Data in Brief 62 (2025) 111949

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Perspective Article

A perspective on the use of personal data stores in data spaces to support multiparty data access as required by the data Act

Tanguy Coenen [a,*], Michiel Fierens [b,*], Beatriz Esteves [c], Esther De Loof [c]

[a] imec, AI and algorithms department, Leuven, Belgium
[b] KULeuven, Center for IT and IP Law, Leuven, Belgium
[c] imec-UGent, Knoweldge at web scale department, Ghent, Belgium

ARTICLE INFO

Article history:
Received 19 December 2024
Revised 9 June 2025
Accepted 28 July 2025
Available online 20 August 2025

Dataset link: Kvasir (Original data)

Keywords:
Data Act
Personal data store
Data space

ABSTRACT

This perspective article explores the requirements arising from the EU's Data Act for multiparty data access in connected products, a key element of the EU's Data Strategy. We categorize these requirements into six dimensions: Data accessibility for users, Information transparency for users, User rights regarding data, Fair and transparent data sharing, Interoperability, and Protection of legitimate interests. For each category, we provide detailed explanations and propose a conceptual model to address them. Central to this model are technological concepts such as identity management, consent management, data access policies, data usage policies, data catalogues, confidential compute, and linked data. We analyze how these technologies can support the Data Act's requirements and integrate them into a cohesive architecture.
To ground these insights, we present an "agentic home" use case, illustrating the deployment of personal data stores in edge environments to enable generative AI applications. This application exemplifies the transformative potential of combining personal data in data spaces with generative AI, showcasing a pathway for the EU to maintain a competitive position in AI while upholding ethical norms and values. We argue that by leveraging the data space paradigm in relation

* Corresponding authors.
E-mail addresses: tanguy.coenen@imec.be (T. Coenen), michiel.fierens@kuleuven.be (M. Fierens).

https://doi.org/10.1016/j.dib.2025.111949
2352-3409/© 2025 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/)

to personal data and within the provisions of the Data Act, the EU can capitalize on generative AI as the "killer app" for data spaces, driving innovation and ethical data practices in the digital economy.

© 2025 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/)

1. Introduction

The EU Data Act[1] is a piece of legislation that will have an impact on data sharing, emphasising users' rights and forms of control over data from IoT and edge devices and related services, while maintaining a balance with the interests of manufacturers or data holders. In addition, it encourages the innovation that comes from the various forms of data sharing[2]. The Data Act defines a connected product as an object that collects data about its use and environment and transmits it electronically[3]. The EU Data Act entered into force on January 11, 2024 and will apply from September 12, 2025. Obligations under Article 3(1) will apply to devices legally defined as connected products and related services put on the market after September 12, 2026. Other articles, such as those relating to contracts, apply only to contracts concluded after September 12, 2025. However, the Act's success hinges on effective implementation, standardisation efforts, and a continuous dialogue to balance user rights, data protection, and innovation incentives. There is no clear insight yet into what technological choices can be made to implement the Data Act in connected products. In this paper, we discuss the requirements that are made by the Data Act regarding access to data in connected products. More specifically, we propose a set of technological choices that can be made to implement the Data Act in this regard. Methodologically, we have rooted this paper in the epistemological tradition of design science research (DSR) [1].

DSR presents a research methodology commonly used in fields like information systems, engineering, and management to develop innovative solutions to practical problems. Its primary goal is to create and evaluate *artifacts*—such as models, methods, frameworks, or tools—that address specific challenges while also contributing to the advancement of knowledge in each domain. In our case, the artifacts that are under study are personal data stores and the advancement of knowledge should be situated in the legal and computer science fields.

2. European Data Strategy and the Data Act

Broadly speaking, the EU aims to create a single market for data, where data flows freely within the EU according to European rules and fundamental rights. In this context, the EU wants to evolve into a data economy with fair rules for data access and re-use. Data from key European sectors is conSOLIDated into shared, interoperable data spaces, and data holders and users are equipped with rights, tools, and capabilities to exercise forms of control or rights over data[4]. A fundamental shift in perspective is required, moving away from the traditional competition for data collection and towards a focus on the development of innovative data-driven services (e.g. personalised medicine through the European Health Data Space, improved mobility through

[1] Regulation (EU) 2023/2854.
[2] Recital 32 Data Act.
[3] Art. 2(5) Data Act.
[4] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 'A European Strategy for Data', COM(2020) 66.

the European Mobility Data Space)[5]. The EU Data Strategy is therefore a multifaceted approach, comprising a series of instruments that collectively contribute to the realisation of the aforementioned objectives. The Data Governance Act (DGA) has been introduced with the objective of addressing the issue of a lack of trust between service providers, which has previously resulted in reluctance to share data with competitors. The DGA introduces the concept of neutral intermediaries to facilitate data sharing between parties. Furthermore, the Data Act (DA) addresses the issue of the release of IoT data and its fair access and use. Similarly, the Digital Markets Act (DMA) facilitates the establishment of a fair and more competitive data economy. This is achieved by subjecting gatekeepers, as defined in Article 3 of the DMA, to device-neutrality regulations on their platforms and prohibiting them from bundling specific core platform services. Furthermore, gatekeepers are subject to interoperability obligations. The Digital Services Act (DSA) aims to gain control over services provided on top of vertically integrated platforms, for instance, by imposing requirements regarding explainability and transparency of their proprietary algorithms.

The Data Act, with its focus on data portability and interoperability, represents an important component of the European Data Strategy. Indeed, it addresses the issue of data that is currently with IoT providers and which requires further fair use and sharing. Users gain the right to access and port data from their connected products, facilitating switching between service providers or using third-party applications[6]. The objective of this enlarged data portability right, in combination with the promotion of interoperability standards, is to eliminate technical barriers to data sharing, thereby allowing for seamless data sharing between diverse systems and platforms[7]. This open ecosystem has the potential to unlock novel data-driven services and applications across sectors. In addition to addressing the use of connected products by natural persons, the Data Act also addresses potential market imbalances in business-to-business relationships. This is achieved by promoting fair and non-discriminatory access to data, thereby fostering a level playing field. In particular, the criteria of fairness, reasonableness, and non-discrimination (FRAND) should be regarded as procedural standards for negotiating terms that can subsequently be employed as a basis for dispute resolution by the relevant authorities[8,9]. Additionally, the Act curtails the direct access of gatekeepers[10] (large online platforms under the Digital Markets Act[11]) to user data, preventing them from leveraging their dominance to stifle competition [1]. By enabling users to exercise forms of control over and share their data, the EU Data Act sets the stage for a more dynamic data economy. The objective is to evolve beyond the platform-based data sharing model, which views data as the primary resource of a single firm. Increased data accessibility and interoperability can drive innovation, leading to the creation of new services, applications, and business models. This can range from data-driven insights for businesses to competitive aftermarket services and advancements in smart cities and the Internet of Things.

It is important to note that the Data Act and its new provisions complement existing legislation [2] such as the General Data Protection Regulation (GDPR)[12], the Free Flow of Non-Personal Data Regulation[13], and the Open Data Directive[14]. Because the Data Act is intended to apply in several different contexts, there are still issues that have yet to be resolved about the simultaneous application of these different regulations. It could be argued that the Data Act expands some

[5] Commission Staff Working Document on Common European Data Spaces, SWD(2022) 45 final.
[6] Art. 4, art. 5 Data Act.
[7] Art. 34, art. 35 Data Act.
[8] Harhoff, D. (n.d.). *Max Planck Institute for Innovation and Competition*. Para 101.
[9] Art. 16 Data Act.
[10] Art. 44 Data Act.
[11] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).
[12] Regulation - 2016/679.
[13] Regulation (EU) 2018/1807.
[14] Directive (EU) 2019/1024.

rights of data subjects to include data users and thus legal persons. This represents a departure from the traditional, "data subject as a natural person", terminology established by the GDPR, whereby the term "user" is defined as "a natural or legal person that owns, rents or leases a product or receives a service" [3]. Furthermore, there is a lack of clarity regarding the interrelationship between the Data Act and the implementation of the GDPR in terms of scope in the context of data sharing. This is due to the fact that the Data Act encompasses both personal and non-personal data, and the current interpretation of personal data is notably expansive[15]. This is evidenced by the fact that personal data is frequently inextricably linked[16] with non-personal data in mixed data sets. As an illustration, the interrelationship between the Data Act and the principles of purpose limitation and data minimisation set forth in the GDPR is a subject of ongoing debate [4]. The objective of this research is not to delve into the more fundamental questions of legal compatibility. Instead, this paper sets forth a multidisciplinary perspective on the development of "regulatory technology" [5], referring to technology that is designed to apply specifically to certain types of regulation. Within this context, the paper sets out a series of technological choices that can be made to implement the Data Act in as many contexts as possible [3]. To pursue this objective, it was necessary to engage in a broad collaborative effort between the legal, computer science and business disciplines.

3. Key Concepts

To facilitate the understanding of the Data Act, it is essential to establish and explain the key concepts that are used in the document. In this analysis, we place comparatively little emphasis on the concepts that are used in other legislative texts, such as the GDPR. This discussion focuses on the Data Act and directs the reader to the other texts for further clarification on concepts such as data controller, data intermediary or lawful grounds of processing. It is nevertheless worth noting that while a semantic and conceptual continuum exists across the various texts drafted by the EU to regulate the European Single Market, the relationship between the definitions set out in the various legislative instruments of the European Data Strategy is a point of considerable debate.

3.1. Key concepts from the data act

3.1.1. Data

The definition of data itself under the Data Act is broad and encompasses any digital representation of acts, facts, or information. Data can be personal or non-personal. The data covered by data sharing obligations comprises raw and pre-processed data. Purely descriptive data (e.g. user manuals or labelling information) is excluded. Furthermore, highly enriched data which has been derived or inferred using algorithms or other additional investments (e.g., proprietary algorithms) is s3imilarly outside the scope of these obligations[17]. In particular, the data in question comprises both that which is intentionally recorded by the user and that which is generated as a by-product of the user's action. Such data may, for example, include diagnostic information. It can be reasonably argued that data generated because of the interaction between the user and the product through a virtual assistant also falls within the scope of the aforementioned data [6]. In practice, however, difficult questions of delineation are likely to continue to arise. For example, it could be argued that the Data Act suggests that combining the output of multiple

[15] E.g., on the advice of the Personal Data Committee Amsterdam and the Dutch Data Protection Authority, all shared mobility data must be regarded as personal data.

[16] Communication from the Commission to the European Parliament and the Council, Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union, COM/2019/250 final.

[17] European Commission, Frequently Asked Questions Data Act, 6-7.

sensors is sufficient to qualify as derived or inferred data and thus fall outside the scope. Nevertheless, regarding future developments, namely the growing ease with which sensor fusions can be carried out, it would seem prudent to place greater emphasis on the amount of additional investment required to qualify data as highly enriched and therefore outside the scope of the Data Act. [7] Otherwise, the scope of the Data Act could be readily circumvented, resulting in the failure to achieve the objective of opening up data for further data sharing and fostering innovation. Lastly, the provisions of the Data Act do not extend to intellectual property rights, including those pertaining to audio or audiovisual content.

3.1.2. Connected product

Connected products are designed to generate, obtain, or collect data pertaining to their performance or the surrounding environment. The term is applicable to a diverse array of products, encompassing e.g. consumer items, industrial machinery, or medical equipment. Typical examples include smart home appliances, consumer electronics, industrial machinery, medical devices, smartphones, and TVs. The Data Act excludes prototypes marketed in the EU. It should be noted that any product whose primary function is to store, process, or transmit data on behalf of a third party is to be considered an exclusion. However, this area is still characterised by a certain degree of ambiguity. The initial intention was to exclude devices such as servers and routers. The initial Data Act proposal sought to exclude smartphones and other products with a primary function of displaying, recording, transmitting or playing content. However, the final text of the Data Act only referred to the primary function of storing, processing and transmitting data as an exclusion. The recent FAQ of the European Commission underscores the limited scope of this exclusion, specifying that if the said products are owned, rented or leased by the data user, they remain subject to the provisions of the Data Act [6].

3.1.3. Related service

A related service will encompass most digital services. In order to be classified as a related service, two principal conditions must be met. Firstly, a two-way data exchange between the related service and the connected product is required. This suggests that applications that merely display data received from a connected product are not eligible for classification as a related service. Secondly, for a related service to be considered, it must affect the connected product's functions, behaviour or operation. This may be exemplified by an application that enables the user to adjust the settings of a connected thermostat. The precise meaning of this function-affecting concept remains an ongoing task for courts and real-life practice alike. In any case, several factors may be taken into consideration to assess the nature of the service in question. These may include the user's expectations, the accompanying marketing, the replaceability of the service, the outcome of contractual negotiations, and whether the service was preinstalled on the connected product. It should be noted that connectivity, power supply and aftermarket services (e.g. auxiliary consulting, analytics and financial services, and regular repair and maintenance) cannot be considered as related services.[18]

3.1.4. Data processing service

The term "data processing services" is defined very broadly under the Data Act[19] and is expected to encompass cloud computing and edge services[20]. It is notable that the definition is consistent with existing definitions of cloud computing, such as the one proposed by the US National Institute of Standards and Technology (NIST) [8]. Firstly, a data processing service is a digital service initially provided through a heterogeneous range of platforms, including web browsers, mobile devices, and workstations. Secondly, it provides access to flexibly allocated computing resources, which are capable of handling fluctuations in demand irrespective of the geographical location of the resources. Moreover, depending on the workload, the resources

[18] Recital 17 Data Act.

[19] European Commission, Frequently Asked Questions Data Act, 7-8.

[20] DA Proposal, Recitals 69 and 71; see also DA Memorandum, p. 3.

available can be rapidly increased or decreased. Thirdly, multiple users have shared access to the service, but the processing is carried out separately for each user. Fourthly, the resources are distributed across multiple networked computers or devices, or alternatively, processing occurs near the data source. Lastly, customers have the ability to independently provision computing capabilities and automatic services, including a dashboard and interface.[21]

3.1.5. Data user

A natural person, legal entity, or a public body that has a legally recognized right to a connected product, whether based on ownership, rental, or lease agreements. Data users remain subject to existing legal constraints regarding the precise further use of the data in question (e.g. relevant data protection legislation and rules regarding trade secrets).

3.1.6. Data holder

The data holder is responsible for managing access to the data they provide. Such entities may include the original equipment manufacturers, software companies, sensor vendors, or other electronic equipment suppliers, if they possess the requisite access to data. In contrast to larger companies, micro and small businesses are exempt from data holder obligations, if they are the sole data holders of connected products and related services and these are therefore not subcontracted to another company.

The obligation of the data holder may be transferred to another legal entity. In certain instances, there may be no data holders, e.g. when data generated by a connected product is stored directly on the device without access from the manufacturer. In certain cases, two data holders can merge. This can be the case for example when a person purchases a connected product and subsequently downloads an application that facilitates the use of the connected product. The merging data holders may then include the entity that initially marketed the product and the entity that offers the related service, which in this case would be the downloaded application.

3.2. Key concepts not defined in the data act

Here, we list certain key concepts that are essential to the reading of this paper yet are not defined in the Data Act itself.

3.2.1. Multiparty data access

This term is not defined by the Data Act itself, but we define it here as the necessity for the data holder to provide access by other parties to the data produced by the connected product if the data user wishes it so. Looking at technological solutions for multiparty data access is the key focus of this paper.

3.2.2. Data subject

In accordance with the GDPR[22], this is an identifiable natural person, defined as an individual who can be directly or indirectly identified, particularly by using an identifier such as a name, identification number, location data, online identifier, or one or more factors specific to the individual's physical, physiological, genetic, mental, economic, cultural, or social identity.

3.2.3. Data controller

In accordance with the GDPR[23], the term "data controller" refers to the natural or legal person, acting alone or jointly with others, who determines the purposes and means of processing personal data.

[21] Recital 80 Data Act.
[22] Article 4 (1) GDPR.
[23] Article 4 (1) GDPR.

3.2.4. End-user

With an end-user, we refer to the person that is performing the management of the data access and usage modalities.

3.3. Data Act requirement categories on multiparty access to connected product data

The EU Data Act lays out several requirements that manufacturers, data holders, and related parties must adhere to for data access in connected products to be compliant. These requirements focus on empowering users[24], ensuring fair access, and promoting interoperability. It is imperative that both businesses and individuals, whether owners, lessees, or renters of such products, be afforded additional forms of control and rights over the data they generate. In this section, we delineate the categories of requirements that define the necessity for enabling multiparty data access to connected products and related services in the Data Act. As the success of the Data Act is contingent upon effective (technical) implementation, the objective of this paper is to present a series of technological options that can be employed to implement the Data Act in a multitude of contexts. These options will be mapped to the categories of requirements listed below. After analysis of the Data Act, we discern the following requirement categories:

1. Data accessibility for users
2. Information transparency for users
3. User rights regarding data
4. Fair and transparent data sharing
5. Interoperability
6. Protection of legitimate interests

Each requirement category has a set of subcategories, which we will explain in what follows.

3.3.1. Data accessibility for users[25,26]

Direct and easy access: Connected products should be designed to provide data users with direct, easy, and secure access to the data they generate, preferably in real-time where technically feasible. This access should be free of charge and not unduly difficult for data holders (e.g., affecting the autonomy of the data user in a particular way). The obligation to provide direct and easy access represents a fundamental aspect of the regulatory framework, which establishes the fulfilment of obligations through by-design mechanisms. This concept is analogous to the established principle of data protection by design in European data protection legislation.

Commonly used and machine-readable format: Data should be provided in a format readily understandable by users and machines, such as CSV, JSON, or RDF, ensuring the data can be easily interpreted and used with other applications.

Metadata for interpretation: Along with the data, relevant metadata should also be provided to facilitate understanding and use of the data. This includes information about the data's context, collection method, and any limitations on its use. Although the GDPR initially sought to provide for metadata the right to data portability[27], it ultimately did not achieve this aim. The Data Act now offers a genuine opportunity to leverage the potential of metadata.

---

[24] For a critical stance on the empowering obligations of the Data Act, see Kerber, W. (2023). Governance of IoT Data: Why the EU Data Act Will not Fulfill Its Objectives. *GRUR International, 72*(2), 120–135. https://doi.org/10.1093/grurint/ikac107.

[25] Regulation (EU) 2023/2854, art. 3, art. 4.

[26] Following article 7 of the Data Act, micro and small businesses, particularly manufacturers or service providers of connected products, are exempt from certain data holder obligations. For example, B2C and B2B data-sharing obligations do not apply to SMEs, provided they meet certain conditions (Article 2 and 3 of the Annex to Recommendation 2003/361/EC).

[27] The Article 29 WP clarified in this regard that the additional use of sufficient metadata to preserve the meaning of the exchanged information was necessary upon data sharing.

Simple request mechanism: Access should be granted through straightforward requests, such as through a user account, a mobile application, or an application programming interface (API) associated with the product.

Verification of legitimate users: The Data Act prioritises user empowerment, data accessibility, and streamlined processes for sharing data. The proposed approach advocates a streamlined process for user verification, advocating the deployment of existing mechanisms, such as user accounts and mobile applications, to facilitate data requests.

3.3.2. Information transparency for users[28]

Pre-contractual information: Before a user buys, rents, or leases a connected product, the vendor must provide clear and comprehensive information about the data it generates. This includes the data type, format, estimated volume, storage location (on-device or remote server), data retention duration, and methods for accessing, retrieving, and erasing the data. Similarly, prior to a user registering for a related service, the provider is obliged to provide detailed information regarding the type, volume, and collection frequency of the product data they will access, data storage arrangements, data retention policies, and procedures for data access and retrieval.

Changes to data access: If updates or alterations to a connected product or related service during its lifetime impact the data generated or the means of access, the user must be informed.

User rights regarding data[29]

Right to use and share: Data users have the right to use the data from their connected product for any legal purpose, including sharing it with third-party service providers, even those in competition with the original data holder.

Right to rectification and erasure: Users have the right to correct inaccurate data and request the deletion of their data under the conditions set out in the GDPR.

No unfair contractual terms: Similarly to unfair commercial practices in consumer law, several scenarios are outlined where contractual terms are presumed to be unfair. A hierarchy of provisions is established, comprising a blacklist of practices that effectively exclude liability for grossly negligent breaches of duty by a party; a grey list of provisions that have the effect of unreasonably limiting the remedies of a party or the liability of the imposing party; and a general clause that may be invoked in instances where a practice constitutes a gross deviation from good business practice. Consequently, such terms are deemed unenforceable. These provisions also aim to support small and medium-sized enterprises (SMEs) in negotiating access to data more effectively. The European Commission will adopt a standard contract that will further define the concept of unfairness and also refer to good commercial practices regarding data access.[30] Until such time as the Commission formally adopts the relevant terms, it would be prudent for parties to consider existing soft law instruments [9], such as the ALI-ELI Principles[31] or the rules for data provision currently being developed by UNCITRAL[32].

3.3.3. Fair and transparent data sharing[33]

Transparency in data sharing arrangements: Any agreements for data sharing, especially in business-to-business contexts, should clearly outline the terms and conditions, including aspects of compensation, liability, and data security.

Prohibition on unfair commercial use of data: Data holders cannot leverage user data for unfair commercial advantage against the data user or authorised third parties. This includes deriving insights from the data to undermine the data user's competitive position in the market [10].

---

[28] Regulation (EU) 2023/2854, art. 3.
[29] Regulation (EU) 2023/2854, art. 4, art. 5, art. 7.
[30] Article 41 DA.
[31] https://principlesforadataeconomy.org/the-project/the-current-draft/data-contracts/.
[32] https://uncitral.un.org/en/working_groups/4/electronic_commerce.
[33] Regulation (EU) 2023/2854, art. 5, art. 8, art. 9.

Reasonable compensation: While data users should be able to access their data free of charge, data holders can request reasonable compensation from third parties for the costs associated with providing access to data, such as data formatting, transfer, and storage. This is justified by the necessity to encourage sustained investment in the generation and availability of data. The level of investment required is dependent on the quality and quantity of the data in question, as well as on a reasonable margin [11]. There is no specific limit. Rather, transparency requirements are deemed to ensure that compensation is calculated based on objective, verifiable criteria [2]. It is important to note that while there is no upper limit to reasonable compensation, when small and medium-sized enterprises (SMEs) or non-profit research organisations seek access as data users, the compensation cannot include a profit margin and shall only include costs incurred in making the data available [6]. Nevertheless, there is still a lack of clarity about the precise meaning of reasonable compensation.[34]

3.3.4. Interoperability[35]

Open standards and common specifications: Data, data-sharing mechanisms, and services related to connected products should adhere to open interoperability standards and common specifications. This ensures that data can be readily transferred and used across different systems and platforms.

Interoperability of data processing services: The Data Act also promotes interoperability for data processing services, making it easier for users to switch providers or use multiple services concurrently without experiencing technical hurdles [2]. This obligation is designed to facilitate switching by removing obstacles while providing functional equivalence. In accordance with Article 29(2), providers of data processing services are obliged to reduce any switching charges from January 11, 2024 onwards. From January 12, 2027 onwards, providers will no longer be permitted to levy charges for switching. In the event of parallel use of data processing services, the provider may still impose costs incurred for data egress, e.g., in a multi-cloud deployment model.

Support for smart contracts: These incorporate a range of technological implementations and are, as such, technology neutral in principle, despite the fact that smart contracts are frequently rooted in distributed ledger technology. Article 2(39) defines a smart contract as "a computer program used for the automated execution of an agreement or part thereof, using a sequence of electronic data records and ensuring their integrity and the accuracy of their chronological ordering." The Data Act stipulates that data holders must give full consideration to the use of smart contracts in the context of long-term or regular arrangements between data holders and data users. It is evident that the term 'regular' pertains to instances of recurring or recurrent business dealings.[36] The EU's stated objective was to reduce administrative burdens in this regard, with a view to facilitating the sharing of data within data spaces. Furthermore, the promotion of interoperability among tools such as smart contracts is also advocated in relation to other transactions. For a smart contract to be deemed valid, it is necessary for it to satisfy the essential requirements set forth in the Data Act, including robustness and access control, continuity, consistency, and a mechanism to terminate the continued execution of transactions. This approach has previously been applied in the context of legislation governing products, medicines, medical devices, and so forth. Subsequently, the European Commission may develop standards, with compliance implying a presumption of meeting the essential criteria.

3.3.5. Protection of legitimate interests[37]

Protected data of the data user: It should not be possible for data holders to gain insights into the economic situation, assets, and production methods of a data user in any way that could

[34] In particular, it is unclear whether non-SMEs are always expected to pay the market value of the underlying data, or only compensation for the costs of compliance with the access obligation [10].

[35] Regulation (EU) 2023/2854, art. 33, art. 34, art. 35.

[36] See recital 47 Data Act.

[37] Regulation (EU) 2023/2854, Article 4(13) (14), art. 5 (9), 7, art. 8, art. 43.

negatively impact the data user's commercial position in the market. Furthermore, sharing such data with third parties is only permissible if it is explicitly permitted by the data user through a contractual agreement.

Development of a competing product: It is not permissible for legal entities that are data users and have gained access to data from connected products under the Data Act to make use of that data for the purpose of developing a competing product.

Security Requirements: Data access mechanisms must prioritise the security of connected products and prevent unauthorised access that could jeopardise the health, safety, or security of individuals. In this regard, data holders are afforded the option of restricting or prohibiting access. Sectoral authorities may provide technical expertise to determine whether restrictions are necessary. Requirements regarding safety or security must always have a foundation in EU or national law.

Protection of Trade Secrets: While promoting data accessibility, the Data Act recognizes the need to protect legitimate business interests, such as trade secrets. Data holders are not required to disclose information that would compromise their trade secrets. In this context, it is important to note that the scope of data sharing obligations does not extend to highly enriched or otherwise inferred or derived data. Data holders are required to ascertain which data is to be considered a trade secret. In addition, they must agree with the third party on the technical and organisational measures that are to be taken for the protection of the shared data. These measures can be divided into several categories: model contractual terms, confidentiality agreements, strict access protocols, technical standards and the application of codes of conduct. It is noteworthy that the data holder is afforded the option of identifying any data as a trade secret. It is therefore not difficult to envisage grey areas where the provisions set forth in the Data Act may result in the unjustified restriction of the free flow of data.[38]

3.4. Personal data store technologies to meet data act requirements

Personal data stores have been in development for many years. For example, the MyData movement has long argued for human-centric data infrastructures that echo the empowerment goals that are inherent to personal data stores. [12]. Indeed, personal data stores aim to allow data users to store their personal data and control the way in which data in accessed and used by third parties. The MyData approach, however, seeks to achieve more than simply controlling who can access personal data. Indeed, its fundamental objective is to empower individuals to leverage their data to enhance their autonomy and quality of life.[39] Such personal data stores also have a significant role to play within the data space paradigm, which aims to allow data sharing between organisations in the public and the commercial realm. Indeed, many use cases involve data that pertain to individuals, making a focus on personal data a necessary complement to the hitherto dominant focus on non-personal data. The development of the data space concept has mainly taken place in communities organised around initiatives such as the International Data Spaces Association (IDSA)[40], GAIA-X[41], the Big Data Value Association (BDVA)[42] and the Data Spaces Support Center (DSSC)[43]. The main interpretations of data spaces have thus far focussed on the sharing of organisational data, yet clearly, many use cases that require data

[38] For a more in-depth study of this topic, see [3]. Under the Data Act, data holders have the option to identify specific data as trade secrets, potentially creating grey areas where data-sharing requirements could be circumvented. This could lead to restrictions on data flow, similar to the GDPR's grey areas regarding data portability and technical feasibility. While dispute resolution mechanisms may address these issues, the broad scope of trade secrets under the Data Act could still present challenges.

[39] https://mydata.org/wp-content/uploads/2025/05/MyData-in-Motion-Evolving-Empowerment-for-2025-and-beyond-layout-v4-1.pdf.

[40] https://internationaldataspaces.org/.

[41] https://GAIA-X.eu/.

[42] https://bdva.eu/.

[43] https://dssc.eu/.

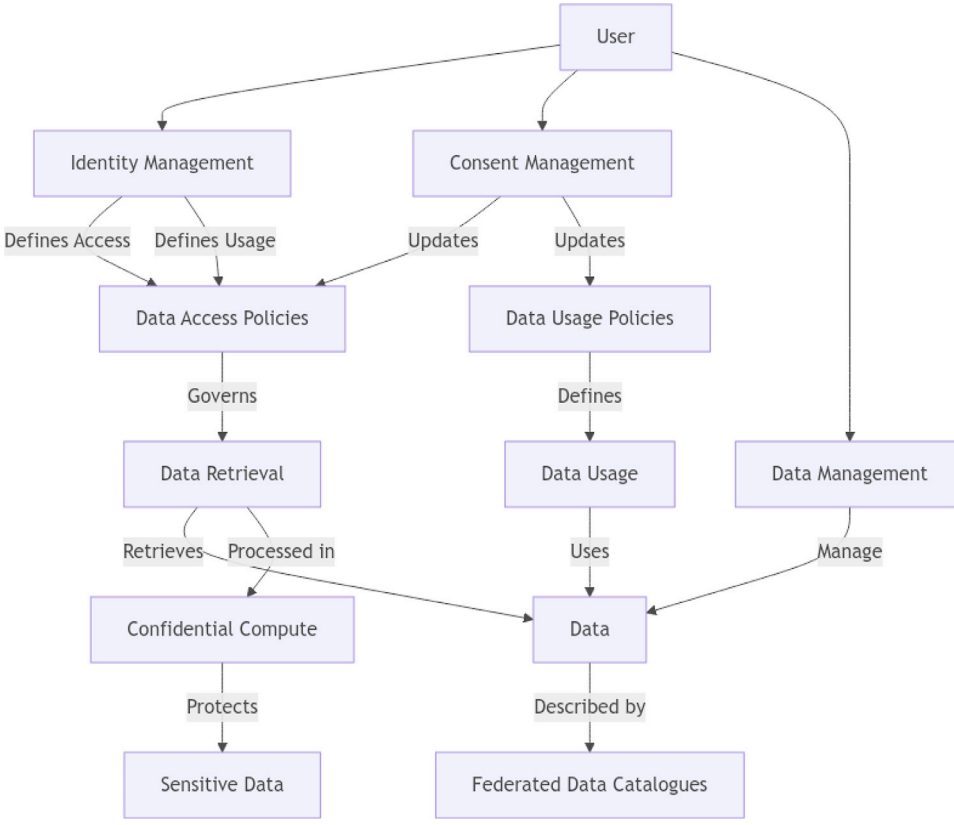

Fig. 1. Conceptual model of the key personal data store concepts.

must allow personal data to be shared as well. As a result, it seems that contributing to the discussion on how personal data stores can fit in the data spaces framework should be beneficial to the EU's strategic aim of creating a level playing field data economy in which innovation can thrive.

Data stored in a personal data store can be shared across different platforms and services, but only when based on a lawful data processing ground and when accompanied by sufficient safeguards following GDPR obligations such as transparency measures. More specifically, data protection and security are further ensured through encryption and privacy-enhancing technologies, protecting sensitive information and ensuring that only authorized parties can access it. Transparency is an additional key feature, as users of connected products and related services are informed about how their data is used by data holders and third parties while tracking access, fostering trust and accountability. Furthermore, international, open standards are critical for practical compliance. Prominent candidates include SOLID for personal data pod APIs and W3C Web of Things for device-side data models and IoT-based "Thing" Descriptions and NGSI-LD for interoperable context and entity management across distributed data ecosystems.

Overall, a personal data store empowers users to actively manage their digital footprint, offering a more private and secure alternative to traditional data storage models where third parties often have significant control over personal data. Fig. 1 describes key concepts of personal data stores and their management, which will be instrumental in addressing the Data Act requirements in what follows.

The conceptual diagram begins with the user engaging with three critical systems: identity management, consent management, and data management. Through the identity management system, users establish and verify their identities, ensuring secure and authenticated access to services. Users interact with the consent management system to configure the policy management system and provide explicit permissions and preferences regarding how their data can be accessed and used, including to provide explicit consent. The data management system manages the data, including the creation of new data assets or the update of data. This triadic interaction empowers users with control over their personal data and how it is managed within the system.

Building upon the user's inputs, the policy management system actively updates both data access policies and data usage policies. These policies are dynamic reflections of the specific conditions under which data can be accessed and used. In this context, the term "access" primarily denotes eligibility under the conditions of the Data Act to access specific data. In contrast, usage policies pertain to subsequent steps and, consequently, further use of the data. In accordance with the GDPR, users are required to specify a legal basis for the processing of personal data, which may include consent, legitimate interest, or performance of a contract. Additionally, they must determine a specific purpose of this processing.[44] Meanwhile, the identity management system plays a crucial role in controlling access to these data access policies, ensuring that only authorized entities can view or modify them. This layered approach to policy management ensures that user preferences are respected, and that data governance remains robust and compliant with relevant regulations.

For data to be retrievable, it must be described and indexed by the federated data catalogues. These catalogues serve as a comprehensive directory, offering metadata and standardized descriptions of data assets across multiple federated systems. By doing so, they facilitate efficient data discovery and access, allowing data users and systems to locate and understand the available data within a broader, interconnected environment. This federated approach enhances collaboration and interoperability between different data sources and platforms. Through the data management system, the user is also able to ensure that the federated data catalogues are up-to-date and request the update of any incorrect data.

To safeguard sensitive information, the data retrieval and data usage process can operate within a confidential computing environment. This secure computing paradigm ensures that data is processed in a protected enclave, preventing unauthorized access during computation. The confidential compute environment effectively protects sensitive personal data, maintaining confidentiality and integrity even as data is actively used. This security layer is critical for handling sensitive or regulated data, providing assurance that all data processing adheres to high standards of security and privacy.

A component that is not explicitly represented in Fig. 1, but which we believe is essential to the personal data store concept, is linked open data. Linked Data refers to a method of publishing structured data in a way that allows it to be interlinked and accessible over the Web. It is based on the principles of Linked Data, which are a set of best practices for publishing and connecting data using standardized Web technologies like Uniform Resource Identifiers (URIs) and the Resource Description Framework (RDF).

3.5. Requirements-technology matrix

To understand how to address the Data Act using the conceptual model addressed above, we set out to map the technological concepts that were described in the previous section to the multiparty data access requirements that were presented as part of the Data Act. Table 1 shows how the various Data Act requirements (columns) can be addressed by the indicated technolog-

[44] See recital 7 and 34 Data Act.

Table 1
Personal data space technologies and requirements matrix, linking the requirements of the Data Act to the key technological personal data space concepts.

Data act requirement categories	Requirements	Identity Management	Consent Management	Data Access Policies	Data Usage Policies	Federated Data Catalogues	Confidential Compute	Linked Data technology
Data Accessibility for Users	Direct and Easy Access	DA1		DA2		DA3		DA4
	Commonly Used and Machine-Readable Format	MR1	MR2	MR3	MR4	MR5		MR6
	Metadata for Interpretation					MD1		MD2
	Simple Request Mechanism	RM1	RM2	RM3	RM4			RM5
	Verification of legitimate users	LU1						
Information Transparency for Users	Pre-Contractual Information			PCI1	PCI2			
	Changes to Data Access			CDA1				
User Rights Regarding Data	Right to Use and Share		RUS1	RUS2	RUS3			
	Right to Rectification and Erasure		RRE1		RRE2			
	No Unfair Contractual Terms				NUC1			
Fair and Transparent Data Sharing	Transparency in Data Sharing Arrangements		TDA1	TDA2	TDA3			TDA4
	Prohibition on Unfair Commercial Use of Data	PUC1			PUC2			
	Reasonable Compensation							
	Procedural not substantive requirement							
Interoperability	Open Standards and Common Specifications	OS1	OS2	OS3	OS4	OS5		OS6
	Interoperability of Data Processing Services							IDS1
	Support for Smart Contracts							
Protection of Legitimate Interests	Potected data of the data user	PD1	PD2	PD3	PD4			
	Development of a competing product							
	Security Requirements	SR1		SR2	SR3		SR4	
	Protection of Trade Secrets						TS1	

ical concepts (rows). We have given each cell in the table a unique code, to easily reference it in the discussion below.

3.5.1. Direct and easy access

Identity Management (DA1) authenticates authorized users efficiently, streamlining access for users. Data Access Policies (DA2) define protocols for data retrieval, allowing the control of the parties that have access to the data. Data Catalogues (DA3) facilitate the discovery and access of data across multiple platforms, enhancing accessibility. Linked Data (DA4) provides data in a structured, linked format, further simplifying access by connecting related data points.

3.5.2. Commonly used and machine-readable format

Linked Data (MR6) is the main way to contribute to the machine readability of the connected device data. Leveraging over 20 years of semantic web and linked data research and development is an evident way to address this requirement. All technologies that we propose to address this requirement, except confidential compute, can be expressed using linked data (cf. MR1-MR5).

3.5.3. Metadata for interpretation

Federated Data Catalogues (MD1) provide comprehensive metadata that helps users understand and effectively utilize the data. Linked Data (MD2) enhances this by embedding metadata within the data itself, facilitating better data comprehension and interoperability between systems.

3.5.4. Simple request mechanism

Identity management (RM1) is a component for managing processes that request data, as it is necessary to understand who or what is requesting the data to be able to establish if this request is permissible. Consent Management (RM2) allows end-users to set the conditions under which data can be shared. Data Access Policies (RM3) regulate the data access process, based on the consent that was given by end-users. Similarly, Data Usage Policies (RM4) describe the way in which data can be used and by whom. Linked Data (RM5) simplifies data queries through standard linked data protocols, reducing complexity in data retrieval.

3.5.5. Verification of legitimate users

This requirement is mainly supported by the proper user of identity management mechanisms (LU1). As explained in the previous section, they allow data sharing when combined with Consent management, Data Access Policies and Data Usage Policies.

3.5.6. Pre-contractual information

By communicating the default Data Access Policies (PCI1) and Data Usage Policies (PCI2), the end-user who is contemplating the acquisition of a connected product can better understand the data sharing processes that may occur. These default policies can later be adjusted by the end-user through consent management, but provide a first indication on the nature of the connected product under scrutiny.

3.5.7. Changes to data access

Such changes can be implemented by changing the Data Access Policies (CDA1). Should a vendor feel that there is a need by end-users to change these policies in the connected products that were sold, these end-users could be notified to change the policies. Alternatively, and if the end-user allows it, the vendor could build automated services that change these policies in name of the end-user. However, the end-user should still be notified of the changes that were made.

3.5.8. Right to use and share

Consent Management (RUS1) controls permissions for data sharing based on user consent, respecting user preferences and legal requirements. Data Usage Policies (RUS2) define how data can be used and shared legally, providing clear guidelines, while Data Access Policies (RUS2) do the same from the perspective of data access.

3.5.9. Right to rectification and erasure

Consent Management (RRE1) handles user requests to correct or delete data efficiently, ensuring compliance with data protection regulations. Data Usage Policies (RRE2) establish protocols for data modification and deletion, providing a clear framework for how such requests are processed.

3.5.10. No unfair contractual terms

Data Usage Policies (NUC1) can be audited to ensure that there are no unfair contractual terms by defining fair and compliant usage terms. They can help in creating contracts that are transparent and adhere to legal standards, protecting both users and providers from unfair practices.

3.5.11. Transparency in data sharing arrangements

Consent Management (TDA1) keeps users informed about how their data is shared, fostering trust. Data Access Policies (TDA2) and Data Usage Policies (TDA3) detail the terms of data sharing, providing clarity on how data is used and with whom it is shared. Linked Data enhances transparency by using linked data structures (TDA4) that make data sharing practices more visible and understandable.

3.5.12. Prohibition on unfair commercial use of data

Consent management (PUC1) and Data Usage Policies (PUC2) prevent the unfair commercial use of data by defining acceptable usage terms and setting boundaries for how data can be used in commercial contexts. This protects users from exploitation and ensures that data is used ethically and legally.

3.5.13. Reasonable compensation

The technologies we propose do not support the requirement that data holders can request reasonable compensation from third parties for the costs associated with providing access to data, such as data formatting, transfer, and storage. Although indications on pricing can be presented through Federated Data Catalogues, we argue that pricing and compensation as described in the Data Act are rather the remit of the business than the technical domain.

3.5.14. Open standards and common specifications

Identity management (OS1), Consent management (OS2), Data Access Policies (OS3), Data Usage Policies (OS4) and Federated Data Catalogues (OS5) can all be addressed using open standards with common specification, as we will detail in the section on architecture. Linked Data (OS6) itself is completely based on open standards and common specifications.

3.5.15. Interoperability of data processing services

Linked Data (IDS1) enhances interoperability through linked data formats that allow seamless integration and processing of data across different services and platforms. The use of shared ontologies in linked datasets can further complement the interoperability requirement.

3.5.16. Support for smart contracts

It is important to note that the reliance on smart contracts as outlined in the Data Act necessitates a nuanced and constrained approach. In accordance with the provisions stipulated within the Data Act, smart contracts are regarded as being fixed in terms of both content and operation. Nevertheless, the safe termination and interruption requirements implied by the Data Act

appear to be challenging to achieve in practice. Furthermore, the necessity for oracles or real-world input data may result in certain challenges regarding the reliability of the sources. The immutability and enforceability characteristics of distributed ledger-based smart contracts are not always compatible with domestic legal systems and civil law in that region [13]. In this regard, the present paper does not supported the use of smart contracts as rooted in distributed ledger technologies. Rather, we argue more broadly that under the technological neutral definition in the Data Act, Consent Management, Data Access Policies, Data Usage Policies and Federated Data Catalogs can play a role in facilitating the usage and publication of smart contracts under the Data Act.

3.5.17. Protected data of the data user

The combination of Identity management (PD1), Consent management (PD2), Data Access Policies (PD3) and Data Usage Policies (PD4) makes it possible for the data user to retain control over the types of data access and usage that can be made by the data holder and by third parties.

3.5.18. Development of a competing product

The technologies and architecture that we propose have no affordances that can solve this requirement.

3.5.19. Security requirements

Identity Management (SR1) strengthens security through robust identity verification, ensuring that only authorized users can access sensitive data. Data Access Policies (SR2) and Data Usage Policies (SR3) contribute to setting the conditions under which data can be shared and used, which can contribute to security. Confidential Compute (SR4) protects data in use by processing it within secure computing environments, safeguarding against unauthorized access during data processing activities.

3.5.20. Protection of trade secrets

Confidential Compute (TS2) contributes significantly to the protection of trade secrets by safeguarding sensitive business information during processing. By ensuring that data remains confidential even while being processed, it helps organizations protect their proprietary information from leaks and unauthorized disclosures.

3.6. Architecture

To implement the requirements-technology matrix discussed above, a high-level architecture can be specified, considering the secure, decentralised, and user-controlled features that should be incorporated. Such an architecture is proposed in Fig. 2, where any user, being a person, agent or service, can interact with multiple interfaces, which in turn can retrieve data from multiple storage units. Moreover, beyond decoupling applications from data storage, a further degree of decentralization can be achieved by having a dedicated authorization server taking care of the access and usage control mechanisms.

Different technological choices can be made to fulfil this architecture. Proprietary standards and systems are certainly possible, but we strongly advocate in favour of using open linked data web standards and vocabularies such as the Open Digital Rights Language (ODRL), the Data Privacy Vocabulary (DPV) and the Data Catalog Vocabulary (DCAT) for metadata modelling, and the SOLID[45] and User-Managed Access (UMA)[46] protocols for the storage and authorization components. Different authentication methods can be utilized to identify users, such as the SOLID-

[45] https://SOLIDproject.org/TR/protocol.
[46] https://docs.kantarainitiative.org/uma/wg/rec-oauth-uma-grant-2.0.html.

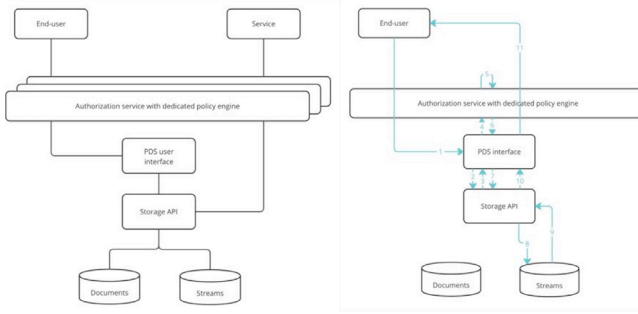

Fig. 2. Left: High-level architecture to support decentralized personal data spaces in support of Data Act requirements towards multiparty data access in connected products. Right: a version of the architecture showing a typical interaction flow between the end-user and the personal data store.

supported OpenID Connect (OIDC)[47] or the Verifiable Credentials (VCs)[48] standard, which allows the modelling of more complex types of identity claims including signatures, provenance and timestamp information.

Fig. 2 showcases the high-level architecture that we propose can address the requirements that emanate from the Data Act. In the right-hand side of Fig. 2, we show a typical interaction between an end-user and a (in this case streaming) data source. Practically, (1) end-users authenticate themselves through an interface, using any authentication method recognized by the authorization server, to request access to protected data. If (2) there is not yet an access token issued by the authorization server to get access to the requested resource, the request made by the interface to the data source (3) will result in a failed attempt to access the resource. Nonetheless, the interface is then (4) redirected to the authorization server, which in return (5) must check whether the existing policies governing the access and usage of the requested resource allow the interface to access it. The Policy Engine component is a protocol independent framework that takes as input the policies that govern the access and usage of a data source and assesses the policies that are active and must be fulfilled. In the presence of a successful request, i.e., access allowed to the resource, the authorization server can then (6) issue a token to be used by the interface (7) to get access to the resource, and finally (8) have access to it and (9), (10), (11) make the data available to the end-user. The same flow can also apply to services.

3.6.1. Identity and authorization management

This architecture allows users, being an end-user (i.e., a person), an agent or a service, to interact with multiple interfaces, for instance with multiple applications or APIs, which in turn can retrieve data from multiple storage units, while the access and usage is being governed by a dedicated authorization server [14]. This decoupling of storage and authorization concerns provides a more flexible and dynamic access and usage control mechanism when compared with traditional access control mechanisms. Moreover, this separation of concerns also allows the possibility of using different authentication methods, e.g., OIDC or VCs. For authorization and policy management, we propose to use the User-Managed Access (UMA) Protocol[49], which is a digital security framework that empowers individuals to control who can access their data and resources, under what conditions, and for how long. Built on the OAuth 2.0 standard, UMA provides a centralized mechanism for managing access permissions across multiple services and platforms. Rather than dealing with separate permissions for each service, users configure all policies in one place, known as an authorization server. This approach enhances privacy, stream-

[47] https://openid.net/specs/openid-connect-core-1_0.html.
[48] https://www.w3.org/TR/vc-data-model-2.0/.
[49] https://docs.kantarainitiative.org/uma/rec-uma-core.html.

lines consent management, and ensures that access decisions align with user-defined preferences.

The protocol operates through a clear division of roles. The resource owner, typically the data user, manages access policies via the authorization server. This server acts as a policy hub, coordinating with resource servers, where the data is stored, and issuing secure access tokens to requesting parties, such as applications or individuals seeking access. UMA also supports real-time, dynamic conditions for access, enabling rules based on factors like location, identity, or time of day.

3.6.2. Linked data standards in support of the architecture

For the proposed architecture, we consider the work of the W3C Open Digital Rights Language Community Group (ODRL CG)[50] and the Data Privacy Vocabularies and Controls Community Group (DPVCG)[51] to be essential, as they are state of the art resources for the expression of policies and representation of information related to data protection requirements, respectively [15]. We now provide a discussion of both resources.

3.6.2.1. Open digital rights language. ODRL is a W3C Recommendation for the expression of data access and usage policies related to the usage of content and services. The ODRL standard contains an information model[52] to represent permitted, prohibited and mandatory actions over assets, to be met by involved parties. Additionally, policies can be further restricted by constraints (e.g., temporal or spatial constraints) and duties (e.g. payments) may be imposed on permitted actions. Moreover, ODRL also has a standardised vocabulary[53] to be used to encode policies in RDF and a mechanism to extend its vocabulary in an interoperable manner using ODRL profiles[54]. As such, the ODRL CG oversees the maintenance and further development of the current standard and of future developments that might lead to a new version of ODRL. Amongst these developments, it is of interest to mention the specification of a formal semantics[55] for ODRL, as it will have a high impact on the interoperability of policy enforcement based on ODRL policies. This specification is currently being finalised and will document the expected behaviour of an ODRL Evaluator. This is a software component that evaluates the access and usage of assets based on a set of policies and the current state of the world and outputs the active permissions, prohibitions, and obligations, as well as any prohibitions or obligations that might have been violated or fulfilled. In the appendix, we have provided an example that represents a data access policy that allows a data user's smart home AI provider to read her smart home IoT data and support her in managing her home.

XACML is an alternative way to express data access policies. Choosing ODRL over XACML provides significant benefits when dealing with digital rights, content usage policies, and obligations associated with digital assets. Its compatibility with Semantic Web technologies, focus on expressiveness and flexibility, and ease of integration with data privacy vocabularies make it particularly suitable for scenarios where detailed usage rights and compliance with regulations like GDPR are essential. While XACML excels in fine-grained access control within enterprise environments, ODRL offers a more specialized and adaptable framework for managing digital rights and obligations in a broader range of contexts.

3.6.2.2. Data privacy vocabulary. The Data Privacy Vocabulary (DPV)[56] enables the representation of machine-readable data related to the processing of personal data based on data protection requirements such as the GDPR. Although DPV is designed as a jurisdiction-agnostic resource,

[50] https://www.w3.org/community/odrl/.
[51] https://www.w3.org/community/dpvcg/.
[52] https://www.w3.org/TR/odrl-model/.
[53] https://www.w3.org/TR/odrl-vocab/.
[54] https://w3c.github.io/odrl/profile-bp/.
[55] https://w3c.github.io/odrl/formal-semantics/.
[56] https://w3id.org/dpv.

adaptable to various legal requirements across different regions, it has been heavily influenced by the GDPR since its inception. As a result, it already includes extensions with GDPR-specific[57] and Data Governance Act specific[58] concepts [16], enabling the concrete linking of legal basis or rights to the relevant legal provisions. We argue that DPV can be instrumental in supporting the Data Act requirements in combination with ODRL, as we will explain below.

The core taxonomies provided in DPV include terms to describe multiple aspects. Entities refer to various roles and parties involved, such as legal roles like data controllers or data protection officers, as well as authorities, organizations, and different types of data subjects. These entities form the backbone of data governance structures, ensuring accountability and adherence to legal frameworks. Purposes include both commercial and altruistic objectives, like marketing or combating climate change under the Data Governance Act (DGA). Data and personal data are central, with defined categories ensuring compliance with particular requirements, e.g. special data categories identified in the GDPR are also classified as such in DPV. Processing operations and their context can be used to detail processing automation, sources, scale, and technologies, with an extension for describing technology functions and capabilities. Context covers duration, necessity, location, and legal considerations for compliance. Legal measures like contracts ensure accountability, while a taxonomy for consent can be used to provide details of processing operations based on this legal ground. Risk management terms help mitigate impacts, and rights-focused terms support transparency and data subject empowerment.

The integration of DPV's taxonomies with the ODRL standard is of great interest for the expression of data access and usage policies that are aligned with legal requirements [16,17]. We have provided an example in the appendix that illustrates how DPV can be applied to the ODRL policy of a data user, allowing her to set the data access and usage policies to her home IoT devices for energy management purposed. By using the concepts from taxonomies like DPV, data users can ensure semantic and technical interoperability in their policy management systems and use the corresponding policies as the basis for the creation of data sharing terms, including contractual or consent-based agreements.

3.6.2.3. Consent records. DPV has been successfully used to create consent records[59] aligned with the ISO/IEC TS 27560:2023 standard[60] on 'Privacy technologies – Consent record information structure' [18]. Maintaining standardised records of consent in a machine-readable format promotes interoperability of solutions, allowing data controllers to fulfil their consent-related reporting duties, while data subjects can be provided with meaningful details on how their data is being used. Such details includes data about their rights. In the appendix, we have provided an example consent record in the context of the smart home, using the required fields from ISO 27560:2023 (cf. Table 3).

3.6.2.4. Management and exercising of rights. A guidance document for the usage of DPV for rights exercise and management[61] is being developed with the goal of providing guidelines and examples on how to express information about the management and exercising of rights. It is related to the data subject rights enacted in the GDPR, using DPV and other semantic standards. In the appendix, an example of a rights exercise record related to a GDPR Right of Access request applied to the smart home example has been provided.

3.6.2.5. Provenance. DPV can be used with the W3C's PROV-O recommendation[62] to track the provenance of activities related to data management, policy management and for other compliance documentation generation. Using this standard, provenance information

[57] https://w3id.org/dpv/legal/eu/gdpr.
[58] https://w3id.org/dpv/legal/eu/dga.
[59] https://w3id.org/dpv/guides/consent-27560.
[60] https://www.iso.org/standard/80392.html.
[61] https://w3id.org/dpv/guides/rights.
[62] https://www.w3.org/TR/prov-o/.

can be represented, concerning the entities with whom the activity is associated, i.e., prov:wasAssociatedWith, or what data was generated, i.e., prov:generated, by the activity. Additionally, the DCMI Metadata Terms standard[63] can be used to record temporal information, descriptions and identifiers of the activities and their creators and/or publishers.

3.6.2.6. Digital catalog vocabulary (DCAT). The W3C's DCAT [64] standard is essential for connecting personal data store systems into a data space, as it was developed to enable interoperability between data catalogues published on the Web, and can be used to describe datasets and data services, using a standard model that facilitates the usage of data and metadata spread across multiple catalogues. Beyond being standards on their own, these resources are also recommended by DPV and ODRL, as well as multiple other standards, as state of the art resources to record information related to provenance, catalogues and their resources, and other contextual metadata-keeping.

3.7. Use case: The agentic home

As a use case to illustrate the architecture in Fig. 3, we have implemented a use case that would comply to the Data Act by design. The use case describes a future-looking application of agentic workflows in the home or the office. An agentic workflow describes a collaboration between agents. Such an agent can be understood as an algorithm that has a certain degree of autonomy in supporting the end-user in taking certain decision and/or taking certain actions. The agents can live both in the cloud or in the edge, depending on the use case. The most prevalent deployment mode for agents is currently cloud-based, yet many reasons exist for deploying agents in the edge. These include bandwidth reduction, data protection, latency reduction, energy expenditure reduction and critical infrastructure resilience. A useful example to consider would be security cameras that monitor the environment of a house. Sending all the image data continuously to the cloud would require a lot of bandwith, introduce privacy vulnerabilities during the movement of the data to the cloud and the storage in the datacentre and introduce latency due to the communication of the data to the datacentre and the communication back to the end-user. From an energy perspective, the communication of the data itself can be energy-intensive and sending the data to the cloud would forego the potential use of renewable energy that could be available to the devices through the energy production capability of the house (e.g. solar panels). Finally, a distributed deployment in the edge is less vulnerable to attacks on critical infrastructure than a cloud deployment from an architectural perspective, as the number of places to attack is higher.

Having the LLM agent itself operate in the edge would mainly be beneficial from a privacy perspective, as the data remains under the full control of the end-user. On the other hand, doing a full edge-based deployment introduces a key redundancy vulnerability: if something happens to the location of the edge deployment, in this case the house, the setup and the data can be lost. Therefore, a robust back system may be needed. Furthermore, cloud-based deployemts offer a convenience to the end-user in terms of management, as the infrastructure is managed and updated by the cloud provider. As a result, we expect that most of the real-world deployments of agentic systems will be hybrid, with different architectural components residing at different locations in the cloud-edge continuum.

In this application, we assume that the agentic workflow is deployed in a home, where multiple residents are living together. The agentic system exposes a generative AI large language model (LLM) based chatbot to the residents that allows them to interact with the different IoT devices that have been installed in the home, such as thermostats, HVAC systems, EV batteries, solar panels, kitchen appliances, etc. These devices are to be considered as connected products

[63] http://purl.org/dc/terms/.
[64] http://www.w3.org/ns/dcat.

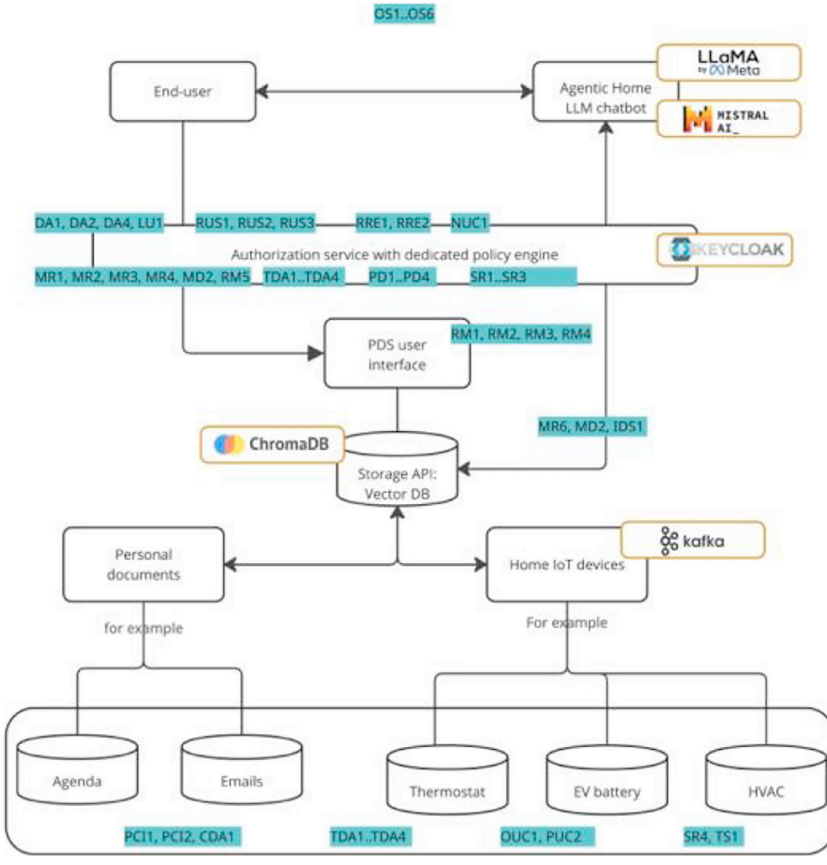

Fig. 3. Adaptation of the architecture to the use case of the agentic home. The codes in green refer to the technology-matrix codes in Table 1 and indicate what parts of the matrix are supported by which elements in the architecture.

who send their data to a central device that runs a large language model and serves as the text or voice-based user interface.

Fig. 3 shows how the architecture in Fig. 2 is adapted to the agentic home use case. In this architecture, the end-user accesses the personal data store (PDS) user interface to set the access and usage policy permissions that allow the LLM chatbot to access the data coming from the IoT devices in the smart home. These are stored as policies that are applied to vectors in a vector database. The choice was made to work on vectors, as these represent the embeddings that are necessary for an LLM to quickly retrieve the data it needs to reply to prompts by the end-user. This is very similar to the Retrieval Augmented Generation (RAG, 31) technique that has become a common practice for production-level implementations of LLMs in real-world deployments. The vector database, containing the numerical representations of the data in the form of long vectors, also known as “embeddings” (we used Chroma[65], an open-source vector DB, for our implementation), contains both the data coming from the IoT devices (in NGSI-LD format) as well as embeddings representing the personal data of the end-user. This is useful, as for example agenda data can indicate when the end-user is home and therefore should be available by the LLM chatbot. For example, this would mean that the chatbot could change the settings of the

[65] https://www.trychroma.com/.

thermostat or the HVAC system, understanding when the end-user is at home or not, to optimise both the quality of experience of the end-user as well as the energy bill. The proposed approach allows the decoupling of the IoT and personal documents from the algorithm, which means that, if the user chooses to, she can connect other algorithms to these data sources, effectively implementing multiparty data sharing as we have described it before. For the authorization engine and the policy engine, we used the Keycloak open-source system. The combination of API's for managing the data streams and the Keycloak component has been made available on GitHub under the name Kvasir[66], which aligns with the SOLID protocol and the ongoing work in in the W3C Linked Web Storage workgroup[67]. As is plain in Fig. 3, the majority of the codes in the technology-requirements matrix from Table 1 is supported by the Authorization server and dedicated policy engine architectural component. This indicates that this is the key component in our proposed architecture and of essential importance to supporting compliance with the Data Act. Not all codes from Table 1 have been mapped to the architecture of the use case. DA3, MR5, MD1 and OS5, which all map to the "Federated data catalogues" technology column in Table 1, have not been implemented in the use case. Indeed, our use case was mainly aimed at providing a proof-of-concept of how the architecture could be deployed for use in a single home. The addition of data catalogues would publish the availability of the data by placing the meta-data in the catalogue and would allow the level data playing field that is aimed for by the Data Act and the Data Spaces paradigm. This could be implemented by a component like the Eclipse EDC federated catalogue[68] and can leverage the DCAT RDF vocabulary[69].

This architecture has been deployed in an edge architecture, where the IoT data streams (published and accessed using Apache Kafka[70] through our Obelisk[71] data service platform) emanate from real devices in a real-life smart home. The data was published as NGSI-LD[72]. The home in which this was deployed and tested is imec / UGent's Home Lab[73] living lab, which is a fully equipped home that can is optimised for experimental IoT deployments. The LLM that was used to interface with the end-user was LLAMA 3.1 7B and Mistral 7B. We deliberately worked with smaller parameter count LLMs ($<$ 10 Billion), to be able to execute them on hardware configurations that could reasonably be expected to be available to consumers in the years to come. For our prototype, we used an NVIDA AGX Orin, which could act as a baseline for edge deployments in the near future. We do not expect most short-term deployments of LLM-based agents will be on the edge, as this introduces some key issues in terms of robustness and ease of use. However, possibilities of new products exists, where an LLM can run on an LLM-capable base station (e.g. integrated in a smart speaker, a set-top box, an home NAS,...), which is why we choose to explore this option.

Fig. 4 shows an example interaction of how the LLM assistant supports the end-user, in this case by changing the temperature in the house. This is possible, because the user had previously allowed the chatbot to access the thermostat in this way, by setting the relevant policy.

4. Discussion

While the question of what precisely entails switching between data processing services and which aspects of switching require further regulatory attention remains unanswered, our architectural framework suggests and offers a potential way forward. Additionally, the Data Act establishes a novel data portability regime. This complements the existing regime set forth by the

[66] https://gitlab.ilabt.imec.be/kvasir/kvasir-server.
[67] https://www.w3.org/groups/wg/lws/.
[68] https://github.com/eclipse-edc/FederatedCatalog.
[69] https://www.w3.org/TR/vocab-dcat-3/.
[70] https://kafka.apache.org/.
[71] https://idlab.ugent.be/resources/obelisk.
[72] https://obelisk.ilabt.imec.be/docs/tech_reference/standards_compatibility.html.
[73] https://homelab.ilabt.imec.be/index.html.

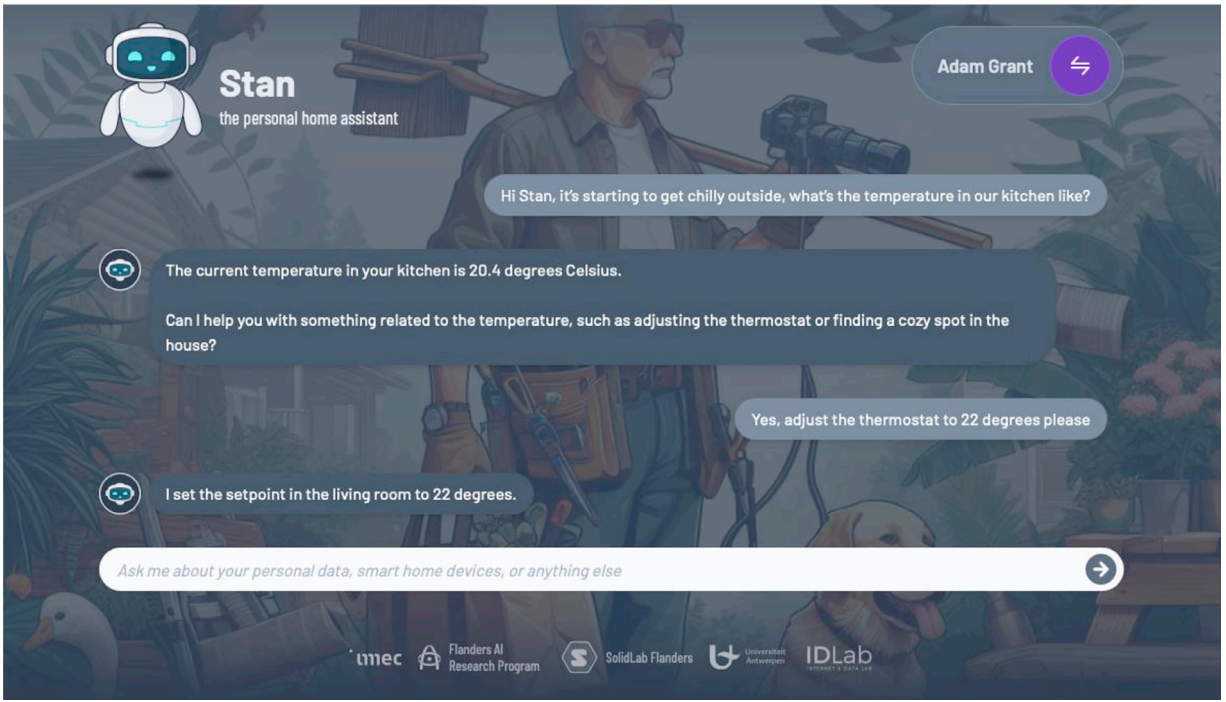

Fig. 4. Screenshot of the resulting LLM assistant interface, showing how the system can actuate the temperature in the user's home.

GDPR, the stricter portability regime for gatekeepers as outlined by the Digital Markets Act, and the data portability regime as proposed in sectorial data sharing regulation such as the European Health Data Space. The objective of our proposed architecture is to provide a comprehensive solution that is consistent with all existing portability regimes.

The Data Act envisions a mandatory, legally imposed framework for enabling data sharing between data holders and users. This creates opportunities for the automation of contracts and obligations using an architectural framework, such as the one proposed in the article, which could facilitate compliance. In future work, an automated data sharing contract and policy management negotiation system would be valuable to negotiate contractual access to and usage of data. The architecture proposed in this paper can provide a basis to do so.

The Data Act does not fully balance the objectives of data holders and data users. For instance, vendors, as data holders, still have the possibility to retain substantial control over generated connected product data due to limitations in user rights, unresolved market failures, and legal inconsistencies. Our proposed architecture suggests that vendors design connected products with the intention of granting end-users direct control over their data. For new products, this could entail affordances for allowing access to the data in a distributed way, where the data access and usage policies are regulated at the device level. For existing products, a middle layer could be introduced that regulates the data access and usage for a whole set of devices under the control of the end-user, using an approach akin to a data space connector. It is customary for such a connector to possess two functional levels. The first is the control plane, which is responsible for identity management, access control, contract enforcement and transaction logging. The second is the data plane, which executes of data transfer itself. Such a middle layer would benefit from the proposed architectural framework, building on recognised standards such as governed by W3C or ETSI and more specifically the work that is currently underway in the W3C Linked Web Storage (LWS) workgroup. In conjunction with DPV and ODRL, for instance, SOLID and LWS can contribute to the establishment of a uniform interface for personal data. It follows that these layers form a basis for different data space connectors to build on and that it would be useful for the personal data store and data space connector paradigms to become more integrated. This approach aligns with the spirit of both the Data Act and the GDPR and presents a next step towards a compliance by design that builds on personal data stores and can be combined with the data space paradigm.

We are looking forward to deployment of SIMPL (Smart Middleware Platform), which is an open-source initiative by the European Commission designed to facilitate secure, interoperable, and scalable data sharing across European data spaces. It serves as a foundational infrastructure to support the EU's vision of a unified digital economy where data can be shared seamlessly across sectors and borders while ensuring data sovereignty and compliance with regulations like the GDPR. The first deployments done at imec of SIMPL components showed promising results in terms of functionality and ease of use. However, preliminary tests of pre-release components showed that these were resource-heavy, with heavy computational and memory demands that may be beyond the reach of small businesses and would certainly be prohibitive towards edge developments, like in people's homes.

However, giving the end-user the possibility to regulate data sharing also represents further challenges, especially from the perspective of usability. Indeed, many end-users may care for the way in which their data is shared, but will not be willing and/or able to invest the time and effort needed to perform the consent management that is needed to configure and maintain data access and usage policies. Therefore, research is needed that supports end-users to do this. The user experience and data processing abilities that are provided by large language models offer an interesting opportunity to do so.

To support the R&D strategy that was set forth by the EU to support the Digital Strategy, it would be essential to further embed the personal data store technology in the data space paradigm. Indeed, there have been multiple initiatives to develop reference architectures (e.g. by the DSSC, IDSA, and GAIA-X) and matching components for creating data spaces in various commercial and societal sectors in the EU. Concepts like identity management, data access and usage policies, catalogues, etc… are common to both the data space and the personal data store

paradigms. Therefore, we believe it to be a feasible and worthwhile challenge to develop and deploy personal data store systems that support the Data Act by design and that can function in the context of a data space. The key component to integrate the world of personal data stores with data spaces is the data space connector. We therefore strongly advocate that these two worlds become more closely aligned and that data space connector technologies could be made compatible with personal data sharing. The W3C's Linked Data Storage effort may be a good forum to make further steps in this direction.

5. Conclusion

In this paper, we have analysed the requirements that emanate from the EU's Data Act towards multiparty data access in connected products as part of the EU's Data Strategy. A first contribution has been the description of the requirements as 6 different categories (Data accessibility for users, Information transparency for users, User rights regarding data, Fair and transparent data sharing, Interoperability, Protection of legitimate interests), each with several subcategories. We have explained each of these categories and have proposed a conceptual model on how to address these requirements. As there is at the time of writing some unclarity about how to interpret the Data Act, as with each new piece of legislation, we hope to have shed some light through our interpretation.

Additionally, we have proposed a technological conceptual model that can be leveraged to address the Data Act's requirements. The key elements in this conceptual model are identity management, consent management, data access policies, data usage policies, data catalogues, confidential compute and linked data. We have analysed how these technological concepts can support the various Data Act requirements and how they can fit together in an architecture. To give more practical resonance to the above, we have provided an insight into a deployment of the architecture in an "agentic home" use case application. Besides serving as a practical example, the use case illustrates how personal data spaces in an edge deployment can allow future generative AI applications, that are highly probable to drive the information technology innovation roadmaps for the years to come. By banking on such technologies, we believe EU organisations, both commercial and non-commercial, can secure the EU's competitive position in the AI domain, while preserving the Union's ethical norms and values. For years, the data space paradigm has been a highly valuable and forward looking effort to prepare the EU's economy for the digital services of the future. This future is now here in the form of generative AI as the "killer app" for data spaces and it is our perspective that both the Data Act and personal data store technologies can play a critical role in enabling it, if embedded in a data space context.

Limitations

A key limitation is that this paper focuses solely on personal data stores, omitting a broader discussion on data spaces in general. While the Data Act encompasses various data types beyond personal data, we have intentionally limited our scope to personal data. Consequently, our analysis is confined to examining how data spaces can address certain requirements related to multiparty personal data access, as outlined by the Data Act.

This approach also reflects a practical limitation: we have selected specific requirements from the Data Act based on their relevance to our technological focus. While this selection process could be critiqued as a teleological approach lacking a broader argumentative basis, the authors believe it establishes a crucial technical foundation for navigating the complexities of Data Act compliance. Still, we have aimed to be as exhaustive as possible in listing and categorising the relevant Data Act requirements.

We acknowledge the role of dispute settlement bodies, which have the authority to address issues related to safety, security requirements, the identification of trade secrets, and disputes

over fair, reasonable, and non-discriminatory (FRAND) conditions, as well as the fairness of contractual terms. However, the nuances of these dispute resolutions and their impact on data-sharing practices are beyond the scope of this paper.

Finally, we would like to provide a disclaimer regarding the incorporation of multiple standards (W3C) in this paper, as a means to comply with certain requirements stemming from the Data Act. Whilst various mechanisms, including standards, have traditionally been employed to facilitate the alignment between principles in the Data Act and their operationalization, the establishment of these standards alone does not fully address the need for effective compliance and legal certainty among the parties subject to the Data Act. An illustration of this restriction can be found in the implementation of ISO/IEC 29134 for conducting risk assessments and conducting DPIAs under the GDPR. The overarching objective of this ISO standard is to establish a framework for conducting privacy impact assessments. In practice, this standard deviates from the logic of an Information Security Management System, which does not fully recognise the legal risks to fundamental rights and freedoms of the individual. This chapter acknowledges the absence of a unified standard or methodology for facilitating compliance with the Data Act's requirements. Alternatively, a plurality of methodologies may be employed. The selection of methodology will be contingent on the sector, the specific concerns, the safeguards taken with regard to the risks to fundamental rights and the data at stake [19].

Ethics Statement

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

Data Availability

Kvasir (Original data) (gitlab).

CRediT Author Statement

Tanguy Coenen: Conceptualization, Methodology, Writing – original draft, Supervision; **Michiel Fierens:** Conceptualization, Formal analysis, Writing – original draft; **Beatriz Esteves:** Conceptualization, Methodology, Writing – original draft; **Esther De Loof:** Writing – review & editing.

Acknowledgments

This research received funding from The Flanders AI research program and SOLIDLab Flanders.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work the author(s) used Microsoft Copilot to improve language and readability. ChatGPT 4o was used to do quick searches of the relevant legislative texts and to format the bibliographic references. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2025.111949.

References

[1] A.R. Hevner, S.T. March, J. Park, & S. Ram. Design Science in Information Systems Research MIS Quarterly, 28(1), 75-105 (2004).
[2] B. Martens. Pro- and anti-competitive provisions in the proposed European Union Data Act, Bruegel Working Paper 01/2023 (2023), https://www.bruegel.org/working-paper/pro-and-anti-competitive-provisions-proposed-european-union-data-act.
[3] M. Hennemann, G.K. Ebner, K. Benedikt, G. Lienemann, & M. Wienroeder. (Eds.) Data Act: an introduction Nomos Verlagsgesellschaft mbH & Co. KG, 35-48 (2024), https://directory.doabooks.org/handle/20.500.12854/139057.
[4] W. Termont, R. Dedecker, W. Slabbinck, B. Esteves, B. De Meester, & R. Verborgh. From resource control to digital trust with user-managed access, arXiv preprint arXiv:2411.05622 (2024), https://arxiv.org/abs/2411.05622.
[5] P. Mell, & T. Grance. The NIST definition of Cloud Computing NIST Special Publication 800-145 (2011), https://doi.org/10.6028/NIST.SP.800-145.
[6] F. Casolari, M. Taddeo, A. Turillazzi, et al., How to improve smart contracts in the European Union Data Act, DISO 2 (2023) 9, doi:10.1007/s44206-023-00038-2.
[7] B. Esteves, & V. Rodríguez-Doncel. Analysis of ontologies and policy languages to represent information flows in GDPR Semantic web, 15(3), 709-743 (2024), doi: 10.3233/SW-223009.
[8] N.T. Wiedemann. Bereitstellung von Daten nach dem Data Act – offene fragen und verbleibende probleme, 157-161 (n.d.).
[9] I. Graef, & M. Husovec. Seven things to improve in the Data act (SSRN Scholarly Paper No. 4051793) Social science research network (2022).
[10] M. Götz, P. in der auf D. spezialisierten K. N. & G. in, & Braunschweig, S. B. ist R. am O. Götz/Blöink: datenvertrag: Lösungsansatz für das Spannungsfeld zwischen Data Act und DS-GVO Multimedia und Recht, 6, 451–456 (2024).
[11] A. Poikola, K. Kuikkaniemi, O. Kuittinen, H. Honko, A. Knuutila, V. Lähteenoja, MyData – An Introduction to Human-Centric Use of personal Data, 3rd revised edition, Ministry of Transport and Communications, Finland (2020) https://mydata.org/wp-content/uploads/2020/08/mydata-white-paper-english-2020.pd.
[12] P. Fehlinger. Enabling the responsible use of technology at scale Sitra (October 12) (2023), https://www.sitra.fi/en/publications/enabling-the-responsible-use-of-technology-at-scale/.
[13] C. Ducuing, T. Margoni & L. Schirru. CiTiP White paper on the Data Act Proposal KU Leuven Centre for IT & IP Law (CiTiP) (2022), https://www.law.kuleuven.be/citip/en/Publications/citip-whitepaperdataact.pdf.
[14] B. Esteves, V. Rodríguez Doncel, H.J. Pandit, D. Lewis, in: Semantics for implementing data reuse and altruism under EU's data governance act in knowledge graphs: semantics, machine learning, and languages, IOS Press, 2023, pp. 210–226, doi:10.3233/SSW230015.
[15] H.J. Pandit, & B. Esteves. Enhancing data use ontology (DUO) for health-data sharing by extending it with ODRL and DPV Semantic Web, 15(4), 1473-1498 (2023), doi: 10.3233/SW-243583.
[16] H.J. Pandit, J. Lindquist, G.P. Krog, in: Implementing ISO/IEC TS 27560: 2023 Consent Records and Receipts for GDPR and DGA In Annual Privacy Forum, Springer Nature Switzerland, Cham, 2024, pp. 228–251, doi:10.1007/978-3-031-68024-3_12.
[17] A. Christofi, P. Dewitte, C. Ducuinc, P. Valcke, Erosion by standardisation: is ISO/IEC 29134:2017 on privacy impact assessment up to (GDPR) standard? in: Personal Data Protection and Legal Developments in the European Union, IGI Global Scientific Publishing, 2020, pp. 140–167, doi:10.4018/978-1-5225-9489-5.ch008.
[18] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, D. Kiela, Retrieval-augmented generation for knowledge-intensive NLP tasks, Adv. Neural Inform. Proces. Syst. 33 (2020) 9459–9474 arXiv:2005.11401.
[19] E. Habich. FRAND Access to data: perspectives from the FRAND Licensing of Standard-essential patents for the Data Act Proposal and the Digital Markets Act IIC - International Review of Intellectual Property and Competition Law, 53(9), 1343–1373 (2022).