Ianese S, 'Challenging Data protection in the age of Data Spaces' (2024) Special Issue EJPLT.
DOI:

EJPLT
ISSN 2704-8012



Challenging Data Protection in the age of Data Spaces

La sfida della protezione dei dati personali nell'era dei Data Spaces

SVEVA IANESE
Lawyer
Ph.D.(c) in Economics and Management
University of Padova

Abstract

Data spaces are new digital ecosystems introduced to support rapid and secure data sharing within the EU. Data protection legislation poses new challenges to such projects, as it is still unclear how its principles should be translated into practice (e.g. privacy roles and responsibilities of the Data Space actors, accountability of the data controllers, risk management, and so on). The purpose of this paper is to analyse how data protection principles and duties are shaped in the context of Data Spaces to establish whether, and eventually how, these rules are challenged by that collaborative ecosystem.

Gli Spazi Dati sono stati introdotti per favorire la condivisione sicura e lecita delle informazioni all'interno dell'UE. La protezione dei dati personali pone sfide nuove nella realizzazione di tali progetti, non essendo tuttora chiaro come tali principi debbano essere declinati nella pratica (es. ruoli e responsabilità degli attori del Data Space, risk management, ecc.). Il presente articolo si propone di verificare come la normativa privacy sia declinata nell'ambito dei Data Spaces al fine di stabilire se, ed eventualmente come, tali regole necessitino di un ripensamento.

© The author(s) 2024, published by Suor Orsola Benincasa Università Editrice.
This contribution is licensed under a Creative Common Attribution 4.0 International Licence CC-BY-NC-ND, all the details on the license are available at:
https://creativecommons.org/licenses/by-nc-nd/4.0/

Keywords: data space, data protection, privacy, GDPR

Summary: 1. Introduction. – 2. Privacy roles within a Data Space. – 3. Stressing data protection in Data Spaces. – 3.1. Processing lawfulness. – 3.2. Privacy by design and data synthesization. – 3.3. A holistic DPIA for Data Spaces. – 4. Conclusions.

1. Introduction.

Today a new perspective in approaching data protection issues has to be adopted. The setting proposed so far has been characterized by a compartmentalized and sectoral logic, in which the data sources were independent and centralized (data silos). This resulted in a duplication of available information and the creation of a set of rules based on one-to-one relationships between data subjects and controllers. All the actors involved in processing operations were clearly identified *ex ante* and did not change over time.

In recent years a revolution occurred, based on a collaborative perspective. The development of Data Spaces and the possibility of reusing information for multiple purposes represents the ultimate goal to achieve. This collaborative approach is one of the main pillars of the European data strategy, which aims to create the legal and technological conditions for creating a single data market, improving trust between actors and the fairness of sharing practices. Data Spaces are one of the models chosen by the European Union to promote the achievement of these goals[1] [2].

'Data Space' is often used as an umbrella term and the literature has elaborated various definitions for it[3]. In its most simplified form, a Data Space is a common infrastructure and governance framework to facilitate data sharing between multiple entities, both public and private[4]. A Data Space represents not only a technical framework[5] but also a set of participants involved in sharing activities and the inter-relations between them[6] [7]. The Data Space is therefore a *hard* and *soft* infrastructure that enables transactions involving personal and non-personal data between the members of a digital ecosystem based on a common governance setting, in a secure manner and on a fiduciary basis.

Despite the innovativeness of this model, it is still regulated by a broad set of 'traditional' principles and rules represented by European values, rights and laws[8]. The most relevant legislation in this context is the data protection Regulation.

---

[1] 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' [2020] <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066> - accessed 28 July 2024.

[2] Statement of the Members of the European Council [2021] 6c). <https://www.consilium.europa.eu/media/48976/250321-vtc-euco-statement-en.pdf> - accessed 28 July 2024.

[3] E Curry et al. (eds.), *Data Spaces: Design, Deployment, and Future Directions* (Springer, 2022) 3. <https://doi.org/10.1007/978-3-030-98636-0_1> - accessed 28 July 2024.

[4] Staff Working Document on Common European Data Spaces [2022]. <https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces> - accessed 28 July 2024.

[5] L Nagel, D Lycklama, *Design Principles for Data Spaces. Position Paper*. (IDSA 2021) 13. <http://doi.org/10.5281/zenodo.5105744> - accessed 28 July 2024.

[6] M Franklin, A Halevy, D Maier, 'From databases to dataspaces: a new abstraction for information management' (2004) 34(4) ACM SIGMOD Record, 27. <https://doi.org/10.1145/1107499.1107502> accessed 28 July 2024.

[7] M Singh, SK Jain 'A Survey on Dataspace' in DC Wyld, M Wozniak, N Chaki, N Meghanathan, D Nagamalai 'Advances in Network Security and Applications. CNSA 2011.' (Springer, 2011) <https://doi.org/10.1007/978-3-642-22540-6_59> accessed 28 July 2024.

[8] A European strategy for data [2020] Sec I.

However, the development of a Data Space - which is based on collaborative logic - poses new challenges to personal data protection, which implies a unidirectional transaction model (1:1) instead. Such a collaborative ecosystem asks for rethinking the foundations of privacy regulation and conceptualizing it in a different manner.

This paper aims to map the main challenges posed by the application of privacy legislation to the Data Space framework. The objective is to establish whether this Regulation should be partially redesigned in order to promote the creation of new ecosystems based on the purpose of information sharing and reuse.

The article is structured as follows. In section 2 the privacy roles of the main actors of a Data Space are listed. Section 3 addresses the major challenges posed to data controllers by the accountability principle. In each subsection the topic is explored by considering the processing lawfulness, the privacy by design principle and the risks management duty. Finally, Section 4 includes some conclusions and final remarks.

2. Privacy roles within a Data Space.

In any Data Space various actors are brought together[9] [10] and take on different roles[11].

The first element of complexity concerning a Data Space regards the stakeholders involved in it (*static approach*). There is no minimum number of participants or a rigid organizational structure that must be respected to develop such projects. Furthermore, its network structure and the number of actors participating in it may change enormously over time. The structure of a Data Space varies hugely from case to case consequently and participants take on different roles at the same time, depending on the data transactions in which they are involved.

A second complexity element concerns the whole setting that a Data Space has (*dynamic approach*). The fluid structure of these ecosystems can favor multidirectional interaction between the participants. Different scenarios come out. The relationship between the participants in the Data Space might be bilateral ("1:1"), multilateral and/or consortial ("n:1" and "1:n")[12], or mediated by the existence of a Data Intermediary[13] (Fig. 1).

[9] M Micheli, M Ponti, M Craglia, A Berti Suman 'Emerging models of data governance in the age of datafication' (2020) 7(2) Big Data & Society <https://doi.org/10.1177/2053951720948087> accessed 28 July 2024.
[10] L Nagel, D Lycklama, *Design Principles for Data Spaces.* (n 5) 84.
[11] Data Spaces Support Centre, Glossary, (2nd, DSSC 2023), Sec 3. <https://dssc.eu/space/BVE/357073697/Introduction> accessed 28 July 2024.
[12] A Duisberg 'Legal Aspects of IDS: Data Sovereignty—What Does It Imply?' in B Otto, M ten Hompel, S Wrobel (eds.) *IDS Reference Architecture Model* (Springer, 2022), 75 <https://doi.org/10.1007/978-3-030-93975-5> accessed 2 August 2024.
[13] See Section 2§5.

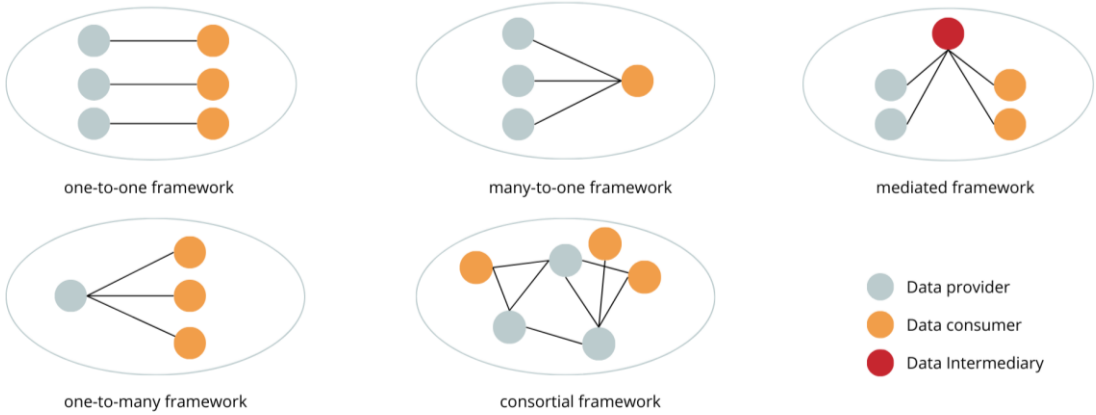


Figure 1 - Eventual interactions between the actors of a Data Space

Therefore an important aspect that must be considered in order to guarantee the privacy compliance of a Data Space concerns its participants - it is necessary to establish which stakeholder takes which roles (controller, processor, and joint controller) in order to assess the related obligations. In the following sections the roles and responsibilities of the main players in a Data Space are described, according to privacy law. Please note that this breakdown is illustrative and refers to a single data transaction, while every subject can play multiple roles at the same time (e.g. Data Holder in one transaction and Data User in the following one).

- Data Holder

The Data Holder is a natural or legal person who technically provides data or grants access to it[14]. In addition, it can perform additional functions including performing data communications and dissemination, recording and storage of personal data, and data transformation. Its main responsibility is to thoroughly describe licensing and conditions for use of data assets[15]. The Data Holder may adopt numerous technological means to exchange information with Data Users (e.g., software), to modify data in format or semantic, or to improve information quality (e.g., data apps)[16].

Due to these activities, the Data Holder usually plays the role of 'data controller' pursuant to the GDPR, being the person who determines the purposes and means of data processing[17]. Exceptionally, the Data Holder acts as a 'data processor' when he/she is asked to follow the hints given by third parties in data proceeding. In these cases, the parties have to sign a Data Processing Agreement[18] for setting out mutual responsibilities.

Finally, it may happen that Data Holders jointly provide, modify or grant access to personal data - e.g., through formal or *de facto* consortia. In these hypotheses all the

[14] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space [2022] art 2.2(y) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197> accessed 29 July 2024.

[15] E Farrell, M Minghini, A Kotsev et al. *European Data Spaces: Scientific insights into data sharing and utilisation at scale* (Publications Office of the European Union, 2023) 71 <https://doi.org/10.2760/400188, JRC129900> accessed 2 August 2024.

[16] B Otto, S Lohmann, S Steinbuss, A Teuscher, Andreas. *IDS Reference Architecture Model* (1st, IDSA, 2018), 18.

[17] Regulation 679/2016 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, and repealing Directive 95/46/EC (General Data Protection Regulation)[2016] OJ L 119/1 art 4(7) <https://eur-lex.europa.eu/eli/reg/2016/679/oj> accessed 29 July 2024.

[18] General Data Protection Regulation [2016] art 28.

subjects have to jointly determine the purposes and means of the processing and stipulate a joint controllership agreement with each other[19].

- Data User

A Data User can be defined as a natural or legal person which has been granted lawful access to data within a specific data transaction[20]. The Data User generally acts as a 'further data controller' in addition to the Data Holder itself. The digital transaction between the Data Holder and the Data User may be qualified as a controller-to-controller data transfer consequently. Usually the Data User does not operate as a 'data processor'.

Data Users could be involved in a consortium too. In this case, they should enter into a joint controllership agreement to define mutual responsibilities and jointly establish the purposes and means of the processing.

- Data Subject

The Data Subject is a identified or identifiable natural person whose personal data are intended to be processed[21]. This party can be a participant of the Data Space or not. A Data Space has to provide Data Subjects the tools to exert their rights, as set out in the GDPR (e.g., the right to decide what data collectors are allowed to do with their data), regardless of their involvement in the ecosystem.

Data Subjects can always demand transparency as to where their data is stored and who could access it. They can revoke their consent, modify access preferences and set out new conditions of how their data can be obtained or used. Furthermore, they can choose to outsource data rights management to third parties (e.g., Data Intermediaries).

- Data Space Orchestrator

The Data Space Orchestrator is a participant in the Data Space that sets out business rules and other conditions to be followed by other members. It implements the technical, legal and organizational setting that enable data transactions. It may act as a digital solution provider or might ensure internal coordination of the digital ecosystem by creating centralized databases and catalogs, keeping track of data sources and processings, updating internal policies and so on[22] [23].

The Orchestrator qualifies as a 'data processor' when it carries out processing activities on behalf of a third party. For example, it occurs when the Orchestrator develops an application used by the Data Holder to allow the User access to data or when it offers data cleansing services. In these hypotheses, the Orchestrator and the data controller must sign a Data Processing Agreement pursuant to Art. 28 GDPR. Sometimes the Orchestrator acts as a 'data controller' as when it anonymizes personal data for its own purposes or transforms its format, determining the methods and purposes of such processings in that way[24].

---

[19] General Data Protection Regulation [2016] art 26.

[20] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space [2022] art 2.2(z)

[21] General Data Protection Regulation [2016] art 4(1).

[22] Spanish Data Protection Agency (AEPD) 'Approach to Data Spaces from GDPR perspective' [2023] <https://www.aepd.es/documento/approach-to-data-spaces-from-gdpr-perspective.pdf> accessed 1 August 2024.

[23] One example is MyHealth@EU, the national contact point for the secondary use of electronic health data under the European regulation on the health Data Space proposal. See ibid art 12.

[24] AEPD 'Approach to Data Spaces from GDPR perspective' [2023] 34.

- Data Intermediary

A Data Intermediary is a trust builder. It enhances commercial relationships for the purposes of data sharing between Data Subjects and/or Holders, on the one hand, and Data Users, on the other[25]. It could provide the technical and legal solutions to do so or enhance the execution of privacy rights from Data Subjects[26]. Its services enable and encourage transactions among the Data Space members but its presence is optional for the development of such ecosystems. In summary, the Data Intermediary plays a fundamental role in protecting personal data within a Data Space by providing centralized consent management solutions, digital wallets for decentralized identity management, and storage of verified credentials[27].

Looking at its role, most of the time a Data Intermediary acts as a 'data processor' as it complies with the instructions provided by a third party to carry out personal data processing activities. For example, it could facilitate personal data management by providing Data Holder with an interface to govern how information is used and by whom within the Data Space[28].

3. Stressing data protection in Data Spaces.

To the extent that a Data Space involves processing of personal data, it is subject to privacy regulations (first of all, the GDPR). This condition occurs very often as processing activities often involves mixed datasets[29] - where personal and non-personal data are inextricably linked[30] - or datasets including personal information. Compliance with privacy Regulation raises many questions covering all the phases of processing activities, from preparation to execution. These challenges correspond to a set of requirements necessary to ensure the development of a lawful and compliant Data Space, defined as *data protection building blocks*.

Each building block is briefly presented in the following subsections.

3.1 Lawfulness of processing.

The first step is represented by the lawfulness of processing operations[31]. The data life cycle within Data Spaces may involve different activities and every

[25] Regulation 868/2022 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) [2022] OJ L 152/1 <https://eur-lex.europa.eu/eli/reg/2022/868/oj> accessed 1 August 2024.

[26] M Micheli, E Farrell, B Carballa Smichowski et al. *Mapping the landscape of data intermediaries* (Publications Office of the European Union, 2023) 11 <https://doi.org/10.2760/261724, JRC133988> accessed 3 August 2024.

[27] E Farrell, M Minghini, A Kotsev et al. *European Data Spaces* ibid 27 ff.

[28] One example is personal information management systems (PIMS), which support users or third parties in establishing business relationships with an indeterminate number of counterparties while guaranteeing users remain in full control of their data. Those services are defined as data intermediation services under the Data Governance Act. See Regulation 2854/2023 of the European Parliament and of the Council of 13 December 2023 on harmonized rules on fair access to and use of data and amending Regulation (EU) 2394/2017 and Directive (EU) 2020/1828 (Data Act) [2023] OJ L Recital 33 <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32023R2854> accessed 3 August 2024.

[29] 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' [2019] 8 ff. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019DC0250> accessed 12 August 2024.

[30] Regulation 1807/2018 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 [2018] OJ L 303/59 Art. 2(2). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807> accessed 12 August 2024.

[31] A processing operation is 'any operation or set of operations which is performed upon personal data or sets of personal data, whether or not by automated means' under Art. 4(2) of the GDPR.

processing needs a specific legal basis to be performed, pursuant to Art. 6 GDPR. As we know, legal basis are the consent, the performance of a contract or the preliminary steps to entering into it, compliance with a legal obligation, the protection of the vital interests of the Data Subject or of another natural person, the performance of a task carried out in the public interest or in the exercise of official authority, the legitimate interests of the controller or a third party. A legal basis must recur for each processing activity and, by virtue of the purpose limitation principle, personal data cannot be further processed in a manner that is incompatible with the original scope[32].

The topic of further proceeding is particularly challenging in the Data Space context. If a data controller has collected information on the basis of legitimate interest, a contract or vital interests, this data can be used for another purpose (but only) after checking that the new purpose is compatible with the original one. In other words, the controller must carry out and be able to demonstrate the positive outcome of a compatibility test[33]. On the other hand, if the controller collected data on the basis of consent or due to a legal requirement, no further processing is possible beyond what is covered by the original consent or the law. Further proceedings would require obtaining new consent or referring to a new legal basis. This final requirement is particularly demanding for the controller, who must (a) verify that the Data Subject has been adequately informed and agreed to such further operation, or (b) if he/she fails, come back to the Data Subject for obtaining a new consent or identify another legal basis.

This aspect is very challenging within the Data Space, where new participants continuously come in, stakeholders are numerous (especially when the project is mature) and the relationships between the Data Space actors evolve very quickly. Here it may be very difficult for controllers to predict the characteristics of further proceedings from the beginning. Moreover the Data Subjects are not always Data Space participants. If the controller needs to ask them for another consent, those people must be detected and their will has to be collected and recorded.

The difficulty of building the necessary conditions to guarantee a lawful further proceeding in practice could limit the power of Data Spaces. As this framework is established precisely for encouraging the reuse of information between the stakeholders, the binding conditions aforementioned may undermine the potential of such collaborative ecosystems.

The Proposal for a Regulation on the European Health Data Space adds another complexity element by introducing the 'secondary use of (electronic health) data'. This concept is defined as 'the processing of (electronic health) data for purposes set out in Chapter IV of this Regulation, other than the initial purposes for which they were collected or produced'[34]. It implies information processing for purposes different from the original ones, like policy making and regulatory activities or development and innovation activities for health products/services[35].

While recalling data protection principles and regulations, the Proposal develops

[32] General Data Protection Regulation [2016], Art. 5.

[33] General Data Protection Regulation [2016], Art. 6(4). The following points should be considered in performing the test:

- the link between the original purpose and the new/upcoming purpose;
- the context in which the data was collected;
- the type and nature of the data;
- the possible consequences of the intended further processing;
- the existence of appropriate safeguards (e.g. anonymization, encryption).

[34] Proposal for a Regulation on the European Health Data Space [2022] Art. 2.2(e).

[35] Proposal for a Regulation on the European Health Data Space [2022] Art. 34.

the concept of 'secondary use of data' alongside the one of 'further proceeding' posed by the GDPR. The document does not define the mutual boundaries of those concepts[36]. While Art. 2.2(e) and 34 of the Proposal aim to encourage the reuse of (health) data for valuable purposes, they generate regulatory uncertainty regarding the conditions necessary to carry out such additional operations. To date, it is not clear how the 'further proceeding' principle applies in these circumstances.

3.2. Privacy by design and data synthesization.

The 'privacy by design' principle is one of the main pillars of the GDPR paradigm[37]. It consists of reducing data processing to a minimum thanks to the adoption of specific technical and organizational measures. It requires data controllers to draw the processing operations during the design phase focusing on the values and requirements set by the Regulation.

It is closely linked to the minimization principle, which asks data controllers to process only the amount of information that is strictly necessary to achieve their goals, without including useless data[38].

These principles must also be applied in the context of Data Spaces but their observance might be tough in collaborative ecosystems. Concerted actions are fundamental for the success of Data Spaces and the strict limitation of circulating information would jeopardize their ultimate purpose. Therefore it is necessary to balance two conflicting needs - limiting the exposure of personal data while making available the minimum information necessary to fulfill the main goal of the Data Space (meaning, information sharing and reuse)[39]. These contradictory needs could be reconciled thanks to Privacy Enhancing Technologies (PETs). PETs constitute a set of technological measures that protect privacy by reducing the amount of processed data or by preventing unnecessary processing operations[40] [41]. PETs vary hugely in their capabilities, functionality, technical design and 'readiness'[42]. However, all of them aim to empower either users and consumers seeking to control the disclosure, use and distribution of their information or businesses and organizations in enforcing their own privacy policies[43].

An important category of PETs concerns 'Data obfuscation tools'[44] which includes data anonymization procedures[45]. Data anonymization consists of a set of models

---

[36] In fact, EDPB EDPS ask for a clarification on that mattern in the Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space. EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space [2022] Sec. 42. <https://www.edpb.europa.eu/system/files/2022-07/edpb_edps_jointopinion_202203_europeanhealthdataspace_en.pdf>

[37] General Data Protection Regulation [2016] art 25.

[38] General Data Protection Regulation [2016] art 5.1(c).

[39] AEPD 'Approach to Data Spaces from GDPR perspective' [2023] 40.

[40] European Data Protection Supervisor (EDPS) *Glossary* (European Commission, 2023). <https://www.edps.europa.eu/data-protection/data-protection/glossary/p_en#pets> accessed 3 August 2024.

[41] UK Information Commissioner's Office *Privacy-enhancing technologies (PETs). Anonymisation, pseudonymisation and privacy enhancing technologies guidance* (ICO, 2022) 5 <https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf> accessed 3 August 2024.

[42] M Hansen, JH Hoepman, M Jensen 'Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies' (ENISA, 2015) <https://www.enisa.europa.eu/publications/pets> accessed 12 August 2024.

[43] Organization for Economic Co-operation and Development (OECD) 'Emerging privacy-enhancing technologies: Current regulatory and policy approaches' (OECD Publishing, 2023). <https://doi.org/10.1787/bf121be4-en> accessed 2 September 2024.

[44] OECD 'Emerging privacy-enhancing technologies: Current regulatory and policy approaches' ibid 15.

[45] JM Auñón, D Hurtado-Ramírez, L Porras-Díaz et al. 'Evaluation and utilization of privacy enhancing technologies—A Data Spaces perspective' (2024) Data in Brief <https://doi.org/10.1016/j.dib.2024.110560> accessed 2 September 2024.

and techniques aimed at protecting personal data through altering, deleting, or encoding identifiers that can directly reveal the individual behind that information[46]. We know that such data falls outside the scope of the GDPR if the initial Data Subject is not or no longer identifiable[47]. However, what penalizes the use of anonymization techniques is their frequent lack of statistical power. In fact, once anonymized, granular data is deprived of some of its insightful properties because those characteristics might reveal the identity of the Data Subject. This aspect is fundamental in the Data Spaces framework, where successful data sharing and reuse ask for the preservation of as many properties of the original dataset as possible. The removal of relevant data properties could hinder an in-depth analysis of such information or the training of artificial intelligence systems. The removal of some dataset's key characteristics due to anonymization risks making such information completely useless within a Data Space, where data is important because of its representativeness.

This problem might be solved through data synthesization. Synthetic data is fictitious information that is created from a 'real' dataset through artificial intelligence techniques. Even though it is unreal, the synthetic dataset maintains the properties and statistical distribution of the real one[48] so that the two samples are indistinguishable from a mathematical perspective[49]. Although synthetic data does not contain 'real' data, it works as an important tool for achieving the main goals of data reuse[50] in the Data Space context. This solution helps to reconcile the two conflicting needs discussed above - the respect of minimization and privacy by design principles, on one hand, and the need to ensure high quality and representativeness of shared information within Data Spaces on the other.

Here privacy rules are shaped by the collaborative logic of Data Spaces in order to ensure the achievement of its final objectives: data reuse and valorization.

3.3. A holistic DPIA for Data Spaces.

A final crucial aspect linked to the principle of accountability in Data Spaces concerns risk management and the drafting of a data protection impact assessment (DPIA). As we know, a DPIA is 'a process intended to describe the processing, assess its necessity and proportionality, as well as help manage the risks to the personal rights and freedoms of natural persons arising from the processing of data by evaluating those risks and determining measures to address them'[51]. The drafting of a DPIA is mandatory whenever processing of personal data entails high risks for the rights and freedoms of natural persons[52], but it can also be drawn up in additional hypotheses depending on the discretionality of the data controller. This assessment must be carried out not only when prejudice to personal data of the Data Subject

[46] IE Olatunji, J Rauch, M Katzensteiner et al. 'A Review of Anonymization for Healthcare Data' (2022) Big Data. <https://doi.org/10.1089/big.2021.0169> accessed 3 August 2024.
[47] General Data Protection Regulation [2016] Recital 26.
[48] N. Savage 'Synthetic data could be better than real data.' Nature. <https://doi.org/10.1038/d41586-023-01445-8> accessed 3 August 2024.
[49] S M Bellovin, P K Dutta, N Reitinger 'Privacy and Synthetic Datasets' (2019) Stanford Technology Law Review 22(1). <https://law.stanford.edu/publications/privacy-and-synthetic-datasets/> accessed 2 September 2024.
[50] B Van Breugel, M Schaar 'Beyond Privacy: Navigating the Opportunities and Challenges of Synthetic Data' (2023). <https://doi.org/10.48550/arXiv.2304.03722> accessed 4 September 2024.
[51] Art. 29 Data Protection Working Party 'Guidelines on Data Protection Impact Assessment (DPIA)' [2017] 4. <https://ec.europa.eu/newsroom/article29/items/611236/en> accessed 4 August 2024.
[52] General Data Protection Regulation [2016] art 35.

might occur but in any case of jeopardizing any right or freedom[53] of any citizen. Those prejudices could arise not only from a pathological situation (e.g., a prohibited action or an unpredictable event) but also from the physiological performance of proceeding activities. In other words, the document has to map every potential source of prejudice for human rights or freedoms.

The degree of risk must be 'high'. The European legislator considers it recurring in a bunch of hypotheses listed in Art. 35 GDPR and integrated by an additional inventory drawn up by the Article 29 Working Group (WP29) in its Guidelines[54]. In my opinion some of them are particularly relevant in the context of Data Spaces. These are:

A) Data processed on a large scale.
   This circumstance certainly emerges in a Data Space, given the potential volume of data generally transacted and the recurrence of processing activities performed in this ecosystem. Furthermore, the sub-requisite of 'geographical scope' set out in the WP29 Guidelines might occur if the Data Space has a large dimension (at the national or international level).
B) Matching or combining datasets.
   This hypothesis occurs when processing operations carried out by many data controllers within a Data Space are performed in ways that go beyond the reasonable expectations[55] of the Data Subjects. It is plausible that, to date, the context in which information is analyzed and combined by data controllers does not allow the Data Subjects to predict the circumstances and boundaries of that proceeding within a Data Space. The implementation of real collaborative projects is so recent that we do not have an in-depth knowledge of its internal dynamics. Therefore the reasonable expectations of the Data Subject may not cover the processing of his/hers data performed within a Data Space.
C) Innovative use or applying new organizational solutions.
   Although in the WP29 Guidelines the Working Group focuses mainly on the use of new *technological* solutions, joining a Data Space involves new *organizational* methodologies too. Controllers are forced to rethink the management and governance logic underlying processing operations. In my opinion, even in this hypothesis the Data Space use case constitutes a *novum* which could open the door to unknown perils to individual rights and freedoms.

Following the suggestions set out in the WP29 Guidelines, where a treatment meets at least two of the criteria listed therein (three in our case) it must be subject to a DPIA. Therefore, in most cases, data processing performed within a Data Space must be preceded by the drafting of a DPIA by the data controller.

Looking at the risks that processing activities could generate, it is possible to identify two different categories of perils: internal and external. These risks cover the entire data life cycle within the Data Space ecosystem, from input to output[56].

---

[53] Art. 29 Working Party 'Guidelines on Data Protection Impact Assessment (DPIA)' [2017] 6.
[54] Art. 29 Working Party 'Guidelines on Data Protection Impact Assessment (DPIA)' [2017] 9 ff.
[55] Art. 29 Data Protection Working Party 'Opinion on Purpose limitation 13/EN WP 203' [2013] p.24 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> accessed 4 August 2024.
[56] I Barbera, C Castelluccia, G D'acquisition et al. 'Engineering Personal Data Protection in EU Data Spaces' (ENISA, 2024) 11 ff. <https://www.enisa.europa.eu/publications/engineering-personal-data-protection-in-eu-data-spaces> accessed 5 August 2024.

The cluster of *internal* risks include all possible dangers underlying any processing operation, also in 1:1 relationships. That is, they might happen whenever processing of personal data is carried out and do not vary depending on the number of subjects involved in such activities. On the contrary, *external* risks are those linked to the collaborative environment of the Data Space. In other words, they hugely depend on the involvement of multiple actors in processing operations. Sometimes the synergistic action of various stakeholders can generate new or unpredictable threats that are difficult to capture by adopting an individualistic point of view. Let us consider the following example: 'encrypted personal data are stored at one data processor in the processing chain and the corresponding encryption key is stored at another data processor. (...) However, if those two data processors happen to use the same data processor for the actual storage, the combination of both instances may become a severe risk to the data processing, as now both the decryption key and the encrypted data are in the hands of the same organization (and all of its potential hackers that have data access)'[57].

The drafting of a DPIA concerning data processing within a Data Space requires the adoption of a holistic approach aimed at evaluating not only risk factors by themselves (internal) but also the possibility of prejudicial factors caused by the simultaneous action of multiple players in the ecosystem (external). It is necessary to detect all the risk sources, including the ones resulting from the interaction among the participants in the Data Space and their network.

Given these premises, the development of a "holistic" DPIA is mandatory. This document covers multiple processings carried out within a Data Space. The GDPR offers data controllers the possibility of drawing up a joint DPIA that examines a cluster of similar treatments, which in turn present similar (high) risks, if this appears reasonable and economical[58]. This chance is particularly useful in the case of multiple treatments that are similar in terms of nature, scope, context, purpose and risks[59]. The creation of a “holistic” DPIA allows the different data controllers to map all the potential risks in a single document and, by doing so, more easily identify synergistic sources of perils. The most important aspect of this chance concerns the possibility of adopting a broad-spectrum approach and understanding which risks may arise in the collaborative ecosystem where processing operations are carried out. The existence and severity of such risks may not be adequately identified in the case of drafting single DPIAs by each data controller. By opting for this solution, it is necessary to indicate the reason why data controllers chose to draft a joint impact assessment and document accessibility has to be guaranteed in an easy way.

The drafting of a "holistic" DPIA does not exempt individual data controllers from the obligation to draw up an impact assessment by themselves, if the processing operations they will perform are not entirely covered by the joint document. In such cases, data controllers have to perform their own impact assessment but the integration of that document with the "holistic" DPIA is allowed and appropriate.

4. Conclusions.

Previous sections support drawing some conclusions regarding privacy

[57] I Barbera, C Castelluccia, G D'acquisition et al. ‘Engineering Personal Data Protection in EU Data Spaces’ (ENISA, 2024) 14.
[58] General Data Protection Regulation [2016] art 35.1; Recital 92.
[59] I Barbera et al. ‘Engineering Personal Data Protection in EU Data Spaces’ (ENISA, 2024) 7.

compliance within the Data Spaces.

Data Spaces constitute a new collaborative model introduced by the European Union with the aim of encouraging information sharing and reuse between different stakeholders. The collaborative shift has only partially occurred so far and will receive a stronger application within data-driven ecosystems in the future. So nowadays it is necessary to ask ourselves how data protection rules and principles can be inspired by this paradigm shift, in order to fully protect individual rights and freedoms in the following years.

Various Data Space stakeholders could be identified. By taking part in a transaction of personal data, they assume one of the well-known privacy roles listed in the GDPR. The definition and allocation of these roles is really important to define the obligations and responsibilities of each actor. The Data Subject could be a participant in the Data Space or not and this circumstance makes the fulfillment of privacy obligations by data controllers and processors more difficult.

The further processing of personal data poses some challenges, where it may be difficult for the data controller to put in place the necessary guarantees required by the GDPR. The Proposal for a Regulation on the European Health Data Space has not fixed this issue. On the contrary, the concept of 'secondary use of (electronic health) data' generates larger uncertainties about the eventual overlap of this concept with the traditional theorization of ‘further proceeding’.

The use of Privacy Enhancing Technologies (PETs) and, in particular, data anonymization techniques enhances data sharing and reuse within Data Spaces in a lawful manner. Data synthesization guarantees a perfect balance between the need to ensure the protection of personal data and the usefulness of maintaining a high level of statistical representativeness in the original dataset, in order to maximize the exploitation of such information by the Data Space actors.

Finally, when the drafting of a data protection impact assessment is required, the adoption of a holistic approach that takes into account both internal and inter-organizational risks allows data controllers to map a series of hidden prejudices that cannot be identified otherwise (e.g., by following a "traditional" individualistic approach).

In conclusion, the novelty and main contribution of this paper is to offer an overview of the main data protection building blocks related to the development of a Data Space project. This broad analysis helps us to define the evolution of data protection rules from a collaborative perspective more deeply. Similar analysis has not been established by the literature yet.

Future research could further investigate this topic and propose quantitative and qualitative analysis in order to verify to what extent the challenges aforementioned prevent the success of Data Spaces projects in practice.

This study also has some limitations. The introduction of the Data Space framework is so recent that it could be difficult to properly manage this topic, given the lack of use cases and practical initiatives to investigate. It prevents us from having a complete perception of the real challenges posed by such projects in the data protection field. The further implementation of Data Spaces will allow us both to verify whether the challenges listed above could be really fixed by the adoption of the suggested building blocks.