The principles of data minimisation and purpose limitation in Common European Data Spaces: sharing is caring?

Arne Vincken and Bente Schockaert

Abstract. This paper analyses the tension between the General Data Protection Regulation (GDPR) principles of data minimisation and purpose limitation and the data-sharing framework of Common European Data Spaces (CEDS), particularly as promoted by the Data Act. While CEDS aim to enable large-scale data sharing across the EU, the GDPR restricts processing to what is necessary and compatible with the original purposes, creating potential legal conflicts. After outlining the EU's data space strategy and regulatory framework, including the recent Digital Omnibus Proposal, the paper examines the reshaped definition of personal data and its implications for GDPR applicability. It then explores how privacy by design can operationalise data minimisation and purpose limitation within data spaces, focusing on privacy-enhancing technologies (e.g., federated learning, secure multi-party computation). The paper identifies gaps in the literature and offers recommendations for interpreting and applying these principles in the context of CEDS.

Keywords: GDPR, Data Act, Data minimisation, Purpose limitation, Privacy by design, Common European Data Spaces (CEDS)

1 Introduction

In its European strategy for data,[1] the European Commission has set the scene for the creation of Common European Data Spaces (CEDS). Against the backdrop of data-driven innovation, data spaces must enable data sharing across the EU, based on common standards and in compliance with EU law. In particular, the GDPR applies when personal data is being processed in data spaces. This strong framework of data protection raises several questions in the context of data spaces, asking for rethinking privacy regulations.[2] The development of data spaces brings new challenges that require interdisciplinary research to address[3]. In particular, while the GDPR limits data processing, regulations regarding data spaces enhance data availability. As a result, tensions may arise between the GDPR principles of data minimisation and purpose limitation on the one hand, and regulations promoting data sharing in data spaces on the other hand.[4] This paper analyses to what extent this tension exist in the context of data spaces and how it may be addressed through operationalizing privacy by design and by default. Doing so, the paper addresses the following research question: "How could the GDPR principles of data minimisation and purpose limitation be ensured and operationalised within the data-sharing framework of Common European Data Spaces?"

First, this paper explains the legal and technical design of data spaces, as well as the importance of the principles of data minimisation, purpose limitation and privacy-enhancing technologies (PETs). Second, it analyses in more detail to what extent a legal tension exist between data sharing regulations and data protection in the context of

[1] European Commission, 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, 19 February 2020, COM(2020) 66 final.

[2] Ianese S., 'Challenging Data Protection in the Age of Data Spaces | La Sfida Della Protezione Dei Dati Personali Nell'era Dei Data Spaces', *European Journal of Privacy Law & Technologies*, no. 0 (2025), 6, https://university-press.unisob.na.it/ojs/index.php/ejplt/article/view/2002.

[3] Fierens M. (2024), Initiating Interdisciplinary Research for Future-Proof Data Protection in the Context of Data Spaces and Semantic Interoperable Data Sharing, Proceedings of the 1st NeXt-generation Data Governance Workshop 2024 (NXDG 2024) co-located with 20th International Conference on Semantic Systems (SEMANTiCS 2024), 2.

[4] Ianese S., 'Challenging Data Protection in the Age of Data Spaces | La Sfida Della Protezione Dei Dati Personali Nell'era Dei Data Spaces', *European Journal of Privacy Law & Technologies*, no. 0 (2025), 8, https://university-press.unisob.na.it/ojs/index.php/ejplt/article/view/2002; Curry E. et al., eds, *Data Spaces : Design, Deployment and Future Directions* (Springer Nature, 2022), 317, https://doi.org/10.1007/978-3-030-98636-0; Fierens M. (2024), Initiating Interdisciplinary Research for Future-Proof Data Protection in the Context of Data Spaces and Semantic Interoperable Data Sharing, Proceedings of the 1st NeXt-generation Data Governance Workshop 2024 (NXDG 2024) co-located with 20th International Conference on Semantic Systems (SEMANTiCS 2024), 2.

CEDS. Lastly, potential technological measures to minimise the tension are thoroughly reviewed, including privacy by design, PETs and Identity Access Management.

2 Conceptual analysis

2.1 Common European Data Spaces

Data sharing is a crucial part of the EU's strategy to foster data-driven innovation and competitive AI development.[5] Access to large volumes of high-quality data is a key driver for productivity and innovation and essential for the training of AI systems.[6] By facilitating data sharing, the EU aims to promote a competitive and fair economy and strengthen its strategic autonomy in the global digital landscape.[7] To this end, the European Commission has introduced Common European Data spaces, described as *"datasharing ecosystems built on cloud infrastructure and clear governance rules defining who can access, use, and share data."*[8] All private and public actors should be able to share data in a secure and trustworthy manner while maintaining control over their data.[9] The EU is rolling out various sectoral data spaces, but the ultimate goal is to create a genuine single market with data as the fifth freedom.[10]

The concept of data spaces was firstly introduced in 2005 by Franklin, Halvey and Maier[11] as a technical solution for data management.[12] In general, a data space can be understood as an interoperable framework built upon shared governance principles, technical standards, and enabling services that facilitate trusted data transactions among participants. There is, however, no universally accepted definition.[13] One definition was formulated by the European Committee for Standardization (CEN) and adopted by the Data Spaces Support Centre[14]: "*Interoperable framework, based on common governance principles, standards, practices and enabling services, that enables trusted data transactions between participants*."[15] A more technical definition was proposed by the Big Data Value Association (BDVA), an organisation that unites industry members and advances value creation of Big Data and

[5] European Commission, Communication from the Commission to the European Parliament and the Council, *Data Union strategy: unlocking data for AI*, 19 November 2025, COM(2025) 835 final.

[6] European Commission, Communication from the Commission to the European Parliament and the Council, *Data Union strategy: unlocking data for AI*, 19 November 2025, COM(2025) 835 final.

[7] European Commission, 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*, 19 February 2020, COM(2020) 66 final; European Commission, Communication from the Commission to the European Parliament and the Council, *Data Union strategy: unlocking data for AI*, 19 November 2025, COM(2025) 835 final.

[8] European Commission, Communication from the Commission to the European Parliament and the Council, *Data Union strategy: unlocking data for AI*, 19 November 2025, COM(2025) 835 final.

[9] European Commission, 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*, 19 February 2020, COM(2020) 66 final; European Commission, Communication from the Commission to the European Parliament and the Council, *Data Union strategy: unlocking data for AI*, 19 November 2025, COM(2025) 835 final.

[10] European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *A European strategy for data*, 19 February 2020, COM(2020) 66 final; European Commission, *Commission staff working document on Common European Data Spaces*, 23 February 2022, SWD(2022) 45 final.

[11] Franklin, M., Halevy, A., & Maier, D. (2005). From databases to dataspaces: a new abstraction for information management. ACM SIGMOD Record, 34(4), 27–33. https://doi.org/10.1145/ 1107499.1107502.

[12] Edward Curry et al., eds, *Data Spaces : Design, Deployment and Future Directions* (Springer Nature, 2022), https://doi.org/10.1007/978-3-030-98636-0, 3.

[13] See for a list of definitions from the literature: Edward Curry et al., eds, *Data Spaces : Design, Deployment and Future Directions* (Springer Nature, 2022), https://doi.org/10.1007/978-3-030-98636-0, 4.

[14] The Data Spaces Support Centre is an association funded by the EU that contributes to the creation of common requirements and best practices: https://dssc.eu/space/Partners/175472674.

[15] The European Committee for Standardization (CEN), *CEN Workshop Agreement*, July 2024, CWA 18125:2024 (E), https://dssc.eu/space/BVE2/1071251613/Introduction+-+Key+Concepts+of+Data+Spaces#1.1-What-about-a-definition?.

AI: "*Data Spaces is an umbrella term corresponding to any ecosystem of data models, datasets, ontologies, data sharing contracts and specialised management services (i.e., as often provided by data centres, stores, repositories, individually or within 'data lakes'), together with soft competencies around it (i.e., governance, social interactions, business processes). These competencies follow a data engineering approach to optimize data storage and exchange mechanisms; in this way preserving, generating and sharing new knowledge.*"[16] As can be seen, there are various definitions, and furthermore, each data space is unique, based on specific principles and governance rules.[17]

Despite the absence of a single, universally accepted definition, most data spaces have several common features consistently identified across various definitions and implementations. Data spaces are based on a decentralised architecture, where the data remains with the owner instead of being stored in a central repository with a single point of control.[18] Founded on the principle of data sovereignty, participants retain full control over their data and its usage terms.[19] In addition, interoperability is implemented through common technical standards and semantic models, ensuring that different participants and systems can work together.[20] Furthermore, data spaces establish governance rules defining roles, responsibilities and processes.[21] This is complemented with mechanisms to establish trust among participants,[22] as well as robust security and privacy mechanisms.[23] Data discoverability is facilitated through standardised catalogues. To define usage conditions, data spaces provide mechanisms to negotiate data-sharing agreements.[24] Also, they promote observability by monitoring data use and transactions in a transparent manner.[25] Together, these features facilitate data sharing, improving data value creation.[26] The EU has already set out an extensive Digital Rulebook promoting data sharing, including, amongst others, the Data Governance Act and the Data Act.[27] Tnene he European Commission published a proposal of a Digital Omnibus in order to simplify and optimise legislation.[28] One of the objectives is to coherently unify the single market rules on data sharing in the Data Act.[29]

---

[16] Big Data Value Association (BDVA), *Towards a European data sharing space: Enabling data exchange and unlocking AI potential*, BDVA Position Paper April 2019, https://bdva.eu/wp-content/uploads/pdfs-legacy/BDVA%20DataSharingSpace%20PositionPaper_April2019_V1.pdf.

[17] Fierens M. (2024), Initiating Interdisciplinary Research for Future-Proof Data Protection in the Context of Data Spaces and Semantic Interoperable Data Sharing, Proceedings of the 1st NeXt-generation Data Governance Workshop 2024 (NXDG 2024) co-located with 20th International Conference on Semantic Systems (SEMANTiCS 2024), 2.

[18] Data Spaces Support Centre (DSSC), *Data Spaces Blueprint, Version 3.0* (DSSC, 2025), "Organisational Form and Governance Authority" and "Access and Usage Policies Enforcement" building blocks, https://blueprint.dssc.eu.

[19] IDSA, *IDS Reference Architecture Model 4.0* (n. 1), sec. 3.2.1, identifying data sovereignty as a foundational design goal requiring participants to retain control over their data and its usage conditions; DSSC, *Data Spaces Blueprint, Version 3.0* (n. 1), "Access and Usage Policies Enforcement" building block.

[20] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data (hereafter: Data Act), OJ L 2023/2854, recital 103 (requiring standardisation and semantic interoperability within and among common European data spaces); IDSA, *IDS Reference Architecture Model 4.0* (n. 1), sec. 3.2.4 (standardized interoperability); DSSC, *Data Spaces Blueprint, Version 3.0* (n. 1), "Data Interoperability" technical building block.

[21] IDSA, *IDS Reference Architecture Model 4.0* (n. 1), secs. 3.2.1–3.2.2 (governance roles and responsibilities, including certification bodies and governance authorities); DSSC, *Data Spaces Blueprint, Version 3.0* (n. 1), "Organisational Form and Governance Authority" building block.

[22] IDSA, *IDS Reference Architecture Model 4.0* (n. 1), sec. 3.2.1, describing trust as comprising identity management, certification, and usage policy enforcement; see also Monika Huber et al., " Building Trust in Data Spaces" in *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*, ed. Boris Otto et al. (Springer, 2022), 147–164.

[23] Data Spaces Support Centre (DSSC), *Data Spaces Blueprint, Version 3.0* (DSSC, 2025), "Access and Usage Policies Enforcement" building block, https://blueprint.dssc.eu.

[24] IDSA, *IDS Reference Architecture Model 4.0* (n. 1), sec. 3.2.1 (contract and policy negotiation as prerequisite for trusted data exchange); DSSC, *Data Spaces Blueprint, Version 3.0* (n. 1), "Contractual Framework" legal building block, which addresses data sharing agreements and usage conditions between participants.

[25] Sebastian Steinbuss, *Observability in Data Spaces* (International Data Spaces Association, June 2025), https://doi.org/10.5281/zenodo.15647484.

[26] Big Data Value Association (BDVA), *Position Paper: Towards a European Data Sharing Space* (BDVA, April 2019), 6.

[27] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), *Oj.L.* 3 June 2022; Recital 4 Data Act.

[28] Proposal for a Regulation of the European Parliament and of the Council amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus), COM(2025) 837 final.

[29] European Commission, Commission Staff Working Document 19 November 2025, SWD(2025) 836 final, 8.

The EU has already set out an extensive data legislative acquis. Different legislative acts were adopted to promote data sharing and set the framework for data spaces. Firstly, the Data Governance Act aims to enhance trust in data sharing by laying down governance requirements and obligations for data intermediation service providers.[30] It also creates the European Data Innovation Board (EDIB), which must coordinate technical standardisation.[31] Secondly, the Data Act goes a step further in promoting access and use of data with the main objective to remove barriers to data sharing.[32] *Inter alia*, the Data act enables users of connected products (IoT) to access the data they produce and share it, regulates business-to-business (B2B) data sharing agreements and set rules on mandatory business-to-government (B2G) data sharing. In particular, the Data Act lays down requirements for participants in data spaces as well as grounds for harmonised standards and interoperability. Thirdly, the AI Act contributes to trustworthy data use in the context of AI systems.[33] Lastly, the eIDAS 2 establishes the European Digital Identity Wallet, which enables trusted sharing of data linked to identity.[34] Digital wallets may possibly serve as a solution to secure access to data spaces. These legislative acts form part of the broader legal framework governing data.[35] The European Commission published a proposal of a Digital Omnibus in order to simplify and optimise legislation.[36] One of the objectives is to coherently unify the single market rules on data sharing in the Data Act.[37]

2.2 The principles of purpose limitation and data minimisation

While promoting data sharing, the EU also provides a strong data protection framework enshrined in the GDPR. Two of the foundational principles governing the lawful processing of personal data under the GDPR are purpose limitation and data minimisation, both enshrined in Article 5(1) of the Regulation. Purpose limitation requires that personal data be collected for specified, explicit, and legitimate purposes and not further processed in a manner incompatible with those purposes.[38] This principle operates at two levels: first, the purposes must be identified with sufficient clarity at the point of collection; and second, any subsequent processing must remain compatible

[30] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), *Oj.L.* 3 June 2022.

[31] Article 29 (1) Data Governance Act; European Commission, Commission decision of 20 February 2023 setting up the European Data Innovation Board, C(2023) 1074 final.

[32] Recital 4 Data Act.

[33] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 12 July 2024.

[34] Recital 7 Regulation (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework, *Oj.L.* 30 April 2024.

[35] The legal framework includes rules on the free-flow of non-personal data were already established: Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, *Oj.L.* 28 November 2018. Also, the re-use of data in the public sector is encouraged by the Open Data Directive: Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), *Oj.L*. 26 June 2019. The Digital Markets Act imposes sharing obligations on "gatekeepers" of large digital platform services: 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), *Oj.L.* 12 October 2022. The Interoperable Europe Act lays down measures for interoperability in the public sector: Regulation (EU) 2024/903 of the European Parliament and of the Council of 13 March 2024 laying down measures for a high level of public sector interoperability across the Union (Interoperable Europe Act), *Oj. L.* 22 March 2024.

[36] Proposal for a Regulation of the European Parliament and of the Council amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus), COM(2025) 837 final.

[37] European Commission, Commission Staff Working Document 19 November 2025, SWD(2025) 836 final, 8.

[38] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (hereafter: GDPR), OJ L 119/1 (2016), art. 5(1)(b).

with those original purposes.[39] Compatibility is not a binary test but rather an assessment of contextual integrity, requiring consideration of factors such as the link between the original and envisaged purposes, the reasonable expectations of data subjects, and the nature of the data concerned.[40] The principle of data minimisation, in turn, demands that personal data be adequate, relevant, and limited to what is strictly necessary in relation to the purposes for which they are processed.[41] This principle demands a deliberate assessment of necessity, where controllers must demonstrate that the data processing is indispensable to achieving the objective and that no less intrusive means are available.[42] The European Data Protection Board clarified that data minimisation is a constitutive element of the data protection by design and by default obligation under Article 25 GDPR, meaning that compliance must be built into processing systems from the outset rather than applied retrospectively.[43] Taken together, these principles serve as a critical safeguard against the indefinite accumulation and repurposing of personal data, which is particularly challenging in modern contexts such as blockchain technology where data immutability can conflict with these requirements.

However, the requirements for the practical implementation of these principles remains unclear. While Article 5(1)(b) requires purposes to be "specified" and "explicit", general descriptions such as "improving user experience", "future research", or "developing data-driven use cases" are widely considered insufficient.[44] WP29 emphasised that purposes must be "clearly revealed, explained or expressed in some intelligible form" no later than at collection,[45] but the assessment remains context-dependent. Paradoxically, overly long or legally dense descriptions may themselves hinder understanding, leaving data subjects and authorities uncertain about the true scope of processing.[46] In addition, the principle of data minimisation in the era of AI and Big Data poses several challenges. The use and sharing of large volumes of data may not be “necessary” in a strict sense.

2.3 Reshaped definition of personal data and importance of PETs

The recent *SRB v EDPS* judgment and the Digital Omnibus Proposal (DOP), have reshaped the concept of personal data.[47] The Court held in the *SRB* ruling that the classification as “personal data” depends on whether the recipient of pseudonymised data reasonably can identify the data subjects.[48] This ruling departs from earlier “absolute” interpretations, where pseudonymised data was generally classified as personal data for each recipient. The DOP's introduces the legal anchoring of the “relative” approach to identifiability.. Entities who reasonably cannot identify the natural persons to whom the data relates, would principally fall outside the scope of personal data.[49]In light of this reshaped definition of personal data, the importance of PETs is reinforced. PETs are no longer simply regarded as additional data protection measures, but become institutionalised elements determining whether the GDPR applies.[50] Thus, the “relative” definition of personal data possibly facilitates data sharing by means of technological measures.

---

[39] GDPR, recital 39; Article 29 Data Protection Working Party, *Opinion 03/2013 on Purpose Limitation* (WP203, April 2, 2013), 11–12.

[40] GDPR, art. 5(1)(b); see also recital 50, which clarifies that processing for a purpose other than that for which the data were collected is permissible only where it is compatible with the original purpose, taking into account factors such as the link between the purposes, the context of collection, and the nature of the data.

[41] Ibid., art. 5(1)(c).

[42] European Data Protection Board, *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default* (adopted October 20, 2020), paras. 73–76

[43] Ibid.

[44] Article 29 Data Protection Working Party, *Opinion 03/2013 on Purpose Limitation* (WP203, April 2, 2013), 15–17; Agencia Española de Protección de Datos (AEPD), *Approach to Data Spaces from a GDPR Perspective* (AEPD, 2023), 28-32, https://www.aepd.es/documento/approach-to-data-spaces-from-gdpr-perspective.pdf.

[45] Article 29 Data Protection Working Party, *Opinion 03/2013* (n. 13), 17.

[46] Regina Becker et al., "Purpose Definition as a Crucial Step for Determining the Legal Basis under the GDPR: Implications for Scientific Research," *Journal of Law and the Biosciences*, no. 1 (2024), 4-11

[47] Case T-557/20, *Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS)*, ECLI:EU:T:2023:219; Digital Omnibus Proposal, 19.

[48] Case T-557/20, *Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS)*, ECLI:EU:T:2023:219, n°

[49] Article 3 Digital Omnibus Proposal.

[50] Kenji Suzuki, “Privacy Enhancing Technologies in the EU Digital Omnibus Proposal: Clarifying the Role of PETs in the Definition of Personal Data under the GDPR” in New Frontiers in *Artificial Intelligence. JSAI-isAI 2026. Lecture Notes in Computer Science*, vol 16608, ed. Satoshi Morinaga et. al. (Springer, 2026), 87-98.

3 Legal assessment of the tension

3.1 Goal compatibility

At the level of stated objectives, the tension is less acute than it first appears. The Data Union Strategy frames CEDS not as an exception to data protection but as an ecosystem operating "within and across sectors" under "clear governance rules," while the GDPR itself has never treated data flows as inherently undesirable - Article 1(3) enshrines free movement of personal data as a Regulation objective standing alongside protection of the fundamental right.[51] Both instruments therefore claim to pursue trusted data circulation rather than unrestricted circulation. The apparent conflict crystallises only once "trusted" is translated into operational terms: for the data economy, trust is generated by scale, interoperability and predictable access rights; for the GDPR, trust is generated by necessity, specificity and the individual's retained control.[52]

Purpose limitation and data minimisation are not incidental technicalities within this scheme - the Court of Justice has repeatedly treated data minimisation as an expression of the broader principle of proportionality inherent in Article 8 of the Charter of Fundamental Rights, meaning that any relaxation of the principle for reasons of economic efficiency must itself satisfy a proportionality test rather than being assumed away by legislative fiat.[53]

This constitutional anchoring is precisely why the tension cannot be dissolved by definitional adjustment alone, such as the Digital Omnibus Proposal's (DOP) relative approach to identifiability: even where data escapes qualification as personal data for one controller, the CJEU's case law on generalised and indiscriminate collection suggests that the underlying proportionality logic re-emerges wherever identifiability risk resurfaces through combination or re-identification, which is structurally endemic to data-space architectures designed for pooling and secondary use.[54]

The compatibility test set out in Recital 50 GDPR offers a further indication that goal compatibility is achievable in principle.[55] Further processing for a new purpose is permitted where that purpose is compatible with the original one, assessed by reference to the link between the purposes, the context of collection, the nature of the data, the consequences for the data subject, and the existence of appropriate safeguards. This multi-factor test is inherently capable of accommodating data-space reuse scenarios, since it does not require identity of purpose but only a demonstrable, contextually reasonable link - precisely the kind of assessment that a well-designed data space, with documented provenance and purpose metadata, is better placed to perform than an ordinary bilateral data transfer. The difficulty is not that the GDPR's compatibility architecture is unsuited to data spaces, but that the current generation of data-space infrastructure rarely operationalises the Recital 50 factors in a structured, auditable way, leaving compatibility assessments to be reconstructed informally, if at all, by downstream data recipients.[56]

A genuine "golden mean" is nevertheless conceivable, though it is narrower than data-space proponents often suggest. The EHDS illustrates the point: Article 66 explicitly binds secondary use to data minimisation and purpose limitation, requiring anonymisation as the default and pseudonymisation only where the secondary-use purpose cannot otherwise be achieved, while Article 71 introduces a reversible opt-out as a structural safeguard against purpose fragmentation.[57] This is not a relaxation of the principles but their operationalisation through ex ante permitting: health data access bodies translate abstract necessity assessments into concrete, purpose-bound data permits, which is precisely the kind of "deliberate assessment of necessity" the EDPB guidance on Article 25

[51] Article 1(3) Regulation (EU) 2016/679 (GDPR) OJ L 119, 4 May 2016; European Commission, Communication, Data Union Strategy: Unlocking Data for AI,19 November 2025, COM(2025) 835 final.

[52] International Data Spaces Association, IDS Reference Architecture Model 4.0(IDSA, 2022) sec 3.2.1.

[53] Case C-446/21 Schrems v Meta Platforms Ireland Ltd [2024] ECLI:EU:C:2024:834, paras 59–60, citing Case C-175/20 Valsts ieņēmumu dienests [2022] ECLI:EU:C:2022:124, para 74.

[54] Article 3(1)(a) Digital Omnibus Proposal, COM(2025) 837 final; Case C-175/20 Valstsieņēmumu dienests (n 5) para 74.

[55] Recital 50 GDPR.

[56] E Curry, S Scerri and T Tuikka (eds), Data Spaces: Design, Deployment and FutureDirections (Springer Nature 2022) 317, noting the immaturity of governance toolingfor documenting purpose and provenance across data-space participants.

[57] Articles 66 and 71, Regulation (EU) 2025/327 (EHDS).

demands.[58] The golden mean, in other words, lies not in loosening the principles but in institutionalising the necessity test through dedicated intermediary governance a solution that is transferable to CEDS more generally only to the extent that comparable access-body architecture exists, which is far from universal across the sectoral data spaces currently being rolled out.[59] Where such institutional intermediation is absent, as in many B2B and B2G data-sharing scenarios envisaged by the Data Act, goal compatibility remains aspirational rather than demonstrated.

3.2 Legal interoperability

The Digital Rulebook is formally interoperable with the GDPR: the Data Act explicitly subordinates itself to data protection law, providing that in the event of conflict the GDPR prevails, and stating that it is "without prejudice" to the powers of supervisory authorities and the rights of data subjects.[60]

The DGA and the EHDS adopt structurally similar deference clauses.[61] This formal hierarchy, however, resolves only the question of priority, not the question of content: none of these instruments defines what "necessary" or "compatible" means for the specific processing operations they regulate, leaving controllers, data intermediaries and national authorities to reconstruct GDPR-compliant purposes for sharing obligations that were drafted primarily from an internal-market, not a fundamental-rights, perspective.[62] This produces two distinct interoperability failures. The first is temporal: the GDPR's principles are technology-neutral and apply ex post to any processing operation, whereas the Data Act and sectoral data-space legislation are drafted around specific data flows (connected-product data, health data, mobility data) and amended piecemeal, meaning that the "static" constitutional layer and the "dynamic" sectoral layer inevitably drift out of alignment as new data spaces are rolled out under Article 33 of the Data Act and equivalent sectoral bases.[63] The second is conceptual: purpose fragmentation is structurally invited by an architecture in which distinct legal instruments authorise distinct categories of onward sharing (B2B, B2G, research reuse) without a shared operational definition of "purpose," so that data legitimately collected under one purpose-specific gateway may be aggregated across gateways in ways that are individually defensible but collectively indistinguishable from the "generalised and indiscriminate" collection the CJEU condemned in Schrems v Meta Platforms Ireland.[64]

The Digital Omnibus Proposal, intended to simplify this landscape, arguably compounds rather than resolves the second failure: by relaxing the identifiability threshold through a controller-relative test, it lowers the practical trigger for GDPR applicability precisely at the moment when data-space architectures are multiplying the number of holders and processing contexts through which the same dataset may travel.[65] Legal interoperability, on this reading, is achieved at the level of formal supremacy but not at the level of substantive coherence a distinction the existing literature on data-space governance has not yet fully drawn out.[66]

3.3 Normative versus practical conflict

The foregoing analysis suggests that the tension is predominantly practical rather than normative, but not exclusively so. It is practical in the sense that nothing in the wording of Article 5(1)(b) and (c) GDPR forecloses large-scale data sharing as such: both principles are purpose-relative rather than volume-based, so that a data space could, in principle, process very large quantities of data lawfully provided the purposes are sufficiently specific and each processing step is demonstrably necessary to them.[67] Biega and Finck's techno-legal analysis supports this reading empirically, showing that data-driven systems typically process substantially more data than their stated purposes require, which locates the problem in implementation design rather than in an inherent

---

[58] European Data Protection Board, Guidelines 4/2019 on Article 25 Data Protection by Design and by Default (adopted 20 October 2020) paras 73–76.

[59] European Commission, overview of the rollout of Common European Data Spaces,https://digital-strategy.ec.europa.eu/en/policies/data-spaces accessed 3 July 2026.

[60] Article 1(5) Regulation (EU) 2023/2854 (Data Act).

[61] Regulation (EU) 2022/868 (DGA); Regulation (EU) 2025/327 (EHDS) art 1.

[62] Recital 4, Regulation (EU) 2023/2854 (Data Act).

[63] Article 33, Regulation (EU) 2023/2854 (Data Act) (interoperability requirements fordata spaces).

[64] Case C-446/21 Schrems v Meta Platforms Ireland Ltd [2024] ECLI:EU:C:2024:834, para 59.

[65] Article 3(1)(a) Digital Omnibus Proposal, COM(2025) 837 final.

[66] S. Ianese, 'Challenging Data Protection in the Age of Data Spaces' (2024) Special Issue European Journal of Privacy Law & Technologies

[67] Article 5(1)(b)–(c) GDPR; Article 29 Data Protection Working Party, Opinion 03/2013on Purpose Limitation (WP203, 2 April 2013) 11–12.

incompatibility between minimisation and scale.[68] Purpose fragmentation, opaque necessity assessments and the absence of harmonised technical standards for expressing minimisation constraints in metadata are, on this view, governance deficiencies capable of correction without touching the substance of the principles.[69]

Yet a residual normative element persists and should not be understated. The CJEU's Grand Chamber in Meta Platforms and Others v Bundeskartellamt, and subsequently the Fourth Chamber in Schrems, have both treated data minimisation as foreclosing indiscriminate aggregation "without restriction as to time or type of data" even where the controller can point to a plausible commercial rationale for broader processing.[70] CEDS, by design, aim precisely at maximising the discoverability and combinability of datasets across sectors; the more successful a data space is in economic terms, the more it resembles the aggregation pattern the Court has found unlawful.[71] There is, therefore, an irreducible ceiling that no amount of governance refinement can remove: purpose limitation and data minimisation impose a normative limit on aggregation as such, not merely on careless aggregation, and CEDS architectures that treat maximal interoperability as an end in itself will collide with that limit regardless of procedural sophistication. **[24]** The tension is thus best characterised as predominantly practical, bounded by a genuine but narrower normative constraint operating at the outer edge of data-space ambition. **[24]**

3.4 Change the law or change the interpretation?

Given this diagnosis, wholesale amendment of Article 5(1)(b) and (c) GDPR is neither necessary nor desirable: the principles derive their content from Article 8 of the Charter, and their dilution would raise fundamental-rights concerns disproportionate to the practical problems identified above.[72]

The more defensible path lies in three complementary interventions. First, amendments to the Digital Rulebook itself - rather than the GDPR - should introduce shared, cross-instrument definitions of "purpose" and standardised, machine-readable necessity metadata, addressing the fragmentation problem identified in the EHDS secondary-use literature without touching GDPR substance.[73] Second, further EDPB guidance specifically addressing data-space architectures, analogous to Guidelines 4/2019 on Article 25, would reduce the current reliance on generic secondary-use commentary and give controllers a workable methodology for demonstrating necessity across multi-actor sharing chains.[74] Third, the CJEU's own case law, most recently in Schrems, already supplies the interpretive tools needed: its insistence on temporal and categorical limitation as constitutive of data minimisation, rather than optional refinement, should be read as directly applicable to data-space intermediaries and access bodies, not only to social-media advertising.[75]

Scholarly proposals to relax the principles through a more permissive "compatible use" test for data-space participants are, on this analysis, difficult to reconcile with the Charter-based reading the Court has endorsed, and should be resisted in favour of governance-level solutions that operationalise, rather than dilute, the existing principles.[76]

A further avenue lies with future preliminary references. Neither Meta Platforms and Others nor Schrems concerned a data space in the CEDS sense; both arose from bilateral platform-user relationships rather than multi-

---

[68] A. J. Biega and M. Finck, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' (2021) Technology and Regulation 44.

[69] ibid; R. Becker and D. Chokoshvili and others, 'Purpose Definition as a Crucial Step forDetermining the Legal Basis under the GDPR' (2024) Journal of Law and the Biosciences.

[70] Case C-252/21 Meta Platforms and Others (Bundeskartellamt) [2023] ECLI:EU:C:2023:537; Case C-446/21 Schrems v Meta Platforms Ireland Ltd [2024] ECLI:EU:C:2024:834, para 59.

[71] Case C-446/21 Schrems v Meta Platforms Ireland Ltd [2024] ECLI:EU:C:2024:834, paras 59–60.

[72] Article 8, Charter of Fundamental Rights of the European Union [2012] OJ C326/391.

[73] Frontiers, 'Secondary use under the European Health Data Space: setting the sceneand towards a research agenda on privacy-enhancing technologies' (2025), reportingstakeholder concerns that existing metadata standards cannot express minimisation constraints.

[74] European Data Protection Board, Guidelines 4/2019 on Article 25 Data Protection by Design and by Default (adopted 20 October 2020).

[75] Case C-446/21 Schrems v Meta Platforms Ireland Ltd [2024] ECLI:EU:C:2024:834, paras 59–60.

[76] M. Fierens, 'Initiating Interdisciplinary Research for Future-Proof Data Protection inthe Context of Data Spaces and Semantic Interoperable Data Sharing' (2024) NXDG2024, CEUR Workshop Proceedings 3891.

actor, intermediated sharing architectures. The Court has therefore not yet had occasion to clarify how the "generalised and indiscriminate" collection standard applies where data is pooled through a governance intermediary - such as a Health Data Access Body - that itself performs a documented necessity assessment before onward disclosure.[77] This is a materially different factual configuration from unrestricted platform-side aggregation, and it is plausible that the Court would apply a more permissive standard where an independent, purpose-bound gatekeeper intervenes between initial collection and secondary use.

Pending such clarification, the most defensible position for data-space architects is to treat the existing case law as establishing a presumption against unbounded aggregation, rebuttable only through the kind of institutionalised, auditable necessity gatekeeping the EHDS model - imperfect and not yet operational as it is - begins to supply. The GDPR need not be amended for this presumption to function correctly; what is required is that the Digital Rulebook be amended to make such gatekeeping structurally mandatory across all sectoral data spaces, rather than an EHDS-specific innovation.[78]

4 Technological approaches to reconciling data sharing and protection:

4.1 Privacy by design in data ecosystems

The aspiration to embed data protection into the architecture of CEDS rather than to append it as an afterthought rests, at its legal foundation, upon Article 25 of the GDPR, which mandates both data protection by design and by default;[79] Recital 78 elaborates that controllers must implement appropriate technical and organisational measures to give effect, in particular, to the principles of data minimisation and purpose limitation.[80] The question, however, is not whether Article 25 applies to CEDS but whether Privacy by Design (PbD) as a legal obligation and an engineering methodology can genuinely reconcile large-scale, multi-actor, cross-sectoral data sharing with those principles, or whether it merely provides a compliance vocabulary that obscures their residual tension.

The concept of PbD, associated with the foundational work of Ann Cavoukian, pre-dates the GDPR and operates on seven foundational principles, including proactive rather than reactive protection and privacy embedded into design.[81] The GDPR translated this doctrinal aspiration into a binding obligation, though it did so in notably open-textured language. Article 25(1) requires measures that are appropriate in light of the state of the art, the cost of implementation, the nature, scope, context and purposes of processing, and the risks to the rights and freedoms of natural persons. This proportionality framing affords controllers substantial discretion, which is simultaneously PbD's strength and its weakness: it renders judicial and regulatory enforcement exceptionally difficult. The EDPB acknowledged in its Guidelines 4/2019 on Article 25 that the provision does not require specific technical solutions but demands genuine effectiveness.[82]

Applied to CEDS, PbD faces structural challenges that the EDPB's framework does not fully resolve. CEDS are, by their legislative design under the DGA and the Data Act, intended to facilitate broad and, in principle, repeated data re-use across multiple purposes and actors. This is architecturally in tension with purpose limitation, which under Article 5(1)(b) GDPR prohibits further processing incompatible with the original purpose. PbD can embed technical access controls to enforce purpose at the point of data sharing, but it cannot cure the legal incompatibility between the *ex ante* specification of purpose required by Article 5(1)(b) and the *ex post* discovery logic inherent in many CEDS use cases, particularly in health and research data spaces governed by the European Health Data Space Regulation.[83]

A more productive framing is to treat PbD not as a mechanism capable of independently resolving the legal tension but as a necessary precondition for its mitigation. Three design properties are particularly salient for CEDS.

---

[77] Articles 65–66, Regulation (EU) 2025/327 (EHDS), establishing Health Data AccessBodies as prior-authorisation gatekeepers for secondary use.

[78] European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (adopted 20 October 2020, version 2.0), paras 73–76.

[79] Article 25 GDPR.

[80] Recital 78 GDPR.

[81] Ann Cavoukian, 'Privacy by Design: The 7 Foundational Principles' (Information and Privacy Commissioner of Ontario, 2009), https://www.ipc.on.ca/sites/default/files/legacy/2018/01/pbd-1.pdf.

[82] European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (adopted 20 October 2020, version 2.0) para 7.

[83] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space COM(2022) 197 final; Article 5(1) GDPR.

First, data space architectures should instantiate data minimisation structurally, by ensuring that data flows by default carry the minimum attributes necessary for any given downstream purpose, rather than transmitting complete datasets from which recipients then extract what they need. This "minimum viable dataset" principle would give Article 5(1)(c) GDPR material operational effect in the data space context. Second, purpose-specific data access channels, technically enforced through policies that reflect specific processing purposes, can operationalise purpose limitation as a real-time constraint rather than merely a contractual obligation in data space agreements. Third, and most critically, PbD must interact with audit mechanisms to remain legally meaningful: a system designed for data minimisation that cannot demonstrate compliance *ex post* satisfies the letter of Article 25 but not its function.

The EDPS has observed that technical safeguards embedded at design stage are not a substitute for legal basis requirements under Articles 6 and 9 GDPR, and that the choice of PbD measures must itself be documented under accountability obligations in Article 5(2).[84] This disciplined reading is important: it prevents PbD from being invoked as a generalised compliance shield in CEDS without addressing the underlying lawfulness question. The CJEU's expansive reading of personal data in *Breyer* and *Fashion ID* further complicates the position, because PbD measures designed for nominally anonymised data spaces may be legally insufficient if participants can reasonably re-identify individuals by combining available information.[85]

The provisional conclusion on PbD is therefore cautiously affirmative but structurally limited: embedded into CEDS architecture from inception, Article 25-compliant design can reduce the tension between data sharing and data protection, but it cannot dissolve it. PbD is best understood as creating the necessary technical preconditions within which more specific legal and technological solutions must operate.

4.2 Privacy-enhancing technologies

Privacy-Enhancing Technologies (PETs) are not a homogenous category. They encompass a spectrum of technical approaches with different functional properties, maturity levels, and relationships to GDPR compliance. Rather than cataloguing them, this section offers a comparative analysis directed at the central research question: which PETs most effectively address purpose limitation and data minimisation in the CEDS context, and do they genuinely reconcile data sharing with GDPR compliance or merely manage residual legal risk?

Pseudonymisation, defined in Article 4(5) GDPR, is the most legally integrated PET: its application is expressly noted in Article 25(1) as an example of a PbD measure and in Article 89(1) as a safeguard for processing for archiving, research and statistical purposes. Its limitation, however, is fundamental. The GDPR and the EDPB make clear that pseudonymised data remains personal data when the key is available to the controller or accessible to the recipient.[86] In a CEDS with multiple participants, the practical likelihood that pseudonymous data can be linked to identities — especially through combination with auxiliary datasets — is high, as the CJEU confirmed in *Breyer*.[87] Pseudonymisation therefore reduces risk and may satisfy data minimisation in limited architectures, but it does not resolve purpose limitation: a pseudonymous dataset shared for one research purpose remains constrained by Article 5(1)(b) if subsequently processed for a different purpose.

Anonymisation, if genuine, would remove data from the GDPR's scope entirely pursuant to Recital 26.[88] The EDPB's predecessor, the Article 29 Working Party, set a high bar in Opinion 05/2014 on Anonymisation Techniques, requiring that re-identification be "reasonably likely" to be impossible, considering both the means available to the controller and to likely third parties.[89] In practice, genuine anonymisation is exceptionally difficult in large, rich datasets characteristic of CEDS. Research in the field of differential privacy has demonstrated that even ostensibly anonymised datasets including mobility data, medical records and financial transaction logs, remain

[84] European Data Protection Supervisor, 'Opinion 3/2020 on the European Strategy for Data' (16 June 2020) para 40.
[85] Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland EU:C:2016:779; Case C-40/17 Fashion ID GmbH & Co KG v Verbraucherzentrale NRW eV EU:C:2019:629
[86] Recital 26 and Article 4(5) GDPR; European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (adopted 20 October 2020, version 2.0) para 31.
[87] Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland EU:C:2016:779, paras 45-49.
[88] Recital 26 GDPR.
[89] Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (WP216, 10 April 2014) 3–4.

vulnerable to re-identification through linkage and inference attacks.[90] The consequence is that anonymisation, while theoretically the most powerful GDPR compliance tool, is operationally unreliable in CEDS and cannot be routinely invoked as a basis for removing data protection obligations.

Federated learning and secure multiparty computation (SMPC) represent a qualitatively different approach: rather than sharing data, they share computation. In federated learning, model training occurs locally on distributed datasets, and only model updates — not raw data — are shared with a central aggregator.[91] SMPC allows multiple parties to jointly compute a function over their combined inputs without any party learning the others' inputs.[92] Both technologies directly operationalise data minimisation — because raw data is not shared — and are structurally better suited to purpose limitation than pseudonymisation, because the computation itself can be purpose-specific. ENISA has identified federated learning as one of the most promising PETs for health and industrial data spaces precisely because it enables collaborative analytics without centralising sensitive datasets.[93]

However, both technologies retain limitations that prevent them from fully resolving the legal tension. Federated learning remains susceptible to model inversion and gradient inference attacks through which an adversary can reconstruct training data from shared model parameters.[94] SMPC protocols, while cryptographically robust in principle, impose significant computational overhead that renders them impractical for many real-time CEDS use cases and for datasets of the scale envisaged by the European Health Data Space or the Mobility Data Space. Furthermore, neither technology addresses the question of legal basis: even if no raw personal data crosses organisational boundaries, the underlying local model training on personal data still requires a valid legal ground under Article 6 or Article 9 GDPR.

Differential privacy (DP) merits separate attention because it provides a mathematically defined privacy guarantee: it ensures that the output of a query or computation is statistically indistinguishable regardless of whether any particular individual's data is included.[95] This property directly supports data minimisation by bounding the information leaked about any individual. The EDPS Technical Report on Differential Privacy for European Statistical Agencies acknowledged its potential but cautioned that the calibration of the privacy budget (epsilon) involves a fundamental trade-off between utility and privacy protection that must be determined in context and documented under accountability obligations.[96] At high utility requirements epsilon values sufficient to preserve usefulness may provide weaker privacy guarantees than the standard implies. Differential privacy is therefore a genuine, mathematically grounded contribution to data minimisation, but it functions along a utility-privacy continuum rather than providing a binary compliance answer.

Secure processing environments (SPEs) represent a different architectural logic: data is brought to a controlled computation environment rather than distributed to participants. Data clean rooms, in particular, have gained significant traction in the digital advertising and health sectors, enabling data holders to allow third-party analysis without exposing underlying datasets.[97] They support purpose limitation by restricting what computations can be performed, and support data minimisation by preventing direct data export. Their legal status under the GDPR remains somewhat undertheorised, however: depending on architecture, a clean room operator may qualify as a controller, processor or joint controller, with different implications for data subject rights under Articles 15-22 GDPR.

Assessing these technologies against the two key GDPR principles, a provisional ranking for the CEDS context emerges. Federated learning, complemented by differential privacy, ranks most highly because it structurally minimises data sharing (supporting data minimisation), can be designed for purpose-specific computations (supporting purpose limitation), and is technically mature enough for deployment at scale in health and industrial data spaces,

---

[90] L. Rocher, J. M. Hendrickx and Y.-A. de Montjoye, 'Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models' (2019) 10 Nature Communications 3069.

[91] B. McMahan and others, 'Communication-Efficient Learning of Deep Networks from Decentralized Data' (2017) Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 1273.

[92] R. Cramer, Ivan Damgård and Jesper Buus Nielsen, *Secure Multiparty Computation and Secret* Sharing (Cambridge University Press 2015).

[93] European Union Agency for Cybersecurity (ENISA), 'Data Protection Engineering' (ENISA, January 2022) 58–63.

[94] L. Melis and others, 'Exploiting Unintended Feature Leakage in Collaborative Learning' (2019 IEEE Symposium on Security and Privacy) 691.

[95] C. Dwork and A. Roth, 'The Algorithmic Foundations of Differential Privacy' (2014) 9(3–4) Foundations and Trends in Theoretical Computer Science 211, 211–12.

[96] European Data Protection Supervisor, 'Preliminary Opinion 8/2020 on the European Statistical Programme 2021-2027' (21 August 2020) para 19; see also EDPS TechDispatch No 1/2021 'Differential Privacy'.

[97] European Union Agency for Cybersecurity (ENISA), 'Data Protection Engineering' (ENISA, January 2022), 71-73.

as demonstrated by EU-funded pilots such as GAIA-X and the TEHDAS project.[98] Secure processing environments with data clean room functionality rank second, because they provide strong governance controls and are already commercially deployed, though their legal classification requires careful structuring. Pseudonymisation, despite its GDPR recognition, ranks below these alternatives for CEDS because its effectiveness is contingent on re-identification risk assessments that are increasingly difficult to sustain in data-rich environments.

The overarching observation is that no single PET fully reconciles data sharing with GDPR compliance. PETs reduce residual legal risk and, in the best cases, operationalise data minimisation and purpose limitation as technical constraints. They do not, however, displace the need for legal basis analysis, data protection impact assessments (DPIAs) under Article 35, or the governance arrangements that must underpin CEDS under the DGA. Technology and law must function in combination: PETs create the technical architecture within which GDPR compliance can be achieved, but they are not a substitute for the legal obligations that remain.

5 Conclusion

This paper set out to answer the following research question: how could the GDPR principles of data minimisation and purpose limitation be ensured and operationalised within the data-sharing framework of Common European Data Spaces? To that end, the paper first outlined the legal and technical design of CEDS and the content of the purpose limitation and data minimisation principles, before turning to the reshaped definition of personal data following the "SRB v EDPS" judgment and the Digital Omnibus Proposal, which reinforces ra-ther than diminishes the importance of privacy-enhancing technologies. The paper then assessed the legal ten-sion between CEDS and the GDPR along four dimensions — goal compatibility, legal interoperability, the normative-versus-practical character of the conflict, and whether the law or its interpretation should change — before evaluating the extent to which Privacy by Design and Privacy-Enhancing Technologies can operational-ise the GDPR principles within CEDS.

The analysis yields a consistent finding: the tension between CEDS and the GDPR is predominantly practical rather than normative, though a narrower normative ceiling on unbounded aggregation persists regardless of governance sophistication. Wholesale amendment of Article 5(1)(b) and (c) GDPR is neither necessary nor desirable; the more defensible path lies in amending the Digital Rulebook to introduce shared definitions of "purpose" and standardised necessity metadata, further EDPB guidance on data-space architectures, and the application of existing CJEU case law to data-space intermediaries. Technology can substantially mitigate, but not resolve, the residual tension: Privacy by Design creates the necessary technical preconditions but cannot cure the legal incompatibility between the ex ante specification of purpose required by Article 5(1)(b) and the ex post discovery logic of many CEDS use cases, while among the Privacy-Enhancing Technologies examined, federated learning combined with differential privacy, purpose-based access control paired with eIDAS 2.0-compliant digital identity infrastructure, and secure processing environments such as data clean rooms offer the greatest potential to narrow the gap, each subject to its own technical or legal limitations.

The critical gap identified throughout this analysis is the absence of a regulatory framework that systematically coordinates the deployment of these technologies within CEDS governance: the DGA establishes data intermediaries and data altruism organisations but does not prescribe technical standards for GDPR-compliant data spaces, the Data Act does not harmonise the PET requirements that should accompany data access obli-gations, and the Digital Omnibus Proposal risks weakening rather than clarifying the GDPR's application to data spaces. What is needed, and what the literature has not yet provided, is a technology-informed legal framework specifying which PETs and identity architectures satisfy Article 25 GDPR obligations in the CEDS context, coordinated with the sectoral rules governing specific data spaces.

Building on this gap, future research could pursue several complementary avenues. First, work could develop the technology-informed legal framework identified above, specifying concretely which combinations of PETs and identity architectures satisfy Article 25 GDPR in CEDS and how such standards should be coordinated across

[98] GAIA-X AISBL, 'Architecture Document' (version 22.10, October 2022); TEHDAS Joint Action, 'Recommendations for Federated Health Data Infrastructure' (April 2022).

the DGA, the Data Act and sectoral instruments such as the European Health Data Space Regulation. Second, once adopted, the Digital Omnibus Proposal's relative approach to identifiability warrants closer as-assessment against the re-identification risks inherent in data-rich, multi-participant CEDS architectures. Third, future preliminary references could clarify how the Court of Justice's "generalised and indiscriminate" collection standard, developed in Meta Platforms and Schrems, applies to data pooled through a governance intermediary that performs a documented necessity assessment — a configuration the Court has not yet addressed. Fourth, further technical research could examine the practical trade-offs — computational overhead, inference-attack vulnerability, and the calibration of the differential-privacy budget — that determine whether federated learning and secure multiparty computation can operate at the scale envisaged by sectoral data spaces. Finally, the legal classification of secure processing environments, including data clean rooms, as controller, processor or joint controller remains undertheorised and merits dedicated doctrinal attention.

References

1. Allen C. (2016) The Path to Self-Sovereign Identity. Life with Alacrity (blog), http://www.lifewithalacrity.com/2016/04/the-path-to-self-sovereign-identity.html.
2. Becker R., Chokoshvili D. et al. (2022) Secondary use of Personal Health Data: when is it 'Further Processing' under the GDPR, and What Are the Implications for Data Controllers?. doi.:10.2139/ssrn.4070716.
3. Becker R., Chokoshvili D. et al. (2024) Purpose Definition as a Crucial Step for Determining the Legal Basis under the GDPR: Implications for Scientific Research. Journal of Law and the Biosciences, doi: 10.1093/jlb/lsae001.
4. Curry E., Scerri S., Tuikka T. (eds) (2022) Data Spaces : Design, Deployment and Future Directions. Springer Nature, doi.:10.1007/978-3-030-98636-0.
5. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (Open Data Directive). OJ L 172, 26 June 2019.
6. Fierens M. (2024) Initiating Interdisciplinary Research for Future-Proof Data Protection in the Context of Data Spaces and Semantic Interoperable Data Sharing. Proceedings of the 1st NeXt-generation Data Governance Workshop 2024 (NXDG 2024) co-located with 20th International Conference on Semantic Systems (SEMANTiCS 2024). In CEUR Workshop Proceedings 3891.
7. Huber M., Wessel S., Brost G., Menz N. (2022) Building Trust in Data Space. In: Otto B. et al. (ed.) Designing Data Spaces: The Ecosystem Approach to Competitive Advantage, Springer, 147–164.
8. Ianese S. (2024) Challenging Data Protection in the Age of Data Spaces. Special Issue European Journal of Privacy Law & Technologies. universitypress.unisob.na.it/ojs/index.php/ejplt/article/view/2002.
9. Steinbuss S. (2025) Observability in Data Spaces (International Data Spaces Association, June 2025). doi:10.5281/zenodo.15647484.
10. Article 29 Data Protection Working Party (2013) Opinion 03/2013 on Purpose Limitation (WP203). Adopted 2 April 2013.
11. Big Data Value Association (BDVA) (2019) Towards a European Data Sharing Space. Position Paper. BDVA, April 2019.
12. Court of Justice of the European Union (2023) Case T-557/20, Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS), ECLI:EU:T:2023:219. Judgment of 26 April 2023.
13. Data Spaces Support Centre (DSSC) (2025) Data Spaces Blueprint, Version 3.0. DSSC. https://blueprint.dssc.eu.
14. European Committee for Standardization (CEN) (2024) CEN Workshop Agreement CWA 18125:2024 (E). July 2024.
15. European Commission (2020) A European Strategy for Data. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 19 February 2020, COM(2020) 66 final.
16. European Commission (2025) Data Union Strategy: Unlocking Data for AI. Communication from the Commission to the European Parliament and the Council, 19 November 2025, COM(2025) 835 final.
17. European Data Protection Board (EDPB) (2020) Guidelines 4/2019 on Article 25 Data Protection by Design and by Default. Adopted 20 October 2020.
18. European Union Agency for Cybersecurity (ENISA) (2022) Digital Identity: Leveraging the Self-Sovereign Identity (SSI) Concept to Build Trust. ENISA, Heraklion.
19. Information Commissioner's Office (ICO) (2023) Privacy Enhancing Technologies. ICO Guidance and Resources. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/privacy-enhancing-technologies/.
20. International Data Spaces Association (IDSA) (2022) IDS Reference Architecture Model 4.0. IDSA.
21. World Wide Web Consortium (W3C) (2022) Decentralized Identifiers (DIDs) v1.0: Core Architecture, Data Model, and Representations. W3C Recommendation, 19 July 2022. https://www.w3.org/TR/did-core/.
22. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). OJ L 119, 4 May 2016 (GDPR).
23. Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union. OJ L 303, 28 November 2018.
24. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance (Data Governance Act). OJ L 152, 3 June 2022 (DGA).
25. 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 (Digital Markets Act). OJ L 265, 12 October 2022.
26. Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data (Data Act). OJ L, 22 December 2023.

27. Regulation (EU) 2024/903 of the European Parliament and of the Council of 13 March 2024 laying down measures for a high level of public sector interoperability across the Union (Interoperable Europe Act). OJ L, 22 March 2024.
28. Regulation (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework. OJ L, 30 April 2024 (eIDAS 2.0).
29. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act). OJ L, 12 July 2024.
30. Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. OJ L, 5 March 2025.