



Perspective Article

From use case to data space, a bottom-up data space design framework leveraging Data Products



Casper Van Gheluwe^a, Gabriele Bozzi^a, Eridona Selita^b,
Nele Daels^a, Tanguy Coenen^a, Laure De Cock^{a,*}

^a imec, Kapeldreef 75 3001 Leuven, Belgium

^b Macq, Rue de l'Aéronef 2 1140 Evere, Belgium

ARTICLE INFO

Article history:

Received 9 December 2024

Revised 13 October 2025

Accepted 30 January 2026

Available online 26 February 2026

Dataset link: [No research data \(Original data\)](#)

Keywords:

Data spaces

Mobility

Use case

Canvas

Capability

Data product

ABSTRACT

Data spaces are emerging as a pivotal mechanism for trusted data sharing within the EU, yet comprehensive tools for their design remain necessary for practical implementation. This paper proposes such a tool, by introducing a bottom-up data space design framework to identify data space capabilities necessary to fulfil specific use cases. The framework uses an intake canvas, inspired by the Open Data Product Specification, to define functional requirements of use cases, and a capability mapping process for data space design. The framework was applied with sixteen data space use cases from nine EU countries and provided valuable insights into data space concepts for stakeholders. However, this bottom-up approach also has limitations, as it might fail to identify certain capabilities essential for a minimum viable data space. Therefore, a combined top-down and bottom-up approach is recommended for comprehensive data space design. While this framework was developed for the deployEMDS project, the deployment action for the common European mobility data space, it is domain-agnostic, making it a valuable asset for diverse applications in data space design.

© 2026 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>)

* Corresponding author.

E-mail address: laure.decock@imec.be (L. De Cock).

Specifications Table

Subject	Computer Science > Information systems
Specific subject area	Data spaces
Type of data	• Capability mapping matrix (Table 3) - primary data • Data space capabilities inventory (Table 2) - derived framework component • Intake canvas questions (Appendix) - framework instrument
Data collection	/
Data source location	/
Data accessibility	Included in the manuscript
Related research article	/

1. Value of the Data

- The method provided in this manuscript is valuable for everyone who wants to design a data space for their field.
- The method can be re-used by other data space practitioners by copying the canvas and capability mapping and adapting it to their domain.
- Researchers can further validate and develop the canvas and capability mapping.
- As the data space use cases influence the data space capabilities, and not the other way around, we call this a bottom-up approach for data space design.
- The framework addresses a gap by providing a systematic tool for translating use cases into data space technical capabilities. As organisations often lack practical decision support tools for data space design, this framework enables stakeholders to systematically identify the required capabilities and adapt their requirements to data space implementations. Validation through 16 use cases across nine EU countries confirmed its practical value for nascent data space initiatives.

2. Introduction

A single market allowing for the free movement of goods, services, capital, and people has long been the cornerstone of the European Union (EU). Similarly, the free flow of data is increasingly recognized as vital for future growth and innovation [1]. The European strategy for data aims at creating a single market for data, as data-driven innovation brings a lot of opportunities, such as improved mobility, better policymaking or upgrading public services [2]. The European Commission (EC) boosts trustworthy data-sharing through broad sets of measures, and by introducing new mechanisms. One of those mechanisms is data spaces.

According to the Commission staff working document on common European data spaces, the role of a data space is to “overcome legal and technical barriers to data sharing by combining the necessary tools and infrastructures and addressing issues of trust by way of common rules” [3]. This aligns with the definition of a data space, given by the Data Spaces Support Centre (DSSC) and the CEN Workshop Agreement Trusted Data Transactions, namely “an interoperable framework, based on common governance principles, standards, practices and enabling services, that enables trusted data transactions between participants.” [4,5]. In the second staff working document, the Commission further reinforced the importance of the common European data spaces, noting that they “play a critical role in providing a steady supply of data (...) to economic actors, thereby creating a fundament for AI innovation in Europe” [6].

It is clear that data spaces are being put forward as the new mechanism for trusted data sharing in the EU, but a deeper understanding of the design of data spaces is necessary to realize this digital innovation in practice [1]. While a lot of documentation exists on data space

conceptualization, it remains unclear how organizations who participate in the data space can be supported in making informed decisions, which lowers their willingness to share data. Jussen et al [7], who studied barriers for inter-organizational data sharing, identified that gap and put the development of supporting tools forward as a potential solution, in particular for the research question 'how to select suitable technical infrastructure for data sharing?' Also, Beverungen et al [8] stress that future research should be devoted to the technical foundations that support data sharing in a data space, and Möller et al [9] identify design options for data spaces as essential future research. This research aims to fill this gap by introducing a framework for bottom-up data space design, which provides tools to identify data space capabilities for use cases, by answering the following two research questions:

- How should suitable technical infrastructure be selected for data sharing in a data space, based on concrete use cases?
- What framework can bridge the gap between domain-specific use cases and the requirements of data spaces to support trusted data sharing?

This framework helps to identify data space design options by providing a structured methodology that translates (potentially domain-specific) use case requirements into concrete technical capabilities. Through its intake canvas and capability mapping matrix, the framework enables organizations to systematically collect requirements and map them to modular, technology-neutral capabilities. This approach allows stakeholders without deep technical expertise to influence data space design while ensuring that the resulting architecture addresses their specific needs. By balancing bottom-up use case requirements with top-down considerations for minimum viable data spaces, the framework helps organizations explore design options that not only meet immediate business requirements but also maintain compatibility with broader data space standards and interoperability guidelines.

The framework is developed and tested in the deployEMDS project. This project is the deployment action for the common European mobility data space [10]. While the framework originated in an urban mobility project, it is domain and technology agnostic, so it can be used to design data spaces in different sectors. The remainder of this paper is structured as follows: in the next section a background is given on existing data space initiatives and research; in the third section the method used to develop the framework is outlined; the different components of the framework are presented in the results section, followed by a demonstration for the Flanders and Barcelona implementation sites in the deployEMDS project, and a discussion of the results.

3. Background

According to Otto et al. (2022) [1] the role of data for enterprises is fourfold: data is an enabler of operational excellence within a company, data has become a product which is sold on the market, data is a source of business innovation, and data is considered a strategic resource for long-term sustainability of the economy. For data to fulfil these roles, it must be used to produce information [11], which is why the use and re-use of data is gaining importance (e.g., in the European strategy for data). The overall goal is to evolve from data exchange to data sharing, which entails a collaboration between organizations for a shared goal [12]. Data exchange refers to the transfer of data between parties in a one-time or transactional manner, ensuring interoperability through structured formats and protocols but without establishing ongoing relationships or shared governance. In contrast, data sharing features a more collaborative approach where multiple parties contribute to and access data within a governed framework, incorporating agreements on usage, common governance structures, and trust mechanisms. Unlike data exchange, which is often limited to specific transactions, data sharing enables continuous access, reusability, and value co-creation, making it essential for ecosystems like data spaces where organizations collaborate to maximize data utility. Data sharing can therefore foster data-driven digital innovation trends, which are significantly shaping the transformation of both business and society today, such as cloud computing, Internet of Things (IoT) or Artificial Intelligence.

Data sharing is essential to enable these innovations, and a key development in this domain is the emergence of data ecosystems that encourage interconnectedness among diverse stakeholders and organizations. Data ecosystems facilitate collaborative efforts to create services driven by data supply on the one hand and data utilization on the other [8,13]. They are rooted in the growing interconnections among organizations, which facilitate data sharing within the context of value co-creation. Consequently, digital transformation has evolved into a phenomenon that extends beyond the organizational level [14]. As a new approach for data sharing, data ecosystems have gained attention in recent years, both in academia and in large-scale public initiatives [15]. Specifically, the concept of data spaces has emerged as a promising mechanism, because it supports data sharing and data sovereignty in ecosystems by introducing a distributed software infrastructure [1,16].

Although the concept of data spaces has been defined 20 years ago [17], the research and practice of data spaces is still emerging [8]. Existing literature focuses on inter-organizational data sharing (e.g., [7,18]), the organization and components of data ecosystems (e.g., [19,20]), use cases that can benefit from data ecosystems (e.g., [21,22]), and technical solutions for data sharing (e.g., [23–25]). While the scientific exploration of data spaces is still in its nascent stages, extensive documentation and project development have emerged, driven by the European data strategy. Recognizing the pivotal role of data spaces to build a robust European data economy, the European Commission has established three programmes to launch these data spaces: the development of technical infrastructure, a data spaces support centre, and the deployment of common European data spaces in nine domains. The first programme focusses on the development of a common technical infrastructure, and this should be realized by the Simpl procurement project,¹ which is developing an open source, smart and secure middleware platform. Besides Simpl, several other initiatives play an important role in data space design, by suggesting reference architectures, by developing open-source software components or by designing communication protocols and standards (e.g., IDSA,² Gaia-X,³ FIWARE⁴). The second programme, and a key initiative for aligning different data space architectures, is the Data Spaces Support Centre⁵ (DSSC). The DSSC offers guidance, best practices, and tools to support the development and management of data spaces, helping stakeholders navigate technical challenges and supporting them be legally compliant. An important deliverable of the DSSC is the Data Spaces Blueprint, a set of guidelines and recommendations for the development of data spaces, which received an update to v2.0 in March 2025 [4]. Other supporting initiatives concentrate on ecosystem engagement (e.g., the Big Data Value Association⁶), and many are integrated into national frameworks, such as the Belgian Data Spaces Alliance,⁷ which also acts as the local hub for Gaia-X and IDSA. The third programme focuses on the deployment of common European data spaces in different domains, such as health, agriculture and mobility. An informative tool to explore the growing amount of data spaces is the Data Space Radar,⁸ which allows to filter on development stage and domain. In the mobility domain, the concept of a common European mobility data space is embedded in the Sustainable and Smart Mobility Strategy, to collect, connect and make data available for sustainability and multimodality enhancement [26]. The deployEMDS⁹ project is one of the initiatives that contribute to the common European mobility data space, with a specific focus on the deployment of the data space in nine implementation sites in urban areas across different member states [10]. The challenge of this project lies thus in the integration of the data space in the existing mobility data ecosystems, aligning with the existing infrastructure,

¹ <https://digital-strategy.ec.europa.eu/en/policies/simpl>.

² <https://internationaldataspaces.org/>.

³ <https://gaia-x.eu/>.

⁴ <https://www.fiware.org/>.

⁵ <https://dssc.eu/>.

⁶ <https://bdva.eu/>.

⁷ <https://www.dataspacesalliance.be/>.

⁸ <https://www.dataspaces-radar.org/radar/>.

⁹ <https://deployemds.eu/>.

and complementing existing processes for data sharing. An overview of >270 existing mobility data ecosystems can be found in the inventory of PrepDSpace4Mobility, the coordination and support action preceding the deployEMDS project.¹⁰

The numerous data space initiatives in the EU, encompassing support, preparation, and deployment actions, aim to elucidate the complex implementation of the data space concept across various domains. This complexity has also been emphasized by several researchers, such as Oliveira et al [19] and Beverungen et al [8]. While digital transformation within a single organization can be a challenging endeavor, transforming as part of a dynamic data ecosystem is even more complex because the ecosystem's transformation evolves beyond the control of a single organization [27]. As a result, many organizations are concerned about the planning, design, implementation, and maintenance of software infrastructure for data spaces [1]. Additionally, there remain many barriers for inter-organizational data sharing, including fear of revealing valuable or sensitive information [7,15]. To deal with these concerns and barriers, organizations need decision support tools [7]. Giess et al [28] developed such a tool, which enables stakeholders to gain a clearer understanding of data spaces. They developed a taxonomy for design options for data spaces, which aims to support both researchers and practitioners to understand, analyse, and develop (novel) data spaces. To build the framework they integrated scientific knowledge, leveraged from a structured literature review, and practical knowledge, leveraged from established data space projects. The framework was reviewed by experts in the field in multiple cycles and demonstrated for three use cases. The approach adopted in this paper for the data space design framework has several parallels with the approach of Giess et al. (2023). The data space design framework also combines scientific knowledge, through canvas and capability driven design, with practical knowledge, leveraging resources such as the DSSC Blueprint. It has also been reviewed by data space experts in the deployEMDS project, and was demonstrated through use cases. Both the framework and the taxonomy aim to fill the same gap, by providing a validated tool for data space design, but while the taxonomy focuses on categorizing data spaces, the framework focuses on defining bottom-up data space capabilities.

4. Design Principles Data Space Design Framework

The data space design framework aims to provide a functional tool for the preparatory stage of the data space co-creation method (Fig. 1). It can be used for two development processes, as defined by the DSSC [4]:

1. The process that identifies and describes individual use cases in detail to clarify the benefits for the data space participants and identifies the data space's functional requirements.
2. The process that translates the functional requirements into a useful data space design.

While the DSSC provides a description of these processes in its proposed data space co-creation method [4], they do not provide specific tools or methods yet, at the time of writing. This research aims to fill this gap by providing the data space design framework, thereby contributing to the DSSC Blueprint. The two pillars of the framework align with the two development processes, i.e., the intake canvas can be used for the first process to identify the functional requirements based on the identified use cases and their data products, and the capability mapping for the second process, to translate the functional requirements to the technical capabilities of the nascent data space. To achieve this, the proposed data space design framework is built on three core design principles (DP) that ensure its practical applicability and theoretical soundness. These principles, which are detailed in the subsections below, ensure that the resulting framework is use-case-driven (DP1), capability-driven (DP2) and framework and technology-agnostic (DP3).

¹⁰ <https://mobilitydataspace-csa.eu/inventory/>.

Co-creation

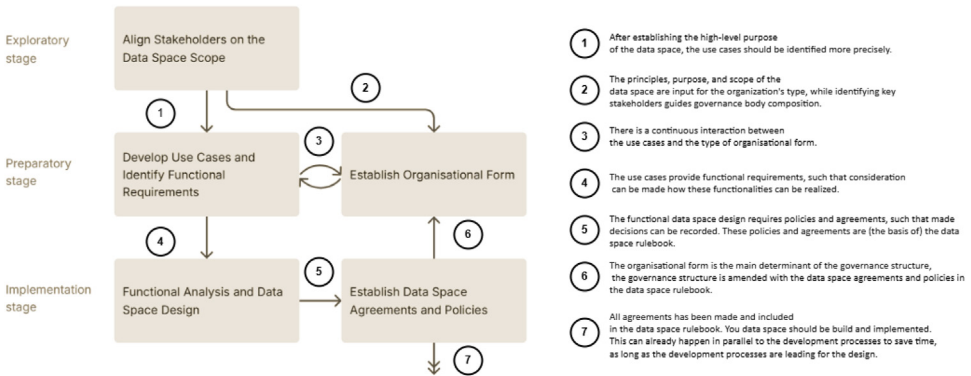


Fig. 1. Development process of data space co-creation [4].

4.1. Design principle 1: use-case-driven

The data space design framework places emphasis on practical use case requirements rather than abstract architectural considerations. It therefore adopts a bottom-up methodology, beginning with the collection of information on the use cases. To this end an intake canvas was prepared. The central concept of the canvas is a Data Product. The DSSC Blueprint v2.0 states that “a data product may include (...) the data products’ allowed purposes of use, (...), access and control rights, pricing and billing information, etc.”, while also noting that the definition of data products is still evolving in the wider data spaces community [4].

Additionally, in the recent CEN Workshop Agreement on Trusted Data Transaction, the Data Product also takes a central role in the conceptual model of a Data Transaction [5], as shown in Fig. 2, making it the ideal foundation for a bottom-up data space design framework. Positioned at the heart of the model, it connects critical elements including permissions, consent structures and usage rights and multiple operational components like Data Catalogues. The Data Product thus serves as both the practical manifestation of the legal and operational aspects of data sharing and as the touchpoint connecting all key stakeholders in a data ecosystem: the data rights holders, producers, providers and users.

Starting with the Data Product as the central concept ensures that the design framework is founded on from where actual value is created rather than from abstract concepts and reference architectures. This approach naturally aligns with bottom-up methodology by building from concrete use cases toward more generalized frameworks, ensuring that the resulting data space design remains aligned with the practical needs of the use cases.

4.2. Design principle 2: capability-driven

The second pillar of the framework is the mapping of the canvas questions to technical data space capabilities. The capability driven design (CDD) approach of Berziša et al. (2015) [29] is used to map the questions of the intake canvas to data space capabilities. According to the CDD meta-model a capability is the ability and capacity that enables an enterprise to achieve a business goal in a certain context. This link between business, context and system is the strength of CDD. This entails that when the business goal or application context change, the information

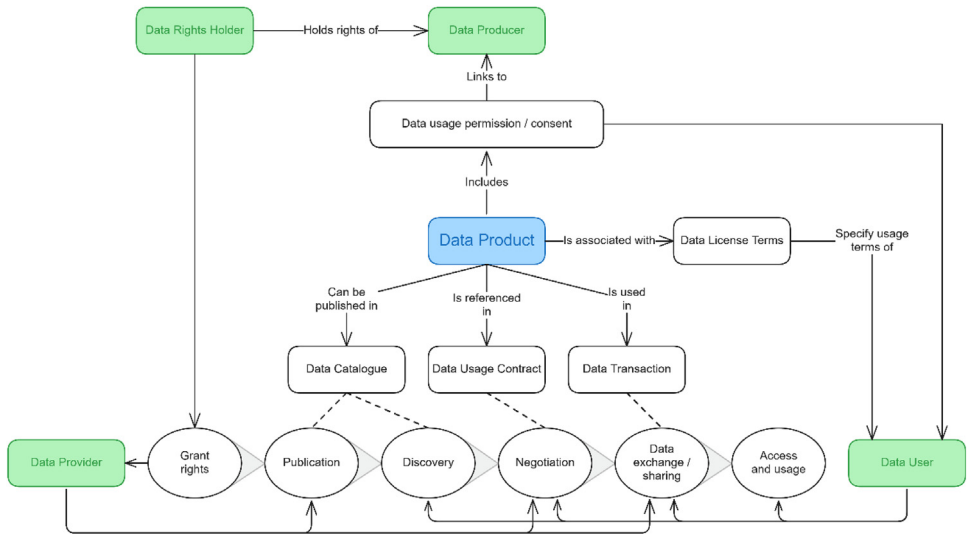


Fig. 2. Conceptual model of a data transaction [5],¹¹ showing the central role of a Data Product in a trusted data exchange.

system components can be changed as well, as new capabilities might be required, resulting in a flexible design of the system. Flexibility is needed in the data space design framework, as it is most likely that business goals or contexts might change during the design and development of a data space, as this often takes multiple years (e.g., three years in the case of the deployEMDS project). In this work, CDD is implemented by gathering input on the business goals and application contexts in the intake canvases. Next, the canvas questions are mapped to technical data space capabilities. These capabilities align with the functionalities of the DSSC Blueprint v2.0 building blocks, but they have been clustered and interconnected in a different way. This deviation was necessary to enable a CDD and follows the guidelines of the DSSC, who recognize that not all data space architectures necessarily benefit from consolidating functionalities into building blocks.

4.3. Design principle 3: technology and domain agnostic

The framework deliberately avoids presupposing any specific data space reference architecture, technology stack or domain constraints, ensuring broad applicability and futureproofing. This technology neutrality is achieved by focusing on functional capabilities rather than specific implementations. For example, identifying the need for "semantic interoperability" without mandating any particular ontology frameworks or vocabulary standards. This approach allows nascent data spaces to select the most appropriate technologies for their context while ensuring compatibility with the broader data space principles. Domain neutrality is maintained through generic question formulations that capture universal data sharing requirements (such as access control, data quality, and governance needs) regardless of whether the use case involves mobility, healthcare, agriculture, or other sectors. The capability inventory draws from cross-domain data space documentation to ensure comprehensive coverage beyond any single domain's specific requirements. This dual neutrality enables the framework to serve as a foundational tool for diverse data space initiatives while maintaining relevance as technologies and standards evolve.

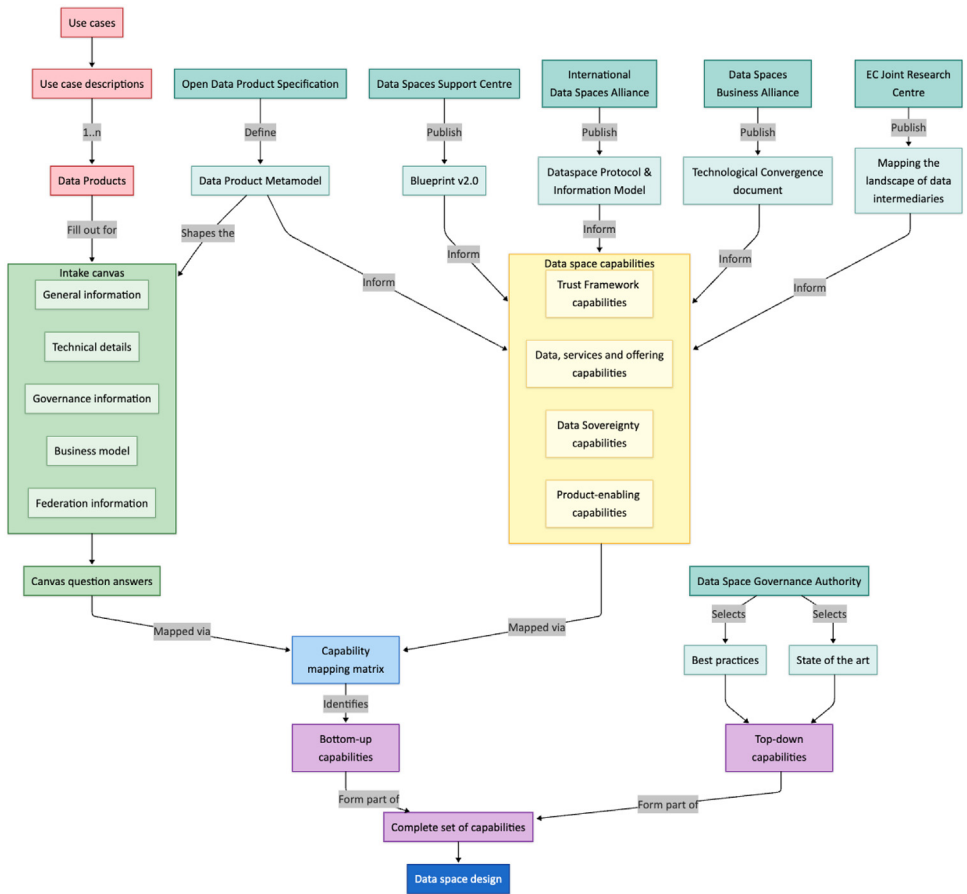


Fig. 3. Overview of the proposed data space design framework.

5. Data Space Design Framework and Components

Fig. 3 presents our comprehensive data space design framework, comprising three primary components: the intake canvas (green), data space capabilities (yellow), and capability mapping (blue). The framework implementation follows a systematic four-step process:

1. Data space use case representatives complete the intake canvas for each required Data Product, generating a set of structured responses. The canvas structure and content are detailed in the next section.
2. A data space architect utilizes the capability mapping matrix to translate canvas responses into specific technical capabilities. These "bottom-up capabilities" (purple in Fig. 4) directly address use case requirements. The subsequent section provides a comprehensive explanation of the mapping matrix and the full list of proposed data space capabilities.
3. The Governance Authority enhances this foundation by incorporating industry best practices and state-of-the-art solutions from the data space technical landscape. This evaluation produces complementary "top-down capabilities" (also shown in purple).
4. The final data space design emerges from the integration of both bottom-up and top-down capabilities, creating a comprehensive technical architecture that caters to the needs of the

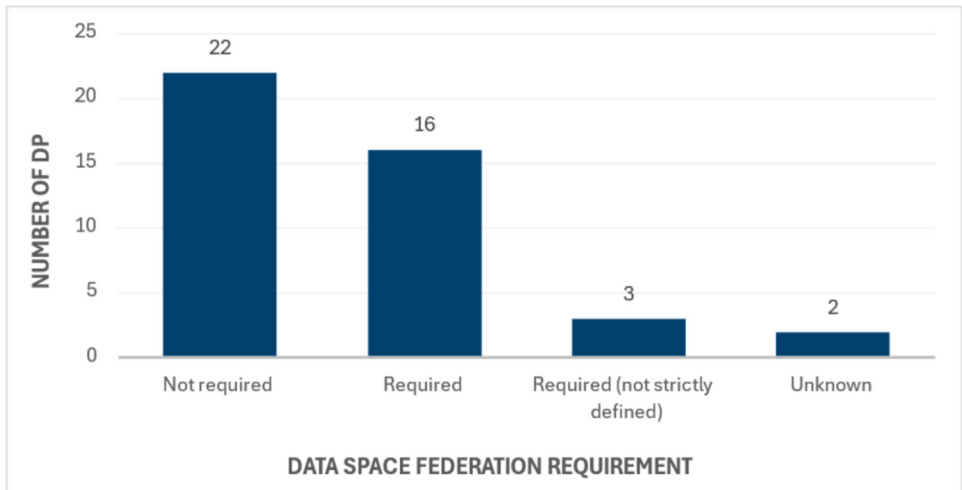


Fig. 4. Number of Data Products in the deployEMDS project that require a federation of different data spaces.

use cases that drive the data space, while also ensuring that it is supported by a sound, state-of-the-art technological infrastructure.

5.1. Component 1: intake canvas

The proposed intake canvas has four sections and an optional fifth section, containing 25 questions in total. The first section contains general info on the use case, such as a description and the actors involved. The second, third and fourth section respectively contain questions on the technical, governance and business aspects of the Data Product (DP). The optional fifth section on data space federation should be completed if the Data Product is offered to other data spaces. Extra information on those other data spaces can be added to the fifth section. The last question of each section is a self-assessment of the maturity of that specific aspect of the DP. This self-assessment of each aspect's maturity enables the participating organizations to identify their own strengths and weaknesses, prioritize improvements, and ensure balanced development across technical, governance, and business dimensions, ultimately leading to more robust and valuable data products. Besides a description of each subsection, the canvas also provides example answers and context information for each question. The canvas should be filled in for each Data Product of the use case.

The ontology of the canvas is based on combining concepts from the DSSC Blueprint v2.0 [4], which enhances alignment with other data (space) initiatives and projects, with the Open Data Product Specification v3.0 (ODPS), which is an open-source vendor-neutral specification that provides a comprehensive metadata model that describes the various technical, business, legal and ethical dimensions of data products [30,31]. The ODPS extends existing metadata standards such as DCAT and schema.org, with new attributes that are key for modern data marketplaces and the broader data economy. The correspondence between intake canvas questions and various Data Product attributes is illustrated through the matrix available in Table 1. When a cell in the matrix is filled, the answer to the question number in the column can directly or indirectly be mapped to the Data Product attribute in the corresponding row. These attributes are categorized according to the framework established by the ODPS. For the sake of conciseness, attributes lacking direct correlations have been excluded from this matrix. The resulting canvas is designed to not presuppose any specific architecture or framework of the data space, so it can be used for use cases in all domains, by anyone and with little data space expertise.

Table 1
ODPS v3.0 attribute mapping to canvas questions.

		Canvas questions																								
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
ODPS Data Product Attributes	Document-level Attributes																									
	details.name																									
	details.valueProposition																									
	details.description																									
	details.categories																									
	details.standards																									
	details.outputFileFormats																									
	details.recommendedDataProducts[i]																									
	Data Service Level Agreement																									
	SLA[i].dimension																									
	SLA[i].unit																									
	SLA[i].displaytitle																									
	Data Quality																									
	dataQuality[i].dimension																									
	dataQuality[i].unit																									
	dataQuality[i].displaytitle																									
	Data Pricing Plans																									
	pricingPlans[i].name																									
	pricingPlans[i].price																									
	pricingPlans[i].unit																									
	pricingPlans[i].offering[j]																									
	Data Licensing																									
	license.scope.definition																									
	license.scope.restrictions																									
	license.scope.geographicalArea																									
	license.governance.ownership																									
	license.governance.confidentiality																									
	license.governance.applicableLaws																									
	Data Ops																									
	dataOps.data.schemaLocationURL																									
	Data Access																									
	dataAccess.type																									
	dataAccess.authenticationMethod																									
	dataAccess.specification																									
	dataAccess.format																									

5.2. Component 2: data space capabilities

Table 2 offers an overview of thirty data space capabilities, divided over four main categories, that were used to map the canvas questions to. These capabilities are organized over four categories, “Trust framework” (TF), “Data, services and offering” (DO), “Data sovereignty” (DS) and “Product-enabling” (PE). These capabilities are proposed after reviewing five fundamental pieces of documentation for data space researchers and practitioners: the DSSC Blueprint v2.0 [4], the DSBA Technical Convergence document [32], the International Data Spaces Information Model [33], the ODPS v3.0 [30,31], and the European Commission Joint Research Centre report on the role of intermediaries in data spaces [34]. Every row in the table presents one capability, along with sources and a brief description explaining the extent of the functionalities that the capability aims to fulfil.

Next, Table 3 contains a matrix that maps the data space capabilities to the canvas questions. When a cell in the matrix is filled, the answer to the question number in the column can have an influence on the capability in the row. Questions 1, 12, 16, 20 and 21 cannot be mapped to a data space capability, as they ask for the name of the DP, the name of an additional data space, or the above-mentioned self-assessment. As the use cases influence the required data space capabilities, and not the other way around, we call this a bottom-up approach.

The data space capabilities inventory was initially developed in close collaboration with i2cat, IONOS, FIWARE, Fraunhofer, and NTTDATA, within the deployEMDS project consortium. This preliminary set of potential data space capabilities then underwent further external review by several organizations including the International Data Spaces Association (IDSA), ITxPT and Akkodis.

Table 2

List of suggested data space capabilities.

Data space capability	Source(s)	Description
Trust framework capabilities		
TF1 – EU Standards for IAA	DSBA, p. 46–51	A capability that implements European Union standards for Identification, Authentication, and Authorization (IAA). This enables standardized identity management across data spaces, ensuring compliance with EU regulations such as eIDAS.
TF2 – Legal persons identifiers	DSBA, p. 42	A system for uniquely identifying legal entities within the data space ecosystem. This capability supports the identification of legal persons through standardized identifier schemas.
TF3 – Organisation identifiers	DSBA, p. 49	A capability for managing organisational identity across data spaces, extending beyond legal identification to include organisational structures, roles, and relationships.
TF4 – Trust Service Providers	DSSC	A capability that integrates qualified trust service providers (TSPs) who issue and validate electronic signatures, seals, timestamps, and certificates. These providers deliver infrastructure for establishing digital trust.
TF5 – Verifiable presentation	DSBA, p. 52	A capability supporting the creation, issuance, and validation of verifiable credentials that can be presented as proof of claims, qualifications, or attributes.
TF6 – Self-issued identifiers	DSBA, p.52	A capability enabling participants to create and manage their own decentralized identifiers (DIDs) without reliance on centralized authorities. This supports self-sovereign identity principles, allowing entities to control their digital identities independently.
TF7 – Trusted Issuers List	DSBA, p.47	A centrally managed registry of organizations and authorities authorized to issue credentials and attestations within the data space ecosystem.
TF8 – Data space registry	IDSa DIM and DSSC	A capability that tracks and maintains the status and essential information of all entities participating in a data space ecosystem. This solution manages participant identities, credentials, and operational states, serving as the authoritative record of active members within the collaborative network.
TF9 – Trusted Participant List	DSBA, p.46	A registry of all trusted participants within a data space. This capability covers the participant onboarding process, verification of eligibility criteria, and ongoing compliance monitoring.
Data, services and offering capabilities		
DO1 – Service level negotiation	ODPS “Data SLA”	A capability enabling automated negotiation and formalisation of service level agreements (SLAs) between data providers and consumers. This includes, among others, mechanisms for expressing service level objectives, quality metrics, performance guarantees, and associated compensations for non-compliance.
DO2 – Notary	DSSC	A capability providing independent, trusted third-party verification and timestamping of transactions, agreements, and data exchanges within the data space. Functioning similarly to traditional notaries, this service creates records of interactions.
DO3 – Observability	DSSC and DSBA, p. 34	A comprehensive capability for monitoring, logging, and analysing data space operations to ensure transparency and accountability.
DO4 – Rating & Billing system	DBSA, p. 19	A capability managing economic transactions within the data space.
DO5 – Marketplace	DSSC, “Data marketplace”	A configurable platform that facilitates the exchange, monetization, and utilization of data. It has interactions with capabilities DO4 and DO7.
DO6 – Vocabulary hub	DSSC and DSBA, p. 22	A centralized repository for standardised vocabularies, ontologies, and semantic models used within the data space. This capability facilitates semantic interoperability by providing terminology definitions, relationships, and mapping between different semantic domains.

(continued on next page)

Table 2 (continued)

Data space capability	Source(s)	Description
DO7 – Catalogue	DSSC “Publication & Discovery – Functional specifications” DSBA, p. 57 “Metadata Broker”	A structured registry capability for publishing and discovering data assets, services, and offerings within the data space. This extends beyond simple listings to include metadata.
DO8 – Semantic interoperability	DSSC and DSBA p. 29 “Data interoperability”	A capability enabling meaningful data exchange by ensuring consistent interpretation of information across different systems and domains. This includes semantic mapping tools, ontology alignment services, and automated translation between different data models. It has interactions with capability DO6.
DO9 – Constructing self-descriptions	DSBA p. 56 and DSSC “Data, services and offerings descriptions”	A capability allowing participants to create machine-readable descriptions of their Data Products, Services, and Offerings characteristics. These self-descriptions enable functionalities such as automated discovery, compatibility assessment, and dynamic integration.
<u>Data sovereignty capabilities</u>		
DS1 – Policy transformation & implementation	DSSC and DSBA p. 55	A capability translating high-level data usage policies into enforceable technical rules across different systems and platforms. This enables the conversion of human-readable contracts and legal requirements into machine-executable policy objects
DS2 – Policy management & governance	DSSC	A comprehensive capability for creating, updating, versioning, and governing data usage policies throughout their lifecycle.
DS3 – Compliance tracking & enforcement proof	DSSC	A capability monitoring and documenting compliance with established data usage policies, generating auditable proof of enforcement.
DS4 – Policy negotiation	DSSC and DSBA p. 53	A capability facilitating the automated or semi-automated negotiation of data usage terms between providers and consumers. This enables dynamic agreement on acceptable conditions for data sharing and usage.
DS5 – Policy enforcement	DSSC and DSBA p. 53	A capability ensuring that all data access and usage within the data space adheres to agreed policies. This includes technical mechanisms for access control, usage restriction, obligation fulfilment, and policy-based data filtering.
DS6 – Local intermediaries	DSSC	A capability providing trusted local components that mediate interactions between participants and the broader data space infrastructure. These intermediaries enforce local policies, manage connections to external services, and provide a layer of isolation for internal systems.
DS7 – Proof of participation	DSBA, p. 45	A capability generating cryptographically verifiable evidence of a participant’s active involvement in the data space. This serves as proof of membership and compliance with data space requirements, for the purpose of access policy enforcement.
<u>Product-enabling capabilities</u>		
PE1 – Data exchange	DSSC and DSBA, p. 32 “Data exchange APIs”	A capability that establishes standardized methods for secure and semantically consistent data transfers within the data space.
PE2 – Data marketplace	DSBA, p. 58	A specialized marketplace capability focused exclusively on data assets, providing comprehensive functionality for data discovery, evaluation, acquisition, and delivery.
PE3 – Service marketplace	DSBA, p. 58	A specialized marketplace capability focused on data processing services, analytics capabilities, and data transformation offerings.
PE4 – App marketplace	DSBA, p. 58	A specialized marketplace capability focused on applications and software tools that operate within the data space ecosystem. This includes application discovery, deployment assistance, configuration management, and compatibility verification.
PE5 – Data intermediaries	JRC report	A capability enabling specialized entities to facilitate data sharing between participants who cannot or prefer not to establish direct relationships. This capability merges three types of data intermediaries identified by the JRC report: data cooperatives, data trusts and data unions.

Table 3
Capability mapping.

		Canvas questions																									
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
Data space capabilities	Trust framework capabilities																										
	TF1 – EU Standards for IAA																										
	TF2 – Legal persons identifiers																										
	TF3 – Organisation identifiers																										
	TF4 – Trust Service Providers																										
	TF5 – Verifiable presentation																										
	TF6 – Self-issued identifiers																										
	TF7 – Trusted Issuers List																										
	TF8 – Data space registry																										
	TF9 – Trusted Participant List																										
	Data, services and offering capabilities																										
	DO1 – Service level negotiation																										
	DO2 – Notary																										
	DO3 – Observability																										
	DO4 – Rating & Billing system																										
	DO5 – Marketplace																										
	DO6 – Vocabulary hub																										
	DO7 – Catalogue																										
	DO8 – Semantic interoperability																										
	DO9 – Constructing self-descriptions																										
	Data sovereignty capabilities																										
	DS1 – Policy transformation & implementation																										
	DS2 – Policy management & governance																										
	DS3 – Compliance tracking & enforcement proof																										
	DS4 – Policy negotiation																										
	DS5 – Policy enforcement																										
	DS6 – Local intermediaries																										
	DS7 – Proof of participation																										
	Product-enabling capabilities																										
	PE1 – Data exchange																										
	PE2 – Data marketplace																										
	PE3 – Service marketplace																										
	PE4 – App marketplace																										
	PE5 – Data intermediaries																										

The capabilities inventory has been periodically refined to reflect the evolving data space technological landscape and to incorporate guidance from updated reference documents, such as the DSSC Blueprint v2.0 released in March 2025.

6. Experimental Design, Materials and Methods

6.1. Methodological foundation and research design

This research adopts design science research (DSR) methodology as its foundational approach, following the guidelines of Hevner et al. (2004) [35] and Peffers et al. (2007) [36] for creating and evaluating IT artifacts. The framework represents a design artifact intended to solve the practical problem of translating use case requirements into data space technical specifications. The research combines DSR with participatory action research principles to ensure practitioner engagement and real-world validation.

The choice to adopt DSR was driven by three factors: (1) the need to create a novel artifact (the framework) rather than analyse existing phenomena, (2) the requirement for practical relevance in addressing real challenges in data space implementation, and (3) the iterative nature of framework development and refinement through stakeholder feedback. The participatory element was essential given the emerging nature of data spaces and the need to uncover the data space functional requirements in a bottom-up manner.

The data collection combined structured workshops with semi-structured collaborative analysis sessions. The implementation sites completed the intake canvases for each of their Data Products, which provided quantitative data, while participating in facilitated discussions with functional analysts and data space architects about capability requirements, which provided some additional qualitative insights and further refined the results of the intake canvases. This mixed approach, supported by Schuurman et al. (2007), ensured both systematic coverage and contextual understanding.

The DSR approach does have some limitations. First, the validation was limited to the urban mobility domain, which potentially limits the transferability of findings to other sectors. This domain bias was mitigated through domain-agnostic framework design principles and grounding the research in literature spanning multiple sectors. Second, the use-case-driven approach may have missed capabilities that are essential for minimum viable data spaces, creating bottom-up incompleteness. This limitation was addressed through complementary top-down analysis that incorporated additional governance requirements. Finally, the results emerging from the deployEMDS project context may not transfer effectively to other data space initiatives. This limitation was addressed by incorporating technology and domain neutrality principles into the framework design.

6.2. Framework application and validation

The data space design framework has been used for the requirements analysis of nine implementation sites in the deployEMDS project. This process produced 43 Data Products and Offering descriptions and identified 416 bottom-up requirements, which were later consolidated into 29 bottom-up capabilities.¹² In this section we will demonstrate how nine of those capabilities were identified, by analysing three intake canvas questions for two Data Products: the Multimodal Traffic Counts of the Flemish implementation site and the Traffic Data of the Barcelona implementation site. The general info of these Data Products can be found in Table 4 and Table 5, respectively.

The main goal of the Flemish implementation site is to optimize the re-use of multimodal traffic measurement data, by onboarding this data as a single Offering in the data space. Important to note is that this data is already onboarded in a data space, as can be seen in the context diagram in Table 4. The Vlaamse Smart Data Space (VSDS) is the Flemish local data space that has been put in place for use cases in the mobility and environmental domain. The Barcelona implementation site uses a similar Data Product for one of their use cases, but those traffic measurements are not multimodal and only count motorized traffic. This Data Product will be used for cross-analysis and prediction, as can be seen in Table 5.

To demonstrate how the framework identifies capabilities, we examine three governance-related questions from the Data Product intake canvas, which was introduced earlier in this work as the first component of the proposed data space design framework. Table 6 presents the answers from the Flanders and Barcelona implementation sites to questions 13 (data model selection), 14 (authentication and identification requirements), and 15 (access control requirements) for both Data Products. According to the mapping, question 13 can be used to identify DO6, DO8, DS6 and PE1; question 14 to identify TF1, TF2, TF3, TF5, TF8 and TF9; and question 15 to identify TF1, TF2, TF3, TF5, TF8, TF9, DS1, DS2 and DS7. When we put this theory into practice for the answers that the implementation sites provided, nine of those capabilities are identified in this case: PE1, DO6, DO8, TF5, TF8, TF9, DO3 and DS2 (Table 7). For question 13 the Flemish implementation site answered that they would like to use the OSLO (Open Standards for Linking Organizations) data model and mobility DCAT-AP metadata model, which implies that these (semantic) data models and interoperability with other data models should be supported

¹² The full report can be found on <https://deployemds.eu/knowledge-hub/publications/>. This report lists the resulting capabilities, but not the detailed mapping as is demonstrated in this section.

Table 4

General info of the multimodal traffic counts DP of the flemish implementation site.

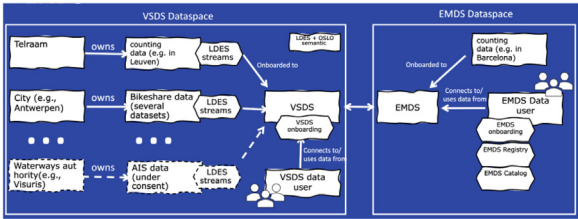
1. Data Product general info	
Implementation site	Flanders
Use case title	Optimizing the (re)-use of traffic measurements
Use case description	Today, many different ways (and standards) exist to exchange data, but most of them are bilateral and ad hoc or through a centralized hub. For traffic measurements specifically, this means that the data is locked up within silos: >500 entities use traffic measurements based on different technologies with their own protocols, which limits data re-use. The value chain from sensor producer to data analysis is linear and closed. Within this use case, we want to make the exchange of traffic measurements understandable, exchangeable, re-usable and future proof; by using standards, linked data, data space technology, building an ecosystem and a clear governance. Our goal is to exchange traffic counts through Linked Data Event Streams (LDES) across different countries, so that concepts and definitions have the same meaning in those countries.
Use case actors	imec (technical lead), Digitaal Vlaanderen (implementation)
Context diagram	
Data Product title	Multimodal Traffic Counts
Data Product description	"Multimodal Traffic Counts" is a data product designed to provide comprehensive, integrated traffic data across various modes of transportation (foot, bike, car, truck and potentially ship) for traffic managers, city administrations, researchers, and other stakeholders. The offering contains nine data sources on traffic counts (including vehicular (cars, buses, trucks), pedestrian, and bicycle traffic). This data is gathered through a mix of sources such as sensors, cameras, both temporary as permanent.

Table 5

General info of the traffic data DP of the Barcelona implementation site.

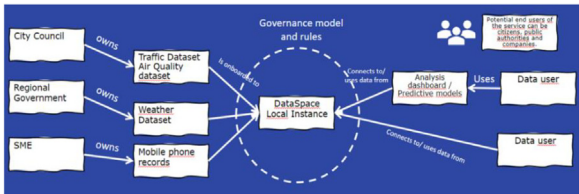
1. Data Product general info	
Implementation site	Barcelona
Use case title	Forecasting system to optimize traffic, based on vehicle flow and air quality
Use case description	This use case will combine several data sources and data models to generate value from traffic data through cross-analysis and prediction. The effect of weather on traffic and the impact of traffic on air quality will be analyzed. Furthermore, a traffic prediction and incident detection model will be developed.
Use case actors	Eurecat (traffic use case lead and data solution provider), Barcelona city council (data owner), regional government (data owner)
Context diagram	
Data Product title	Traffic Data
Data Product description	Traffic information from sensors. This is the number of vehicles detected every 15 min in the position where the sensor is located.

Table 6
Questions 13 to 15 of the intake canvas, as answered by the Flemish and Barcelona implementation sites.

2. Governance			
Background	The information in this section will help identify standard practices in data management and product compliance, including industry/domain standards and governance models relevant for various use cases. Data product owners should outline their trust-building processes, which will guide our decision on supporting identity management and data sovereignty, and to see if a fully managed trust model is needed by the data space.		
No	Question to be answered	Answer Flanders	Answer Barcelona
13	Which data model (if any) would you like to use?	OSLO and mobilityDCAT-AP.	None for the moment.
14	Are there any requirements for authentication and identification of participants?	Authentication and identification in the EMDS must be aligned with the control plane of our local data space (VSDS).	None for the moment. However, it could be used in the future to understand the use of the DP.
15	Are there any requirements for access control to the data product?	We are considering to only allow consuming data, when you also provide data yourself.	None for the moment.

Table 7
Capabilities identified based on the answers of the Flemish implementation site for questions 13 to 15.

Implementation site	Question number	Capability code	Capability name	Detailed capability
Flanders	13	PC1	Data exchange	MobilityDCAT-AP metadata model and OSLO data model must be supported
Flanders	13	DO6	Vocabulary hub	Must be in place and provide the OSLO vocabulary
Flanders	13	DO8	Semantic interoperability	The data space must be able to semantically link the OSLO vocabulary to other traffic counting standards
Flanders	13	DO8	Semantic interoperability	MobilityDCAT-AP metadata model and OSLO data model must be supported
Flanders	14	TF5	Verifiable presentation	Authentication must be verifiable and interoperable with (to be defined) authentication in VSDS
Flanders	15	TF8	Data space registry	A central authority does the identification
Flanders	15	TF9	Trusted Participants List	It must be possible to only give access to participants who also share data
Barcelona	14	DO3	Observability	Basic logging of transactions needed to understand who is using the data
Barcelona	15	DS2	Policy management & governance	Policy needed to support open/unlimited data usage

by the data space, e.g., by a vocabulary hub (DO6). The Barcelona implementation site, on the other hand, does not plan to use a data model, so the same capabilities are not required for their use case. For question 14 on identification of participants, the Flemish implementation site refers to the identification that will be in place in their local data space. This implies that the identification process of the new data space should be interoperable with the one in the local data space (TF5). Furthermore, The Flemish implementation site considers to only allow participants to consume data, if they also provide data themselves. To enable this access control, the data space needs a registry and a Trusted Participant List (TF8 and TF9). The Data Product of the Barcelona implementation site does not require identification nor access control, which means that this is open data, so the data space will need a policy to support the use of open data (DS2). However, in the future they would be interested in logging who uses the data, which requires some form of observability (DO2). This demonstration shows that the framework can identify different capabilities for similar Data Products, depending on the requirements for the use case.

7. Discussion

This research contributed to existing literature by creating a bottom-up data space design framework, a tool to identify the data products of use cases and the required data space capabilities to enable these use cases. The two components of the framework (i.e., the canvas and capability mapping) are designed to meet the need for practical tools that can support the decision for technical data space components [7,8]. To the best of our knowledge, no such framework exists today. The framework aligns with the concepts and principles of the DSSC Blueprint [4], and it complements the blueprint by building the bridge between use cases from the field and conceptual data space components. Moreover, the framework is both grounded in scientific methods, e.g. for capability driven design, and validated in the field, i.e. in the deployEMDS project, which makes it a fundamental addition to the rather limited body of data space research [8]. As data spaces are being put forward as the preferred mechanism for trusted inter-organizational data sharing (e.g., by the EC), we believe that the framework might be of value for anyone initiating a data space project.

According to the framework of Antunes & Tate [37], a canvas consists of two levels: a surface level, which focuses on representing concepts by using ontological components and their implicit relationships; and a deeper level, which focuses on theorizing about these concepts with a holistic view and at an intermediate level of abstraction. Antunes & Tate [37] acknowledge the frequent use of canvases in many domains (e.g., business modelling [38] and innovation [39]), fostered by the need for a lightweight conceptualization, planning, communication and co-creation tool for dynamic environments. A canvas allows to create a concrete representation of a complex phenomenon, which makes it very useful for data space design. The data cooperation canvas,¹³ which was created during the preparatory action of the Data Space for Smart and Sustainable Cities and Communities (DS4SSCC), is a good example. The canvas for the deployEMDS data space aligns with the framework of Antunes & Tate [37], as it contains the two levels that they mention. The first level is the surface level. It aims to create a representation of a Data Product. The canvas introduces an ontology on the main subcomponents of a Data Product, such as governance, and links it with other concepts, such as data source. This representation helps the local implementation sites to understand what a Data Product is, and why it is essential to bridge the gap between a use case and a data space. Secondly, at the deeper level, the canvas lets the implementation sites theorize about the Data Products they need to accomplish their use cases. It stimulates them to form a holistic view of their Data Products, by asking them to identify the needed technology, data sharing protocols and access control mechanisms at an intermediate level of abstraction. Our findings align with the work of Antunes & Tate [37]), which states that a canvas helps to cope with a dynamic, complex environment. By focusing on the concept of a Data Product, the implementation sites learned how their use cases will be implemented in the data space. These insights led some implementation sites to refine their initial use case, adapting it more effectively to the data space way of working.¹⁴ As this approach was proven to be successful, it is also being used in several mobility data space trainings for a broad audience.¹⁵

A second benefit of the data space design framework is that it generates not only a lot of insights into the required data space capabilities, but also in the use cases and their Data Products themselves.¹⁶ These insights are crucial in the early stages of any data space project. They provide an initial impression of how the data space will take shape, such as the scope of the data to be onboarded, but more importantly, they highlight aspects of the use cases that are still unclear and require further definition. For example, one of the insights obtained by using

¹³ <https://www.datacooperationcanvas.eu/>.

¹⁴ See also the take aways in Chapter 7 of the deployEMDS deliverable D2.1: <https://deployemds.eu/wp-content/uploads/2024/05/D2.1-Requirements-analysis-of-the-technical-infrastructure-3.pdf>.

¹⁵ <https://factual-academy.com/course/mobility-data-spaces>, <https://factual-academy.com/course/deployemds>.

¹⁶ See chapter 4.2 of the deployEMDS deliverable D2.1 for all insights: <https://deployemds.eu/wp-content/uploads/2024/05/D2.1-Requirements-analysis-of-the-technical-infrastructure-3.pdf>.

the framework in the deployEMDS project is shown in Fig. 4. This graph shows that for a large amount of Data Products federation is required into multiple data spaces, meaning that these Data Products will be onboarded in more than one data space (as is the case for the 'multi-modal traffic counts' discussed in the previous section). The graph also shows that for five Data Products this is not yet known or strictly defined. This was an important insight for the project, as it entailed that the deployEMDS data space would be the first data space federation of that size in the mobility domain.

8. Limitations and Future Research

While the framework has been validated for a mobility data space, all questions are domain agnostic, so the framework can be used for data space design in any domain. Additionally, the framework does not presuppose any data space reference architecture (e.g., IDSA or Gaia-X), so the capabilities can afterwards be implemented in any architecture by any technology. However, both the use outside of the mobility domain and the translation of the capabilities to data space reference architectures and technologies have not yet been validated. The latter is defined by Berziša et al. (2015) [29] as a 'capability delivery pattern', which describes how a certain capability should be delivered within a certain context and what processes and resources are needed. While the data space design framework incorporates the first two pillars of capability driven design; i.e., enterprise and capability modelling, and capability delivery context modelling; it deliberately omits capability delivery patterns to maintain technology neutrality. Defining the capability delivery patterns is a focus of future research to be undertaken in the deployEMDS project. The planned course of action is to define customer journeys, based on the identified capabilities, and to assess the performance of different data space technologies for these customer journeys. This approach allows to find the most mature technology, i.e., the one that has most of the required capabilities, but also to identify gaps, i.e., capabilities that are not provided yet and need to be developed in the project.

The bottom-up approach of the framework (i.e., by starting from the use cases) generated substantial benefits for the implementation sites, but it also has an important limitation: not all capabilities of a minimum viable data space (MVDS) may be identified through a bottom-up approach. The DSSC refers to the IDSA for defining an MVDS, which is described as "a combination of components to initiate a data space with just enough features to be usable for secure and sovereign data exchange." An MVDS reduces the implementation time of a data space project by offering an initial functional version for the development team to refine, assess, and address the assumptions regarding the data space requirements, and it is therefore often used as a best practice in data space projects. According to the IDSA, in an MVDS, every data provider should be able to set usage policies for their Data Products, which can be enforced either legally or technically, e.g., by deleting data after a specified period. The downside of a bottom-up approach for data space design is that it only identifies data space capabilities needed for specific use cases. Consequently, if a use case does not require usage policies, such as when only open data is exchanged, these capabilities may not be recognized by the framework. This could result in the data space not meeting the MVDS definition. This can be remediated by complementing the bottom-up data space capabilities, originating from the use cases, with top-down capabilities, implied by the governance authority of the data space, which are essential to reach an MVDS. To identify top-down capabilities, it is crucial to determine how the strategic roadmap of the data space will be managed, and by whom. Top-down capabilities are data space specific, rather than use case specific and can for example include compliance to existing technical reference frameworks, the business model, and the governance of the data space. This approach was tested and deemed successful during the validation of the framework in the deployEMDS project, by complementing the identified 29 bottom-up capabilities with 44 additional top-down capabilities. A selection of top-down capabilities is shown in Table 8. We recommend that all founding partners of a data space first agree on the governance authority and organizational form, and on the definition of an MVDS and its required minimum capabilities. After using the data space de-

Table 8

Selection of top-down capabilities that were added in the deployEMDS project.

Capability code	Capability name	Detailed capability
TF1	EU Standards for AAI	The technology stack that implements credentials and their mechanisms should be interoperable with evolving frameworks that support the EU Identity framework: eIDAS credential attestation ^a and the EBSI VC framework ^b .
DS1	Policy transformation & implementation	Choose a Policy Definition Language that has sufficient expressivity to cover most of the Usage Control use cases.
DO7	Catalogue	Each Participant exposes a Local Catalogue of Self-Descriptors, which can be made public or protected via an access policy. The Local Catalogue is accessible through a standardized Catalogue API hosted on Participant's Connector.
DO3	Observability	The data space incorporates a comprehensive logging infrastructure that centralizes logs from Participants' Connectors, the Data Space Registry, and the Catalogue. While optional, this log collection capability is essential for enabling value-added services such as clearing houses, QoS monitoring, and marketplaces. The framework mandates a standardized collection mechanism with potential deployable agents to accommodate diverse sources. All logs are secured with transport and storage integrity controls and maintained in audited storage under the Data Space Governance Authority's management.

^a <https://digital-strategy.ec.europa.eu/en/library/european-digital-identity-wallet-architecture-and-reference-framework>.

^b <https://hub.ebsi.eu/vc-framework>.

sign framework to identify bottom-up capabilities, it should be assessed if additional top-down capabilities are required according to the MVDS definition.

A third limitation of the framework is that its focus is restricted to the business, technical and governance components of a Data Product, as defined by the conceptual model of the DSSC. This focus is deliberately kept narrow, as this was the best fit for the project, but it might be that other data space projects need to gather additional information from their use cases, for example, when the use cases are very mature. Other potentially interesting additions to the framework could involve the infrastructure or the underlying technical stacks that host data sources and data space components of a participant, the (quality of) data sources and the needed transformations, and the lifecycle or long-term implementation and mechanisms that are needed for change management of a data product. We intend to apply the framework, as presented in this work, to the design of additional data spaces as part of our future work, which will provide a more comprehensive assessment of its robustness and completeness.

9. Conclusion

In this paper, a comprehensive framework has been presented to address the complexities and challenges of data space design. The framework, tested in the deployEMDS project, provides essential tools for identifying data space capabilities based on specific use cases. The approach involves an intake canvas for defining functional requirements and a capability mapping process for technical design.

With the framework established in this study, the two research questions posed in this paper can be answered as follows. Regarding RQ1 "How should suitable technical infrastructure be selected for data sharing in a data space, based on concrete use cases?", the framework provides a systematic bottom-up methodology that translates domain-specific use case requirements into concrete technical capabilities through its intake canvas and capability mapping components. For RQ2 "What framework can bridge the gap between domain-specific use cases and the requirements of data spaces to support trusted data sharing?", the proposed dual-process approach, which combines the intake canvas for capturing functional requirements with capability mapping for technical translation, effectively serves this bridging function while maintaining alignment with established data space principles and standards.

Despite the notable benefits, such as improved comprehension of data space concepts and detailed insights into use cases and required capabilities, the framework also has limitations. Specifically, a purely bottom-up approach may not identify all necessary capabilities of a minimum viable data space (MVDS). Therefore, a combined approach that includes top-down considerations is recommended to ensure comprehensive data space functionality. While initially developed for the mobility domain, the framework is versatile and applicable across various domains and technologies, offering a robust tool for the preparatory stages of data space co-creation.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirming that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

Casper Van Gheluwe: Conceptualization, Methodology, Formal analysis, Investigation, Writing – review & editing; **Gabriele Bozzi:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – review & editing; **Eridona Selita:** Conceptualization, Methodology; **Nele Daels:** Conceptualization, Methodology, Writing – review & editing; **Tanguy Coenen:** Writing – review & editing; **Laure De Cock:** Conceptualization, Methodology, Writing – original draft.

Data Availability

No research data (Original data) (No research data).

Acknowledgments

We would like to thank the Flemish and Barcelona implementation sites of the deployEMDS project, and especially the partners Digitaal Vlaanderen, Movias and Eurecat, as they provided the input for the use case demonstration.

This work has received funding from the Flemish Government under the “Onderzoek-sprogramma Artificiele Intelligentie (AI) Vlaanderen” programme, and was co-funded by the [European Union](#) under Grant Agreement No [101123520](#). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union or the granting authority can be held responsible for them.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Laure De Cock reports financial support was provided by the Flemish government and the European Union. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendices

Intake canvas

1. DP general info				
Implementation site				
Use case title				
Use case description				
Use case actors				
Context diagram				
DP title				
DP description				
2. Functional and technical				
Background	The data product is the implementation of the data product in the data space and determines how a data product becomes a data asset. The data product will be implemented in the data space connector, and handles usage control, formats, data assets publication, data asset catalogue, ... In other words, it defines how the data product is offered to the data space.			
No	Question to be answered	Answer	Example answer	Context
1	Can you provide a name for your data product?		Bus real-time positioning	Define/specify the data product being analysed in this sheet.
2	Can you provide a functional description of your data product?		50 data sources of GPS traces, 1 for each bus in Lisbon, with a time interval of 1 min and going back 1 year in time.	Functional description of the data product.
3	What is the scope of the data product?		Bus	Used for grouping and quantitative analysis.
4	What is the geographical data product scope?		Lisbon Metropolitan area	Used for grouping and quantitative analysis.
5	What is your data product type?		1) The project partners publish the data directly onto EMDS, as they are data owners	1) The data owner publishes the data product in the data space (without intermediary service) 2) The data owner onboards the data product on an intermediary offering 3) You are yourself a data intermediary (see glossary)
6	What are the data sources that will use this kind of data product?		Li.03.04	A data product can be an aggregation of more than one data source (and a data source can have multiple data products), therefore it's best to standardise the data product in a way that it can be reused. To answer this question, please refer to the dataset Nr. in the overview excel.
7	Is there personal data involved in your data product or are there any ethical considerations?		This is personal data when it comes to the bus driver, so they have the option to exclude their data from the data source.	Determine if mechanisms for consent or traceability are required when the data product includes personal data.

(continued on next page)

8	Is your current data model (if any) in conformance with existing standard(s)? If yes, which?	GTFS-RT, DCAT	This question refers to the conformity of the data sources at present. Possible answers are for example: OSLO, MMTIS, MDS, DATEX-II, TOMP, GTFS, GTFS-RT, NeTex, ... or there is no standard used now. If there are any current metadata standards (DCAT, DCAT-AP, Mobility DCAT-AP, DOI, ...) in use, you may also list them here.
9	What is the underlying protocol used to transfer your data product?	The data can be accessed through a REST API	We need to know which network protocols are used to infer which data planes will be necessary in deployEMDS. Examples are Kafka topic, LDES, HTTP, TSP, Websocket API, FTP, SSH, SFTP etc....
10	When integrating a data product (back end) with a connector, are there specific access control mechanisms or workflows needed to publish the data product?	You need an API key to access the data that is granted by the transport authority after a thorough screening.	We need to know which access control mechanisms (which might be specific for your organisation) are needed to decide how the control plane (see glossary) should look like.
11	Are there performance concerns about your data product such as latency, high volumes, certain service level agreement(s)?	After a latency of 10 min the GPS traces are available through the API	The current SOTA of data space connectors is that they are vertically scalable, but they are not designed to manage hyperscale architectures. An out of band data plane overcomes scalability issues. We need to understand the effective needs or if there are sufficient reasons for more experimentation and testing.
12	Self asses your data product maturity.	3: We have already a lot of experience with the data product, but not with the data product. The data owner is on board.	We would like you to think carefully on the maturity of the data product and to indicate this on a scale from 1 (not discussed yet) to 5 (all partners agree on the roadmap). Please also elaborate on why you chose this number. This self-assessment will allow us to identify mature and early-stage components of the data product. The results can be used to prioritise certain building blocks, organise trainings, cluster implementation projects, etc.

(continued on next page)

3. Governance

Background The information in this section will help identify standard practices in data management and product compliance, including industry/domain standards and governance models relevant for various use cases. Data product owners should outline their trust-building processes, which will guide our decision on supporting identity management and data sovereignty, and to see if a fully managed trust model is needed by the data space.

No	Question to be answered	Answer	Example answer	Context
13	Which data model (if any) would you like to use?		GTFS-RT, mobility DCAT-AP	This question refers to the desired conformity of the data sources. Possible answers are for example: OSLO, MMTIS, MDS, DATEX-II, TOMP, GTFS, GTFS-RT, NeTeX, ... or none.
14	Are there any requirements for authentication and identification of participants?		Participants must be able to prove their nationality, certified by the government of that country	We need to know if participants are bound to specific identity management governance or processes and verifications when they want to use the data product. Think of claims participants could need, specific identity standards they must adhere to (e.g., EIDAs), domain registries that can identify them (e.g., a company registration registry). Which certification bodies and identity providers are involved?
15	Are there any requirements for access control to the data product?		Only European citizens may access this data.	We need to know which claims can be verified by which trust anchor.
16	Self assess your data product governance maturity.		3: up until now we have only collected open data, so little experience with governance, but the team agrees on the roadmap.	We would like you to think carefully about the maturity of the data product governance and to indicate this on a scale from 1 (not discussed yet) to 5 (all partners agree on the roadmap). Please also elaborate on why you chose this number. This self-assessment will allow us to identify mature and early-stage components of the data product. The results can be used to prioritise certain building blocks, organise training, cluster implementation projects, etc.

(continued on next page)

4. Business model(s)				
Background	The information in this canvas will help determine how diverse the participants' business expectations are. Some business models require specialised components at data space level. We also want to determine if a shared business model is viable, or rather fostering an eco-system of value returns. The data space should provide the necessary building blocks to support this diversity.			
No	Question to be answered	Answer	Example answer	Context
17	Can you specify your expected/intended business model for the data product and relevant restrictions?		Portuguese citizens can access this data product freely, but participants of other European countries pay per volume.	The goal is to determine specific supporting infrastructure for enabling the intended business model. Possible answers are open public, open subscription, value exchange, recurring time, period based, one-time payments, pay-as-you-go, revenue sharing, data volume, dynamic pricing, freemium, value-based, on Request, ...
18	Does the data product fall under local legislation that might differ from and/or precede over EU and foreign legislation?		In Portugal there is a law that says when buses are not on time in 20 % of the cases, the transport authority gets a fine. This is yearly checked by the government by manually timing, so not using this data.	The GDPR applies to the processing of personal data of living individuals in the EU, but the implementation might slightly differ in each member state (e.g., in Italy the heirs of a deceased person can request data erasure, in Belgium they cannot). Additionally, there might be rules in place in your country that differ from other countries (e.g., the age of legal majority). In other words, we need to know which local legislation is in place that requires specific data space functionalities.
19	What is the value proposition (if any) of the data product?		This data product will be used to inform passengers, PTAs, municipalities and other stakeholders on delays, which will improve the travel experience but also allows for a thorough analysis and possible re-organization. Up until now this was closed data so all stakeholders will be very interested in accessing it.	Why is this data product interesting for the community, what value does it bring to the ecosystem? Why will participants want this data product?
20	Self asses your business model maturity.		1: The public transport operators are not aligned.	We would like you to think carefully about the maturity of your business model and to indicate this on a scale from 1 (not discussed yet) to 5 (all partners agree on the roadmap). Please also elaborate on why you chose this number. This self-assessment will allow us to identify mature and early-stage components of the data product. The results can be used to prioritise certain building blocks, organise trainings, cluster implementation projects, etc.

(continued on next page)

5. Optional: data space federation				
Background	There are cases where data products are already part of an existing data space. The information must help us to find the most flexible and sustainable way to integrate them in the data space.			
No	Question to be answered	Answer	Example answer	Context
21	In which other data space is your data product included?		Local open data space	Name the data space. Describe if managing or part of an existing data space. Describe the data space standard used.
22	Does this data space have an asset catalogue? And if so, which technology does it use?		The open data catalogue of the public transport authority will be used, it can be accessed through API	We need to ensure that existing asset catalogue technologies and standards can interoperate with the deployEMDS asset catalogue.
23	Is there a data space mechanism to univocally identify participants? If so, which one?		Participants are registered by creating an account on the data space portal.	We need to ensure that the mechanisms to identify participants in your existing data space can interoperate with the deployEMDS systems.
24	How is trust established in this data space?		Each participant is manually screened by the data space authority.	We need to know if a trust anchor is involved, how credentials are issued, which authorities are involved, etc. This is to ensure that the data space can interoperate with these trust mechanisms.
25	Does the data space implement data usage control policies and enforcement? Does it use standards?		GTFS-RT data model, no metadata model, the control policy is now done manually, and each request is individually assessed before access to the API is granted.	We need to ensure that the existing usage control policies and enforcement methods are compatible with the deployEMDS mechanisms.

References

- [1] B. Otto, M. ten Hompel, S. Wrobel, Designing Data Spaces, Springer International Publishing, 2022, doi:[10.1007/978-3-030-93975-5](https://doi.org/10.1007/978-3-030-93975-5).
- [2] European Commission, "A European strategy for data," Brussels, 2020.
- [3] European Commission, "Commission staff working document on common European data spaces," Brussels, 2022.
- [4] Data Spaces Support Centre, "Data spaces blueprint v2.0." Accessed: Apr. 04, 2025. [Online]. Available: <https://dssc.eu/space/BVE2/1071251457/Data+Spaces+Blueprint+v2.0+-+Home>
- [5] CEN, "CEN workshop agreement (CWA 18125)," Brussels, Jul. 2024.
- [6] European Commission, "Commission staff working document on common european data spaces," Brussels, 2024.
- [7] I. Jussen, F. Möller, J. Schweihoff, A. Gieß, G. Giussani, B. Otto, Issues in inter-organizational data sharing: findings from practice and research challenges, Data Knowl. Eng. 150 (2024), doi:[10.1016/j.datak.2024.102280](https://doi.org/10.1016/j.datak.2024.102280).
- [8] D. Beverungen, T. Hess, A. Köster, C. Lehrer, From private digital platforms to public data spaces: implications for the digital transformation, Electron. Mark. 32 (2) (2022) 493–501, doi:[10.1007/s12525-022-00553-z](https://doi.org/10.1007/s12525-022-00553-z).
- [9] F. Möller, et al., Industrial data ecosystems and data spaces, Electron. Mark. 34 (1) (2024), doi:[10.1007/s12525-024-00724-0](https://doi.org/10.1007/s12525-024-00724-0).
- [10] European Commission, "Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions, creation of a common European mobility data space," Brussels, Nov. 2023.
- [11] B. Otto, Quality and value of the data resource in large enterprises, Inf. Syst. Manag. 32 (3) (2015) 234–251, doi:[10.1080/10580530.2015.1044344](https://doi.org/10.1080/10580530.2015.1044344).
- [12] C. Cappiello, A. Gal, M. Jarke, J. Rehof, R. Aachen, T.U. Dortmund, Data ecosystems: sovereign Data exchange among organizations, Dagstuhl Rep 9 (9) (2020) 66–134, doi:[10.4230/DagRep.9.9.66](https://doi.org/10.4230/DagRep.9.9.66).

- [13] J. Pennekamp, et al., Towards an infrastructure enabling the internet of production, in: Proceedings - 2019 IEEE International Conference on Industrial Cyber Physical Systems, ICPS 2019, 2019, doi:[10.1109/ICPHYS.2019.8780276](https://doi.org/10.1109/ICPHYS.2019.8780276).
- [14] O.A. El Sawy, A. Malhotra, Y.K. Park, and P.A. Pavlou, "Research commentary—Seeking the configurations of digital ecodynamics: it takes three to tango," <https://doi.org/10.1287/isre.1100.0326>, vol. 21, no. 4, pp. 835–848, Nov. 2010, doi: [10.1287/isre.1100.0326](https://doi.org/10.1287/isre.1100.0326).
- [15] J. Lohmöller, et al., The unresolved need for dependable guarantees on security, sovereignty, and trust in data ecosystems, *Data Knowl Eng* 151 (2024), doi:[10.1016/j.datak.2024.102301](https://doi.org/10.1016/j.datak.2024.102301).
- [16] E. Curry, *Real-time Linked Dataspaces*, Springer International Publishing, Cham, 2020, doi:[10.1007/978-3-030-29665-0](https://doi.org/10.1007/978-3-030-29665-0).
- [17] M. Franklin, A. Halevy, D. Maier, *From databases to dataspace: a new abstraction for information management*, *ACM Sigmod Rec.* 34 (4) (2005) 27–33.
- [18] S. Opriel, F. Möller, U. Burkhardt, B. Otto, Requirements for usage control based exchange of sensitive data in automotive supply chains, in: Proceedings of the Annual Hawaii International Conference on System Sciences, IEEE Computer Society, 2021, pp. 431–440, doi:[10.24251/hicss.2021.051](https://doi.org/10.24251/hicss.2021.051).
- [19] M.I.S. Oliveira, G. de F. Barros Lima, B. Farias Lóscio, Investigations into data ecosystems: a systematic mapping study, *Knowl Inf Syst* 61 (2) (2019) 589–630, doi:[10.1007/s10115-018-1323-6](https://doi.org/10.1007/s10115-018-1323-6).
- [20] J. Zrenner, F.O. Möller, C. Jung, A. Eitel, B. Otto, Usage control architecture options for data sovereignty in business ecosystems, *J. Enterp. Inf. Manag.* 32 (3) (Jun. 2019) 477–495, doi:[10.1108/JEIM-03-2018-0058](https://doi.org/10.1108/JEIM-03-2018-0058).
- [21] H. Drees, D.O. Kubitz, J. Theissen-Lipp, J. Lipp, S. Pretzsch, and C.S. Langdon, "Mobility data space—first implementation and business opportunities," 2021.
- [22] M. Jarke, "Culture data space: a case study in federated data ecosystems," in *JointWorkshops At 49th International Conference On Very Large Data Bases (VLDBW'23) — Data Ecosystems (DEco)*, Vancouver, 2023. [Online]. Available: <http://ceur-ws.org>
- [23] A. Gies, M. Hupperz, T. Schoormann, F. Möller, "What does it take to connect? Unveiling characteristics of data space connectors, in: Proceedings of the 57th Hawaii International Conference on System Sciences, University of Hawaii at Manoa, 2024.
- [24] F. Schafer, J. Rosen, C. Zimmermann, and F. Wortmann, "Unleashing the potential of data ecosystems: establishing digital trust through trust-enhancing technologies," in *ECIS 2023 Research Papers*, Kristiansand, 2023.
- [25] I. Akaichi, et al., Interoperable and continuous usage control enforcement in dataspace, *The Second InternationalWorkshop on Semantics in Dataspace*, co-located with the Extended SemanticWeb Conference, 2024 [Online]. Available <http://ceur-ws.org>.
- [26] European Commission, "Sustainable and Smart mobility Strategy – putting European transport on track for the future," Brussels, 2020.
- [27] M. Poniatowski, H. Lüttenberg, D. Beverungen, D. Kundisch, Three layers of abstraction: a conceptual framework for theorizing digital multi-sided platforms, *Inf. Syst. e-Bus. Manag.* 20 (2) (Jun. 2022) 257–283, doi:[10.1007/s10257-021-00513-8](https://doi.org/10.1007/s10257-021-00513-8).
- [28] A. Giess, F. Möller, T. Schoormann, B. Otto, *Design options for data spaces, Thirty-first European Conference on Information Systems (ECIS 2023)*, 2023.
- [29] S. Berziša, et al., Capability driven development: an approach to designing digital enterprises, *Bus. Inf. Syst. Eng.* 57 (1) (2015) 15–25, doi:[10.1007/s12599-014-0362-0](https://doi.org/10.1007/s12599-014-0362-0).
- [30] J. Moilanen, J. Niilahti, T. Luhti, T. Santakivi, and A. Loukiala, "Open data product specification v3.0," <https://opendataproducts.org/v3.0/#open-data-product-specification-3-0>.
- [31] J. Moilanen, "Open data product specification," 2022.
- [32] Data spaces business alliance, "Technical convergence discussion document," 2023.
- [33] C. Mader, J. Pullmann, N. Petersen, S. Lohmann, and C. Lange-Bever, "International Data spaces information model," <https://international-data-spaces-association.github.io/InformationModel/docs/4.2.0/index.html>.
- [34] M. Micheli, E. Farrell, B. Carballa-Smichowski, M. Posada-Sánchez, S. Signorelli, and M. Vespe, "JRC Science For Policy Report, Mapping the landscape of data intermediaries emerging models for more inclusive data governance," Luxembourg, 2023. doi: [10.2760/8943](https://doi.org/10.2760/8943).
- [35] A. Hevner, S. March, J. Park, S. Ram, *Design science in information systems research*, *MIS Q.* 28 (1) (2004) 75–105.
- [36] K. Peffers, T. Tuunanen, M.A. Rothenberger, S. Chatterjee, A design science research methodology for information systems research, *J. Manag. Inf. Syst.* 24 (3) (2007) 45–77, doi:[10.2753/MIS0742-1222240302](https://doi.org/10.2753/MIS0742-1222240302).
- [37] P. Antunes, M. Tate, Examining the canvas as a domain-independent artifact, *Inf. Syst. e-Bus. Manag.* 20 (3) (2022) 495–514, doi:[10.1007/s10257-022-00556-5](https://doi.org/10.1007/s10257-022-00556-5).
- [38] A. Osterwalder, PhD thesis, *Université de Lausanne*, 2004.
- [39] D. Schuurman, A.-L. Herregodts, A. Georges, O. Rits, *Innovation management in Living Lab projects: the Innovatrix framework*, *Technol. Innov. Manag. Rev.* 9 (3) (2019) 63–73 [Online]. Available <http://leanstartupmachine.com>.