

Serving as secure gateways, enabling systems and organizations to access a Data Space securely.

02

Federated services

Offering various functionalities, such as validation or cataloguing of services.

03

Data Space registries

Registering the participants of a Data Space.

The three main categories of technical services in the Common Technical Ground typically operate within the control plane of the DSSCs blueprint [46]. As described in Chapter IV, the operationalization of individual PET implementations may benefit from this common ground by being unburdened from having to implement existing and re-usable basic capabilities for discoverability, trust and data interoperability as part of the PET governance processes.

Alignment with Data Spaces on these capabilities can allow PET implementations to focus on their specific added value and service offerings instead of having to take into account the basic and generic capabilities for data sharing, thus paving the way towards their large adoption. The common capabilities contain trust mechanisms such as identity verification, data contracts and usage policy definition and enforcement [49]. They facilitate the extension of Data Spaces components and the introduction of new ones to allow multi-source processing and orchestration between a multitude of stakeholders as required for supporting PETs.
Building upon the common technical ground in an aligned Data Space and PET architecture leads to a role model of stakeholders as described in the text box.

Common technical ground for Data Spaces and PETs: role model

An aligned Data Space and PET (reference) architecture is based on the role model of stakeholders and the building blocks (capabilities) they provide. A role corresponds to a primary activity in the overarching processes of providing the PETs access to sensitive data sharing and sharing the PET algorithms between the stakeholders.

A role model for supporting the technical grounding in the aligned Data Space and PET (reference) architecture can be derived from the role models as provided in [48]. It is depicted in Figure 6: PETs are provided access to sensitive data, and PET algorithms (in the form of "data apps") are shared between the stakeholders.

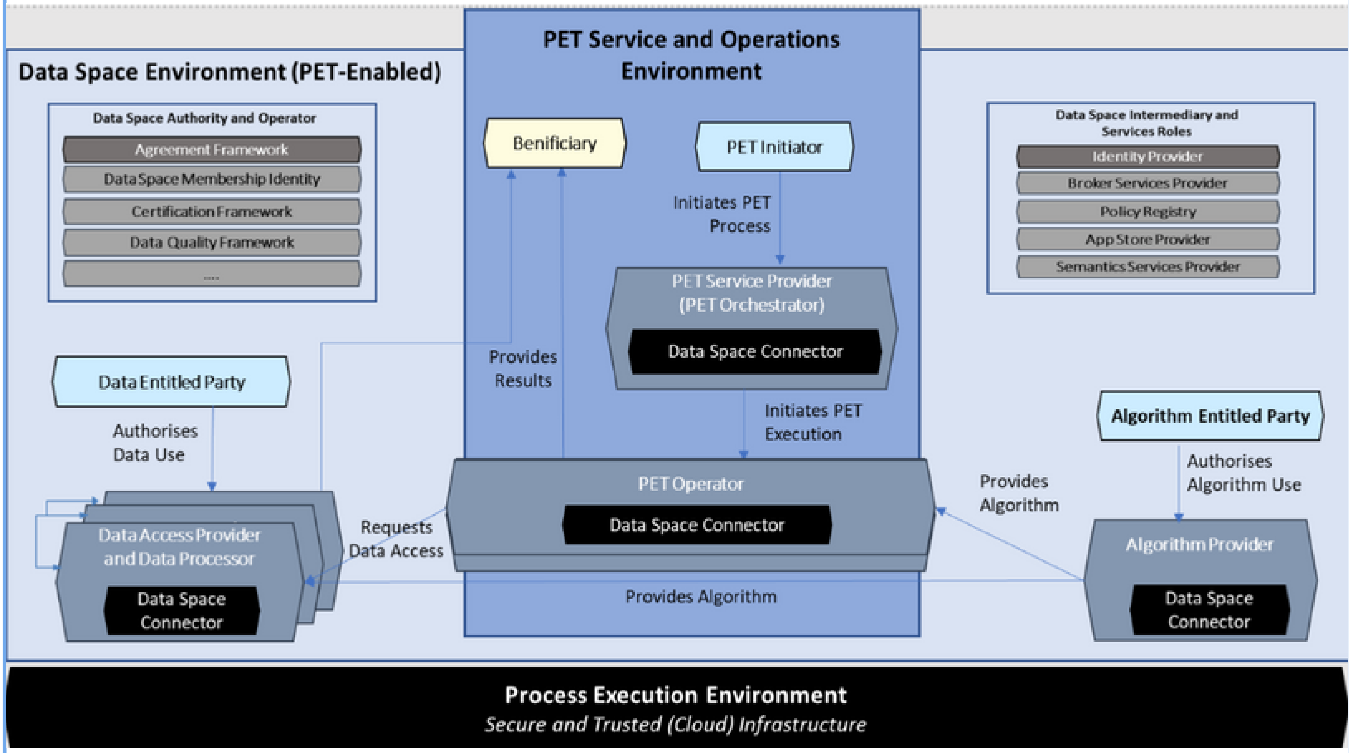


Figure 6: The role model for interoperability in an aligned Data Space and PET (reference) architecture.

The roles in the "Data Space Environment" are the generically applicable roles for many types of Data Spaces. They are derived from and described in [50][51]. It is to be noted that the role which is generically referred to "Data Provider" has been adapted into the role of "Data Access Provider and Data Processor" as, for the case of supporting PETs, this role doesn't provide or share sensitive data with other participants, but rather provides access to data for local processing by decentralized PET-algorithms.

In addition, within the "PET Service and Operations Environment" the following PET-specific roles have been identified:

- The ***PET Initiator*** is responsible for initiating a PET interaction via the PET Orchestrator for the benefit of the PET Beneficiary.

- The ***PET Beneficiary*** is interested in a result of the PET-supported interaction. The Beneficiary receives the results either directly from a Data Access Provider and Data Processor or from the PET Operator.

- The ***PET Service Provider*** (also referred to as PET Orchestrator) is the single-point-of-contact to both the PET Initiator and the PET Beneficiary. It orchestrates the interactions with all core business roles and the services they provide and ensures that the PET-algorithm yields the intended results for the PET Beneficiary. Furthermore, it manages the applicable policies for what it orchestrates, e.g. the usage policies on data sources and algorithms.

- The ***PET Operator*** is responsible for the execution of PET-processing on sensitive data according to the orchestration of data and PET algorithms as set out by the PET Orchestrator. As such it manages the workflows of execution tasks of PET algorithms and data sharing. This includes both the local processing by decentralized PET-algorithms on sensitive data in the domain of Data Access Providers and Data Processors and centralized parts PET-algorithms in its own domain. Multiple PET Operators may have to collaborate in jointly providing the result. The PET Operator may provide the results of the PET process to the PET Beneficiary on behalf of the PET Service Provider.

- The ***Process Execution Environment*** provides a secure and trusted environment for execution of workloads, e.g., for processing PET algorithm on sensitive data. It provides a capability (building block) that is referred to as the "Application Container Environment (ACE)" in which the security gateway and the PET-supported algorithms are executed with the required data in order to produce the intended results of the algorithm. The Process Execution Environment may be secure and trusted cloud environment to be used by various roles. Alternatively, the individual roles may choose to use their own trusted processing environment for secure execution of PET algorithm processes on sensitive data.

Each role in the figure can be assigned to one of the categories proposed by the IDSA role model structure [14]. The ***Data Space core roles*** encompass the core participants who are involved and required every time data is exchanged (such as data providers and data consumers). The ***Data Space intermediary roles*** encompass trusted intermediary entities that are commonly considered as "platforms" and assume a rather central role compared to the great number of core participants. The building blocks for implementing the Data Space intermediary roles are currently under discussion as part of the DSSC blueprint development [6]. ***The Data Space governance roles*** have the authority and the task of setting and enforcing guidelines to standardize data exchange, to create trust, and enable sustainable operation of the Data Space.

Depending on the use case and the type of PET that needs, the role of a data app / algorithm provider is foreseen as the figure shows. In the applicable case, the PET data app / algorithm is provided by the data app provider so that data providers and consumers involved can utilize it. Crucially, Data Spaces also provide certification and accreditation mechanisms for the PET data apps / algorithms offered to their participants, thus guaranteeing the quality and trustworthiness of the PETs [51]. From the Data Spaces perspective, existing reference architectures such as the IDS-RAM 4.0 [14] already take into consideration the availability of a service marketplace where third-party service providers can publish their services for use by other participants. The publishing process of these apps also includes a certification scheme to verify the quality and trustworthiness of the published services [52].

To address the challenge on making the complex multi-organizational cooperative data analysis more transparent and traceable, there should be a methodology elaborated on how to make such analysis completely auditable, so that different stakeholders can see the status of the analysis and operations on data they own.

As part of the development towards an aligned Data Space and PET (reference) architecture and role model, standardization is required for managing the registration and distribution / sharing of PET algorithms (e.g., as data apps). Extra attention should be spent on the bridge between data providers and the PET algorithms providers, specifically on the automation of matching the needs of providers with the controlled access rules on data.

V.3. Common operations model: PETaaS and Data Spaces

Another perspective to look at in the alignment of PETs and Data Spaces is the provisioning and operational model of PET services. In this sense, and taking into account the reference architectures (IDSA) described previously in this document, we propose in this section various scenarios of the integration of PETs as a service (PETaaS) within a Data Space to form an integrated and intertwined offering. The scenarios also describe an operational model for how these PETs can be executed within the Data Space, based on current capabilities as well as proposing new ones for more complex scenarios.

These scenarios are further elaborated in the text box, which also includes a number of figures depicting the integration scenarios and their necessary components.

PETaaS and Data Spaces: integration scenarios

Figure 7 depicts various scenarios about PETaaS within a Data Space and shows the interactions between three main stakeholders: the PET provider, the Data Providers, and the Data Consumers of the Data Space. The reference architecture IDS-RAM 4.0 [14] is used as an example for these scenarios, particularly the Connector and App Store components. Nonetheless, these scenarios can be extrapolated to other reference architectures with similar building blocks, such as the Data Space Blueprint published by the DSSC.

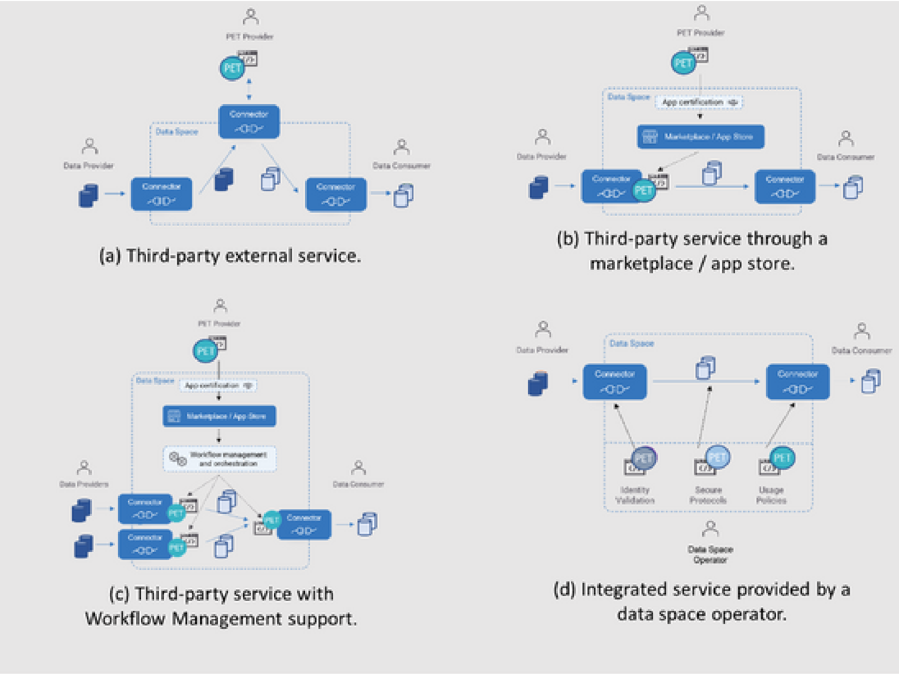
The basic functionality required to provision PETaaS within a Data Space is the ability to execute third-party services within the Data Space itself. This is again contemplated in IDS-RAM 4.0 as well as the Data Spaces Blueprint. Building on top of that functionality, the following scenarios to provision PETs as a service can be defined:

A: PET as an external third-party service: The Data Space does not provide any PET-specific functionalities nor an app store / marketplace through which the PET can be published and consumed. In this case, a PET is to be provided directly by the PET Provider to interested Data Space participants. This scenario implies that the original untreated data is sent by the Data Provider to be processed by the PET Provider before being forwarded to the Data Consumer and does not include local execution of the PET service. In spite of its shortcomings, this scenario is very straightforward to apply and would be beneficial for some quick, preliminary exploratory use cases on the integration of PETs and Data Spaces.

B: PET as a third-party service through a marketplace / App Store: The Data Space provides a marketplace / App Store through which the PET service can be published and consumed. In this case, a PET is to be implemented by means of its own data apps (e.g. for workflow management and data processing) as separate overlay on top of an existing Data Space. However, the App Store / Marketplace offers two important functionalities, absent in the previous scenario. On the one hand, it allows participants interested in the PET to consume, download, and install the PET service locally, thus guaranteeing a higher level of sovereignty over their data considering that they do not need to share the raw data. On the other hand, the App Store can optionally provide the PET Provider with the possibility to certify its PET service prior to publication, a process overtaken by a third independent party and which increases the level of trust of other Data Space participants in the PET service.

C: PET as a third-party service with workflow management support: The Data Space includes a workflow management engine capable of executing, coordinating and aggregating results of distributed / decentralized PET services. This scenario gives the possibility of offering more complex PET services, such as federated / decentralized learning or common Privacy Patterns to the Data Space participants, where multiple instances or parts of the service can be executed locally across different participants and the results can then be aggregated by the workflow management engine. It should be noted that this workflow management capability is not contemplated by current Data Space reference architectures, and thus significant work is required for it to be implemented.

D: PET as an integrated service provided by the Data Space operator: this is the most advanced scenario in which the Data Space Operator integrates with a PET Provider role, jointly providing an aligned / integrated PET and Data Space service offering to their participants. Jointly they minimize the effort for participants in configuring and applying PET implementations and can vary their offerings with new or improved PET services over time. In this case, the Data Space Operator can also cooperate with external PET Provider to offer PET services jointly, giving further flexibility to the offered services.



(a) Third-party external service.

(b) Third-party service through a marketplace / app store.

(c) Third-party service with Workflow Management support.

(d) Integrated service provided by a data space operator.

Figure 7: The various scenarios of providing PETaaS within a Data Space.

It should be noted that these scenarios focus on the technical integration of PETaaS within Data Spaces and should be looked at jointly with the additional considerations for a holistic picture on an aligned (reference) architecture for PETs and Data Spaces. For instance, offering PETaaS in any of the scenarios would not make sense without considering discoverability, trust and data interoperability as discussed in the previous section. Similarly, providing common Privacy Patterns can benefit from these integration scenarios, particularly scenario (c) with a workflow management engine to support the execution of these patterns.

In addition, the operations model describes the various service providers responsible for deploying and operating the Data Space and PET initiatives. The prevailing operational model is the four-corner model, originally developed for the Pan-European Public Procurement Online (PEPPOL) network [53] to standardize and simplify international procurement across borders. It has for instance also been successfully deployed in the Smart Connected Supplier Network (SCSN) Data Space [54] for the smart industries sector. Given its success, the four-corner model is a viable consideration for operating Data Space and PET initiatives as well. This model identifies three distinct types of service providers:

INFRASTRUCTURE-AS-A-SERVICE PROVIDERS

Providing intermediary roles that jointly enable a Data Space, e.g. the intermediary roles as described in the previous section. It is expected that the Infrastructure-as-a-Service providers will emerge to offer their services in a generic manner for multiple sectors, not only for mobility. This approach provides options for economies of scale and ensures interoperability when federating multiple Data Spaces.

In the context of aligning the PETs and IDS initiatives, the role of the Infrastructure-as-a-Service Provider is in providing the converged technical grounding as operational basis, whilst taking care of interoperability. It is expected that this will lead to a consolidated / limited number of service providers. In optimizing the efficiency of the infrastructure, a main driver is to develop the converged technical grounding to reap the benefit from economies of scope, by simultaneously supporting multiple types of data sharing alongside PETs, see Chapter IV.

CONNECTING SERVICE PROVIDERS

That connect data providers and data consumers to the Data Space, for example through specific data apps on a generic Data Space connector. This is a rather generic IT service; service providers may emerge that will provide their services in a generic manner for multiple sectors.

In the context of aligning Data Spaces and PETs, the role of the Connecting Service Provider is to connect data providers and data consumers to the common technical grounding Data Space by means of the connector whilst simultaneously integrating the data apps needed to support the PET-services, e.g. including the workflow management capabilities.

VALUE ADDING SERVICE PROVIDERS

That provide value adding services in the application domain, which may consider becoming part of a Data Space (see Section 9.5 in the chapter on data value creation).

In the context of aligning Data Spaces and PETs, the role of the Value Adding Service Provider specifically relates to the roles in the "PET Service and Operations Environment" as described in the previous section and depicted in Figure 6: the PET Initiator, the PET Service Provider and the PET Operator.

VI. CONCLUSIONS AND RECOMMENDATIONS

VI.1. Conclusions

As described throughout this white paper, many mutual benefits may arise from a solid alignment between Data Spaces and PETs:

1

Data Spaces can derive advantages from PETs, as PETs extend and enhance the set of data sharing services within a Data Space. This broadens the spectrum of supported use cases. Moreover, they provide participants (both data providers and consumers) with single-point-of-entry to a diverse set of data sharing services with associated economy of scope benefits for Data Spaces and onboarding efficiency for its participants.

2

PETs can find benefits in their association with Data Spaces, simplifying their deployment compared to various other environments. As such, Data Spaces can facilitate the scalability of PET implementations. Notably, predefined stakeholder roles and operational processes in Data Spaces can be used for PET operations as well. Furthermore, the availability of integrated PETs reduces implementation risks and minimizes vendor lock-in for data providers. The providers of PET services also capitalize on increased attention for and the impetus to implement Data Spaces as part of the EU Data Strategy.

Despite these potential benefits, alignment of the Data Space and PET developments also brings a potential weakness due to interdependencies in development and deployment with the risk of a longer development time towards implementation structures and complex and interdependent governance structures.

Additionally, since PETs typically address a specific type of analysis, it is also realistic to expect that there remains room for 'point solutions' in which shared data analysis is implemented by means of PETs without involvement a Data Spaces. Vice versa, the initial Data Spaces currently emerging are not (yet) developed to support PET use cases.

In the areas of multi-organizational data sharing and data spaces, the standards for specifying and sharing data models are currently being developed and also partially being implemented. However, that is less the case for the sharing of data processing models. Standardization of data processing models and alignment with business processes and workflow management concepts would both stimulate adoption of PETs and align with Data Space development. For instance, interface standards for input and output information could or should be developed. In addition, the need and benefits for a PET-specific data plane and associated interfaces as part of the Data Space research and development should be considered.

Finally, in this white paper the topic of alignment of Data Spaces and PETs has been addressed, with focus on re-use of operational processes and building blocks. However, at the same time it is more than probable that the overall implementation of PETs will consist of more than one technology. This means that the PETs themselves need to be interoperable as well. Interoperability between PET implementations based on different approaches and from different vendors may need to become interoperable and federated as well. This last aspect will raise its own challenges and justifies its own research and development roadmap. Moreover, although many PETs are mature and available as a commercial market offering, at the same time there is also much research and development still being done, thus it is expected that new PETs will emerge, requiring continuous market monitoring. Their embedding in Data Spaces (in addition to existing PET solutions) will pose an additional challenge.

VI.2 Recommendations

While the potential for mutual benefits is evident, achieving alignment between Data Spaces architectures and PET solutions is a complex task, necessitating significant efforts. Building upon the structure from the DSSC taxonomy (as described in Chapter III and depicted in Figure 1), this endeavor should encompass future work on both organizational and business aspects and on technical aspects. On each of these aspects recommendations are therefore provided.

VI.2.a. Organizational and business

The organizational and business recommendations on Data Space and PET alignment include.

- **Define a harmonized (business) role model for Data Space and PET alignment**

  Establishing a joint role model for Data Spaces and PETs will furnish guidance and the foundation for a unified and common development of both business models, governance models and reference architectures. An initial proposal for such a harmonized (business) role model for Data Space and PET alignment has been presented in Chapter V. It distinguishes multiple roles for PET-enabled Data Spaces, each can be fulfilled by an individual stakeholder and supported by its own business model.

- **Align on common operations processes**

  Although Data Spaces and PETs seem to have their own architectural concepts, they also appear to have operational processes in common as addressed in Chapter IV, e.g., with respect to on/off-boarding, data access, and trust (including identity management and usage control mechanisms).

- **Define a legal and governance structure and framework**

  An appropriate governance framework is essential for the overall operation and to obtain an aligned Data Space and PETs approach. It defines the rules and practices that govern the management, sharing, and utilization of data and processing models. Its rules agreements and (best) practices must adhere to legal requirements, ethical standards, and ensure interoperability. It should adopt a multi-level governance model, incorporating subsidiarity principles whilst adhering to the broader (European) strategy on data sharing, e.g., conforming to the ambition of the EU Data Strategy expressed as the common European Data Spaces. Private and public interests need to be balanced.

- **Identify and describe a common scope and terminology**

  Alignment of Data Space and PET approaches needs a common understanding on scope and terminology. For instance, a standardized way of defining privacy-sensitive data operations helps to link PET concepts to the Data Spaces architecture and concepts. Of specific importance is the upcoming ISO/IEC 20151[6] "Dataspace concepts and characteristics". It will enable a common basis of understanding between the multitude of stakeholders involved.

- **Set up a community on Privacy Patterns & models for PET and Data Space interoperability**

  Common and aligned Privacy Patterns and models may provide the glue between Data Spaces and PETs, thus allowing for the creation of a basis for interoperability. Moreover, joint development thereof would enable people from the Data Space and PET communities to closely collaborate and establish common visions, goals and architectures. The establishment of such a joint community may be considered in the context of the BDVA, IDSA, or DSBA. An ECLIPSE interest group on models for privacy has also been established to foster its community. Standardization (and perhaps even certification) will foster trust in PETs provided by third parties. Specifically, standards as identified in the Data Act could be set up to support this.

VI.2.b. Technical

The technical recommendations on Data Space and PET alignment include.

Develop a market place for both data sources and data processing algorithms

In an aligned Data Space and PET approach, both data sources and data processing algorithms need to be made available and discoverable. Hence data and data process algorithms market places are required to support reusability and accessibility. For Data Space the IDSA metadata broker and Gaia-x clearing house building blocks have been developed and are becoming (open-source) available. Similarly, the adoption of PETs will be fostered by making PET functionalities and PET data processing algorithms available through a PET marketplace, e.g., implemented by means of an 'app store' as being developed as part of the Data Space architectures.

Provide a common technical foundation

A common technical foundation is needed to support the common operations processes for Data Spaces and PETs as identified in this white paper. This common technical foundation may be based on the common technical foundation as being defined in the DSSC blueprint and of which the building blocks are expected to be developed by the SIMPL procurement initiative.

As part of the common technical foundation it is to be further addressed what PET-specific data and algorithm sharing usage policies are required and whether to develop into a set of templates. A standardized manner for describing the algorithm processing capabilities for the various types of PETs would further enable their integration into the Data Spaces environment.

Define the appropriate abstraction layers and interfaces between Data Space and PET responsibilities

In order to achieve privacy with acceptable performance overhead, PETs usually require fine-tuning by experts for given applications. Moreover, combinations of PETs are often required. Therefore, applying PETs in the context of Data Spaces requires clear separation of responsibilities (enabled by an aligned (business) role model, e.g., as depicted in Figure 6) supported by means of well-defined, and preferably even standardized, interfaces between roles and capabilities. These interfaces must define the right abstractions that enable good trade-offs in abstraction and re-usability between the various roles. Moreover, standardized interfaces may support PET-interoperability across approaches, solutions and vendors, lowering the barriers for adoption, together with a in a PET-friendly way to access the data.

Identify common Privacy Patterns (archetypes)

Privacy patterns offer a method to establish a connection between a shared technical foundation and the operational processes within a Data Space, facilitating the orchestration processes of PETs for information protection. This promotes technical interoperability and simplifies deployment. Establishing a joint role model for Data Spaces and PETs will furnish guidance and the foundation for a unified business model and reference architecture. Hence, there is a need to analyze Data Spaces capabilities and information protection architecture requirements in a standardized way, by adopting a concept like Privacy Patterns which will also help to specify the PET compiler.

Support Business Process and Workflow Management and Orchestration tooling

Accessing data and execution PET algorithms means more than only making the data findable and available. It may involve the use of data processing algorithms coming from different organizations. Moreover, implementing and configuring a PET in many cases needs a workflow which handles more than one PET model and algorithm, which needs to be put in the right order and control all data inputs and outputs. It requires not only governance on data, but smart workflow management and orchestration of the services, running on the premises or within the security domains of various organizations. That means that effort should be spent into analysis and requirements on business process and workflow management and orchestration models for managing PETs in a Data Space context. For PETs, there is currently a generic lack in transparency of the undergoing complex analysis and its state of the model execution workflows and the (transformation of) the sensitive data from different data providers. There is a need to be able to trace the state of running analysis and all operations on data through the whole workflow of the PET models and algorithms, supporting traceability and auditability.

Develop a tool-set for interoperability and integration

To stimulate adoption of the combination of Data Spaces and PETs, the barriers for joint implementations should be made as low as possible. Development of scenarios with associated (open-source) tooling can enable this. A minimum viable product can form the starting point, based on an initial set of representative scenarios for combination of Data Spaces and PETs. This may take the form of PET-tailored Data Space connectors (data apps) supporting the specific requirements of the PETs, e.g., on workflow management and pre-defined privacy patterns.

VII. CALL FOR ACTION

The list of challenges to align Data Space and PETs is substantial. Nevertheless, steps for aligning the operational processes seem feasible to address on the shorter term. The need to do so is motivated by the observation that the convergence of PETs and Data Spaces is already beyond fundamental research. Data Spaces are being deployed now and operated by businesses, they should treat sensitive data in an appropriate way, the technology to do so is available and the mutual drivers for Data Spaces and PETs to align seem obvious. Moreover, opportunities and programmes for funding are available to address further research and development topics, e.g., from the Horizon Europe programme.

Therefore, we now call upon the joint Research, Development and Innovation community in the EU to adopt the topic of Data Space and PET alignment as strategic focal point and to set up a research and development roadmap for it.

VIII. REFERENCES

[1] European Commission (2020). “A European strategy for data”. https://digital-strategy.ec.europa.eu/en/policies/strategy-data.

[2] EU PrepDSpace4Mobility CSA (2023). "Preparatory Studies for the European Mobility Data Space (EMDS) - Analysis of building blocks and requirements". https://mobilitydataspace-csa.eu/wp-content/uploads/2023/10/deliverable-3.1.pdf.

[3] EU Digital Europe Programme. “Data Spaces Support Centre (DSSC)”. https://dssc.eu.

[4] EU Data Spaces Support Center (DSSC). "DSSC Glossary", March 2023. https://dssc.eu/space/Glossary/55443460.

[5] EU Data Spaces Support Center (DSSC). “Building Block Taxonomy – Version 0.5”. https://dssc.eu/space/BBE/178421761/Building+Blocks+%7C+Version+0.5+%7C+September+2023.

[6] EU Data Spaces Support Center (DSSC). “Blueprint – Version 0.5”. https://dssc.eu/space/BBE/178422228/Technical+Building+Blocks.

[7] EU Digital Europe Programme, “SIMPL: cloud-to-edge federations and Data Spaces made simple”. https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-spaces-made-simple.

[8] EU PrepDSpace4Mobility CSA (2023). "First Public Stakeholder Forum". https://mobilitydataspace-csa.eu/wp-content/uploads/2023/03/psf-28february.pdf.

[9] European Commission (2022). “European Data Act”. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113.

[10] European Commission (2022). “European Data Governance Act”. https://digital-strategy.ec.europa.eu/en/policies/data-governance-act.

[11] European Commission (2022). "European Digital Services Act". https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en.

[12] European Commission (2022). "European Digital Markets Act". https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.

[13] European Union (2021). "European Artificial Intelligence Act". https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

[14] International Data Spaces Association (IDSA) (2022)."International Data Spaces: Reference Architecture Model Version 4". https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0.

[15] EU Gaia-X Initiative. "Gaia-X Federation Services - GXFS". https://www.gxfs.eu/specifications.

[16] EU Gaia-X Initiative. Gaia-X - Architecture Document - 22.04 Release. https://gaia-x.eu/wp-content/uploads/2022/06/Gaia-X-Architecture-Document-22.04-Release.pdf.

[17] FIWARE. "Components". https://www.fiware.org/catalogue.

[18] Dutch Neutral Logistics Information Platform (NLIP). "iSHARE Data Sharing Initiative". https://www.iSHAREworks.org/en.

[19] iSHARE Foundation. "(Benefits) For Data Spaces". https://ishare.eu/ishare/benefits/for-data-spaces.

[20] Data Space Business Alliance (DSBA). "Unleashing the European Data Economy". https://data-spaces-business-alliance.eu.

[21] Data Space Business Alliance (DSBA). “Technical Convergence Discussion Document”. https://data-spaces-business-alliance.eu/dsba-releases-technical-convergence-discussion-document.

[**22]** International Data Spaces Association (IDSA). “Data Space Radar - Faster IDS breakthroughs are within range”. https://internationaldataspaces.org/adopt/data-space-radar.

[**23]** German Mobility Data Space. "The data catalogue: The entire data offering at a glance". https://mobility-dataspace.eu/data-catalogue.

[**24]** European Union (2017). “New European Interoperability Framework (EIF) – Promoting seamless services and data flows for European public administrations”. https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf.

[**25]** EU European Union Agency for Cybersecurity. “Data Protection Engineering”. https://www.enisa.europa.eu/publications/data-protection-engineering.

[**26]** OECD (2023), “Emerging Privacy Enhancing Technologies – Current regulatory and policy approaches”, OECD Digital Economy Papers, No. 351. https://www.oecd.org/publications/emerging-privacy-enhancing-technologies-bf121be4-en.htm.

[**27]** The Royal Society (2023), “From privacy to partnership – The role of privacy-enhancing technologies in data governance and collaborative analysis”. https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies

[28] United Nations Big Data (2023), “The United Nations Guide on Privacy-Enhancing Technologies For Official Statistics”. https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf.

[29] Adams, C. (2021). “A Collection of Tools: The PrivacyTree”. Introduction to Privacy Enhancing Technologies. Springer, Cham. https://doi.org/10.1007/978-3-030-81043-6_2.

[30] Nesrine Kaaniche, Maryline Laurent, Sana Belguith. “Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey”. Journal of Network and Computer Applications, Volume 171, 2020, 102807, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2020.102807.

[31] Weise, Martin and Andreas Rauber. “A data-visiting infrastructure for providing access to preserved databases that cannot be shared or made publicly accessible.” International Conference on Digital Preservation (2021).

[32] EU-Project Trusted Secure Data Sharing Space (TRUSTS). "TRUSTS project deliverable D2.6: Architecture design and technical specifications document". https://www.trusts-data.eu/wp-content/uploads/2021/09/D2.6_Architecture-design-and-technical-specifications-document-I.pdf.

[33] Dutkiewicz, L., Miadzvetskaya, Y., Ofe, H., Barnett, A., Helminger, L., Lindstaedt, S., & Trügler, A. (2022). “Privacy-Preserving Techniques for Trustworthy Data Sharing: Opportunities and Challenges for Future Research”. Data Spaces: Design, Deployment and Future Directions, 319-335. http://dx.doi.org/10.1007/978-3-030-98636-0_15.

[34] EU Platoon Project. https://platoon-project.eu.

[35] MyData Initiative. “Understanding MyData Operators ”. https://www.mydata.org/publication/understanding-mydata-operators.

[36] CaPe. “A Consent Based Personal Data Suite ”. https://github.com/OPSILab/Cape.

[37] Spanish Data Protection Authority (AEPD) (2023), “Data Spaces, sovereignty and privacy by design”. https://www.aepd.es/en/prensa-y-comunicacion/blog/data-spaces-sovereignty-and-privacy-by-design.

[38] EU Enershare project (2023). "The Energy Data Space for Europe - Bringing together energy and data value chains to enable the energy transition". https://enershare.eu.

[39] W. Abbasi, P. Mori, A. Saracino and V. Frascolla, "Privacy vs Accuracy Trade-Off in Privacy Aware Face Recognition in Smart Systems," 2022 IEEE Symposium on Computers and Communications (ISCC), Rhodes, Greece, 2022, pp. 1-8, https://dx.doi.org/10.1109/ISCC55528.2022.9912465.

[40] Center-of-Excellence Data Sharing and Cloud (CoE DSC, 2023). "A design for scalable governance for MPC data collaborations - Based on the case study for the data collaboration in the Dutch elderly care sector". https://coe-dsc.nl/governance-framework-for-mpc-data-collaborations-based-on-the-use-case-in-the-dutch-elderly-care-2.

[41] Privacy Patterns Organization. “Privacy patterns”. https://privacypatterns.org.

[42] EU Open DEI project. “Reference Architectures and Interoperability in Digital Platforms”. https://www.opendei.eu/case-studies/reference-architectures-and-interoperability-in-digital-platforms.

[43] Wikipedia. "Software design pattern". https://en.wikipedia.org/wiki/Software_design_pattern.

[44] E. Gamma, R. Helm, R. Johnson and J. Vlissides. "Design Patterns: Elements of Reusable Object-Oriented Software". Addison-Wesley, 1994.

[45] P. Papadimitratos *et al.*, "Secure vehicular communication systems: design and architecture," *IEEE Communications Magazine*, vol. 46, no. 11, pp. 100-109, November 2008, https://ieeexplore.ieee.org/document/4689252.

[46] IRT SystemX. “Data privacy issues in C-ITS: acceptability of private mobility data sharing”. https://youtu.be/KjZ2zrhqPxw

[47] TAM-project. "Trust Autonomous Mobility". https://www.irt-systemx.fr/en/projets/tam.

[48] EU Data Spaces Support Centre (DSSC) initiative. "Control plane versus Data plane". https://dssc.eu/space/BBE/178422298.

[49] International Data Spaces Association (IDSA). “RAM Version 4 – Certification Perspective”. https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/blob/main/documentation/3_Layers_of_the_Reference_Architecture_Model/3_4_Process_Layer/3_4_6_Policy_Enforcement.md.

[50] The Netherlands AI Coalition (NL AIC) Working Group Data Sharing (2021). “Towards a federation of AI Data Spaces - NL AIC reference guide to federated and interoperable AI Data Spaces”. https://nlaic.com/wp-content/uploads/2023/04/NL_AIC_Towards_a_federation_of_AI_data_spaces.pdf.

[51] The Netherlands AI Coalition (NL AIC) Working Group Data Sharing (2022). “Reference guide for intra AI Data Space interoperability – Developing individual AI Data Space instances”. https://nlaic.com/wp-content/uploads/2023/04/NL-AIC-intra-AI-Data-Space-Interoperability-v3.1.pdf.

[52] International Data Spaces Association (IDSA). "RAMv4 - Publishing and using IDS Apps". https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/blob/main/documentation/3_Layers_of_the_Reference_Architecture_Model/3_4_Process_Layer/3_4_5_Publishing_and_using_Data_Apps.md.

[53] Holmlund, Per (2022), "Understanding the Peppol four-corner model of business exchange", Blog post, https://qvalia.com/blog/understanding-the-peppol-four-corner-model-of-business-exchange.

[54] Smart Connected Supplier Network (2023), “Four-corner model”. https://smart-connected-supplier-network.gitbook.io/processAmanual/architecture/four-corner-model.

[55] Centre of Excellence for Data Sharing and Cloud. www.coe-dsc.nl.

[56] Big Data Value Association (BDVA), www.bdva.eu.

Acknowledgement

The work presented in this white paper has been collaboratively done by the Dutch National Centre of Excellence on Data Sharing and Cloud (CoE-DSC [53]) and the Big Data Value Association (BDVA [54]). We would like to thank both initiatives for providing us the opportunity to perform this interesting research and the individual participating partners for providing valuable inputs within a stimulating and cooperative setting.

ABOUT BDVA

About the Big Data Value Association

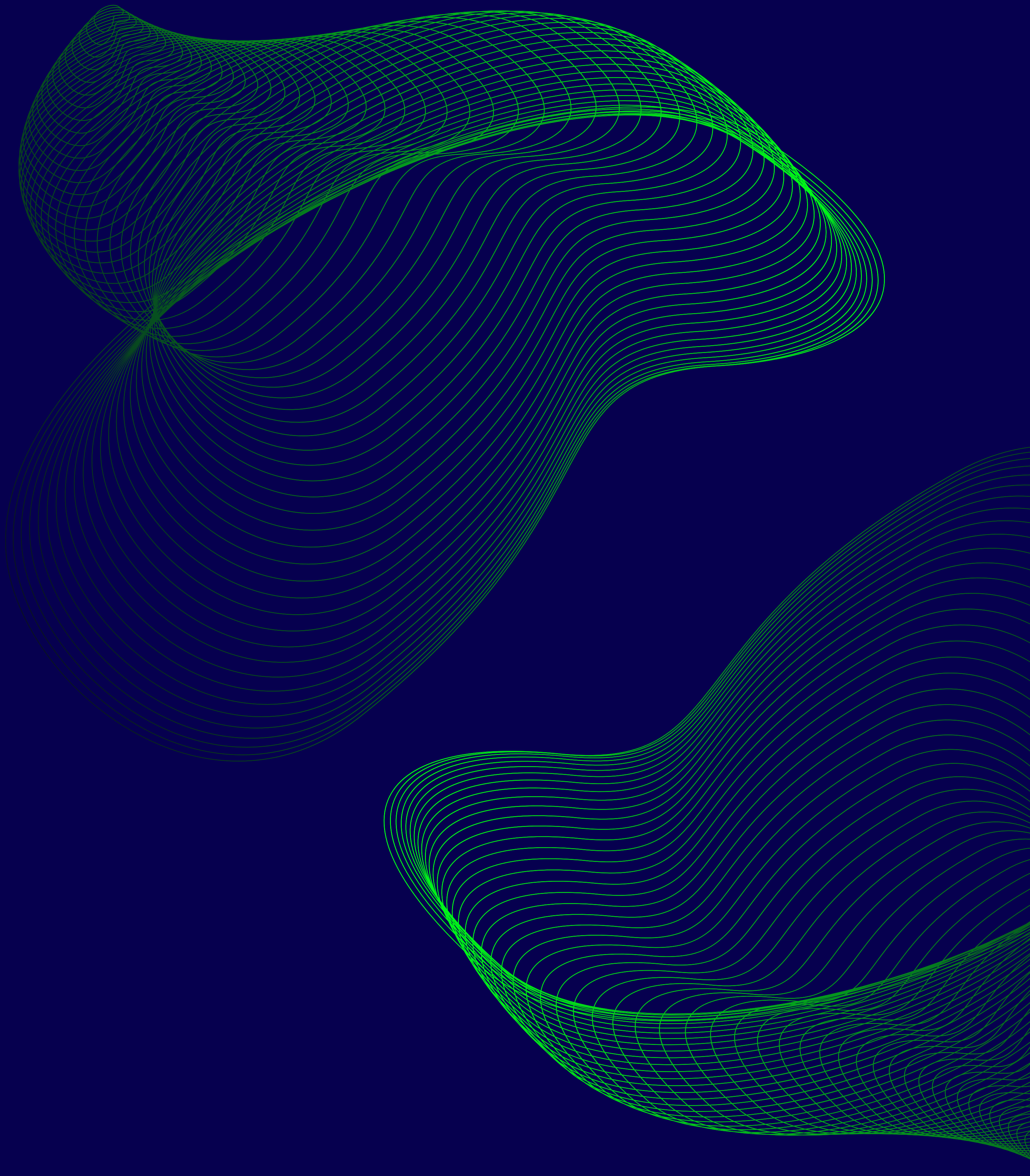
BDVA is an industry-driven international not-for-profit organisation with 250 members all over Europe and a well-balanced composition of large, small and medium-sized industries as well as research and user organisations. Our mission is to develop an innovation ecosystem that enables the data-driven digital transformation of the economy and society in Europe, delivering maximum benefit. To reach this goal, we focus on advancing areas such as big data technologies and services, data platforms and data spaces, industrial AI, datadriven value creation, standardisation and skills.
BDVA enables existing regional multi-partner cooperation, to collaborate at the European level through the provision of tools and know-how to support the cocreation, development and experimentation of pan-European data-driven and AI applications and services and know-how exchange.

Through BDVA, our members contribute to the European data and AI R&I agenda and develop guidelines and strategic roadmaps for industry and policymakers in BDVA Task Forces and our events give opportunities to build new collaborations and co-create new projects. Being part of the BDVA community, the members gain higher visibility on the European level and our services are designed to give timely updates on all the latest developments in the fields of data and AI.

BDVA believes in collaborations! BDVA has been the private side of the H2020 partnership Big Data Value PPP, it is a private member of the EuroHPC JU and it is a founder member of the AI, Data and Robotics Partnership. BDVA has developed a strong and growing cooperation with Gaia-X, IDSA and FIWARE through the Data Spaces Business Alliance (DSBA), it is a partner of the Transcontinuum Initiative (TCI) and collaborates with many industry-driven AI national initiatives and other European communities.

BDVA is open to new members!

Visit BDVA.EU to learn more about members and activities. You can contact us anytime at info@bdva.eu.



ABOUT COE-DSC

About the Centre of Excellence for Data Sharing and Cloud (CoE-DSC)

Data sharing has the potential to generate new economic and societal value. Traditionally, organisations establish bilateral connections or share their data through a central platform controlled by a single party. However, these approaches carry the risk of vendor lock-in, fragmentation, and limited value creation.

Data spaces are the next evolution in data sharing

The next evolution in data sharing is the concept of data spaces. A data space is a decentralised infrastructure that organisations can use to make their data accessible to others based on specific agreements. Any party adhering to these agreements can participate and exchange data. Data spaces offer the scalability that traditional methods canoot offer, enabling greater innovation and economic and societal value. Examples of existing "live" data spaces in the Netherlands include SCSN and HDN, as well as the Mobility Data Space in Germany.

The CoE-DSC supports the realisation of data spaces

Significant barriers still exist when establishing data spaces, such as building trust between participants, complying with regulations, developing governance structures, and defining business models. The CoE-DSC aims to lower these barriers and help participants realise the full potential of data sharing initiatives.
Reach out to us via **info@coe-dsc.nl** if:

You want to implement a use case at scale.
You need technical or operational support for your data sharing initiative.
You want to join our community, which consists of over 500 participants.

Whilst supporting the realisation of data spaces, the CoE-DSC closely monitors the impact of the European Data Strategy and proposed regulations like the Data Act to ensure compliance, and collaborates with the Data Spaces Support Centre (DSSC).

Note

This document should be referenced as follows: Leveraging the Benefits of Combining Data Spaces and Privacy Enhancing Technologies; Big Data Value Association (BDVA) and Centre of Excellence for Data Sharing and Cloud (CoE-DSC); 2024

BDV BIG DATA VALUE ASSOCIATION

CoE-DSC