APPROACH TO DATA SPACES FROM GDPR PERSPECTIVE

v. May 2023

EXECUTIVE SUMMARY

Technology enables both private companies and Public Administrations to use personal data on an unprecedented scale in conducting their activities. The digital transformation and the increasing use of digital services also entails new risks and challenges for individual recipients of the services concerned, businesses and society as a whole.

Within this framework, the purpose of the GDPR is to provide a solid and coherent framework for the protection of fundamental rights in relation to data protection in the European Union and thus to ensure a uniform, homogeneous and high level of protection throughout the Union. In this way, by ensuring control by individuals over their own personal data, trust is created and legal and practical security for individuals, economic operators and Public Administrations is strengthened. Effective guarantees of the fundamental right to the protection of personal data and a set of homogeneous principles, rights and enforcement tools across the states of the European Union will result in the free flow of personal data within the European Union, and the development of the European digital market. The EU legal framework in the area of personal data protection is, therefore, an enabler, not an obstacle, to the development of data economy that corresponds to the values and principles of the Union, and is the basis on which to build a European model of data governance.

A Data Space can be defined as a federated and open infrastructure to allow sovereign access to data, based on governance, policies, rules and standards that define a framework of trust for all parties involved. The European and national Data Space initiatives propose processing models of great organisational and technological complexity, as well as a large scale in terms of the number of subjects concerned, the diversity of categories of data processed, the social strata involved, the geographical scope, the retention periods, the number of parties involved and others. These initiatives are not considered as a reduction or a compromise of the rights and freedoms of natural persons in relation to the protection of their personal data, but rather open up a horizon of possibilities which, in order to guarantee sustainability in relation to the European model of rights and freedoms, require an objective and critical analysis of their implementation from the design that is in accordance with the impact of the processing.

This document is a first approach to GDPR compliance of Data Spaces by applying the principles of proactive accountability and data protection by design. Without seeking to transpose the text of the GDPR to this document, nor to be exhaustive, the document addresses the set of definitions from the GDPR, the various European standards, specific standards and vocabulary in the field of Data Spaces. This is followed by a brief list of the basic regulatory framework, and in development at the time of drafting this document, which affects Data Spaces when they involve the processing of personal data. This is followed by an approach to the processing of personal data in the framework of Data Spaces.

The document closes with two main chapters. The first is on the applicability of data protection by design in Data Spaces. To this end, it should be taken into account that the

access to data allowed by a Data Space is defined as any use of data in accordance with specific requirements, but, importantly, without necessarily implying the transmission or downloading of the data. In this respect, there are different technological resources that allow the re-use of personal data with data protection guarantees offering more options than anonymisation. The document closes with a chapter on issues related to data protection in Data Spaces with regard to such important aspects as the involvement of Data Protection Officers, or the management of the risk to the rights and freedoms of natural persons, both from the point of view of their individuality and from their social perspective, among others.

The 'Approach to data spaces from a data protection law perspective' is addressed to controllers and processors involved in Data Spaces as well as to Data Protection Officers, data protection advisors and all those involved in a data sharing model who process, authorise, supervise or facilitate the processing of personal data within the framework of a data sharing model.

Key words: data space, data protection by design, risk management, European data strategy, GDPR, data governance regulation, DGA.

Prepared in collaboration with and reviewed by PhD. Alberto Palomo Lozano (Head of the Data Office of the State Secretariat for Digitalisation and Artificial Intelligence (SEDIA, by its initials in Spanish) of the Ministry of Economic Affairs and Digital Transformation (MINETD, by its initials in Spanish)), Mr. Carlos Alonso Peña (Director of the Data Office Division of SEDIA-MINETD), Mr. Rafael Pérez Galindo (Deputy Director General for the Digital Society of SEDIA-MINETD, Mr. Jesús Jiménez López (Director of the Council for Transparency and Data Protection of Andalusia (CTPD, by its initials in Spanish)), Mr. Manuel González Seco (Head of the CTPD Compliance Office), PhD. Sara Degli-Esposti (Research Scientist at the del IFS-CSIC of the Scientific Research Council), PhD. Rafael Pastor (Director/Dean of the Higher Technical School of Computer Engineering of the National University of Distance Education) and by PhD. Ricard Martínez Martínez (Associate Professor at the Department of Constitutional Law of the Universitat de València).

CONTENTS

Index of figures

Acronyms

AEPD:	Spanish Data Protection Authority (by its initials in Spanish: Agencia Española de Protección de Datos)
AIA:	Artificial Intelligence Act
CJEU:	Court of Justice of the European Union
DA:	Data Act
DGA:	Data Governance Act
DMA:	Digital Market Act
DMZ:	Demilitarised Zone
DPIA:	Data Protection Impact Assessment
DPO:	Data Protection Officer
DSA:	Digital Services Act
ECHR:	European Court of Human Rights
EDPB:	European Data Protection Board
EDPS:	European Data Protection Supervisor
EHDS:	European Health Data Space
ENISA:	European Union Agency for Cybersecurity
ENS:	National Security Scheme
ETL:	Extract, Transform, Load
EU:	European Union
GDPR:	General Data Protection Regulation
IoT:	Internet of Things
LOPDGDD:	Spanish Organic Law on the Protection of Personal Data and the guarantee of digital rights. (by its initials in Spanish: Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales)
PET:	Privacy-Enhancing Technology
PIR:	Private Information Retrieval
SDK:	Software Development Kit
SMPC:	Secure Multi-Party Computation

I. INTRODUCTION

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, and repealing Directive 95/46/CE (hereinafter General Data Protection Regulation, or GDPR) is the law that protects the fundamental rights and freedoms of natural personas and, in particular, their right to the protection of personal data[1].

The protection of natural persons with regard to the processing of their personal data is a fundamental right. Article 8(1) of the Charter of Fundamental Rights of the European Union ('the Charter') and Article 16(1) of the Treaty on the Functioning of the European Union (TFUE) provide that everyone has the right to the protection of personal data concerning him or her[2].

Technology allows both private companies and public administrations to use personal data on an unprecedented scale in the conduct of their business[3]. The digital transformation and the increasing use of digital services also entails new risks and challenges for individual recipients of digital services, businesses and society as a whole[4]. Therefore, the GDPR stems from the need to reinforce and specify the rights of Data Subjects and the obligations of those who process and determine the processing of personal data[5].

The purpose of the GDPR is to provide a robust and coherent framework[6] for the protection of fundamental rights in relation to data protection in the European Union and thus to ensure a uniform, consistent and high level of protection across the Union[7]. The aim of the regulation is to ensure control by natural persons over their own personal data[8], to build confidence and to strengthen legal and practical certainty for natural persons, economic operators and public authorities. To achieve this objective, in addition to the existence of the regulation, strict enforcement[9] and supervision exercised in an equivalent manner[10] between Member States will be necessary.

Effective guarantees of the fundamental right to the protection of personal data and a set of homogeneous principles, rights and enforcement tools[11] across EU States will result in the free flow of personal data within the Union[12]. If there are divergences between Member States as to the level of guarantee of the right to data protection, or the way it is supervised, this would impede free movement and the proper functioning of the internal

[1] Article 1(2) of the GDPR
[2] Recital 1 of the GDPR
[3] Recital 7 of the GDPR
[4] Recital 1 of the DSA
[5] Recital 11 of the GDPR
[6] Recital 7 of the GDPR
[7] Recital 10 of the GDPR
[8] Recital 7 of the GDPR
[9] Recital 7 of the GDPR
[10] Recital 11 of the GDPR
[11] In contrast to the implementation of Directive 95/46/CE of the European parliament and of the Council of 24 October 1995, on the protection of individuals with regard to the processing of personal data and on the free movement of such data
[12] Recital 13 of the GDPR

market. Therefore, requiring[13] the same level of compliance with the GDPR is essential for the free movement of persons[14], goods and services, and for the development of the European digital market[15]. The EU legal framework in the area of data protection is an enabler, not an obstacle, to the development of a data economy that corresponds to the values and principles of the Union, and is the basis on which to build a European data governance model[16].

This document analyses, from a data protection perspective, a technological model for the efficient implementation of the 'internal data market'[17] called Data Spaces, as it involves the processing of personal data[18]. A Data Space has no single definition, although it is possible to define it as a federated and open infrastructure to enable sovereign access to data, based on governance, policies, rules and standards that define a framework of trust for all interveners[19]. In regulatory terms, legal differences arise with regard to the definition of the interveners, the limits of the processing and the necessary safeguards, depending on whether it is a matter of re-use of data held by public sector bodies[20], services to establish commercial relations between the parties involved[21] or, for example, Data Spaces in specific sectors[22]. A Data Space is distinct from centralised information storage, data lakes[23], data warehouses[24], bilateral data sharing or neutral points, although the underlying technologies for the creation of a Data Space may in many cases overlap with the technologies with which the above solutions are implemented.

The European and national initiatives of Data Spaces, and their regulatory developments, propose processing models of great organisational, legal and technological complexity, as well as of a large scale in the number of subjects affected, in the diversity of categories of data processed, in the social strata involved, in the geographical scope, in the conservation periods, in the extension in time of the processing, in the number of intervening parties and others. The Data Spaces open up a horizon of great opportunities, and in order to guarantee them, as well as to guarantee sustainability in relation to the European model of rights and freedoms, they require an objective and critical analysis of their implementation from the design that is in accordance with the impact of the processing.

---

[13] Recital 13 of the GDPR

[14] Schengen Acquis EUR-Lex - I33020 - EN - EUR-Lex (europa.eu)

[15] An equivalent high level of protection in all states is what guarantees the free flow of data, not only from a data protection point of view, but also as provided for in other areas by the DGA, the DMA, the DSA and the DA y de IAA proposals. The lax and limited application of the levels of protection set out in legislation would be the factor that effectively impedes the single market.

[16] Paragraph 20 of the document 'Joint EDPB-EDPS Opinion 3/2021 on the Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) [10 March 2021]'.

[17] Recital 2 of the DGA

[18] In Data Spaces where there is no processing of personal data, e.g., only data from industrial environments, this document would not apply.

[19] 'What is a Data Space? Definition of the concept Data Space. White Paper 1/2022. (Gaia-x – Hub Germany) [September 2022]' Whitepaper Definition Datenraum (gaia-x-hub.de)

[20] Chapter II of the DGA, with the limitations developed in Recital 12, and Directive 2019/1024.

[21] Chapter III of the DGA

[22] As could be the EHDS proposal in the health data sector.

[23] A single repository of heterogeneous data (structured, unstructured or other) that allows it to be catalogued and transformed to be used for tasks such as reporting, visualisation, advanced analytics and machine learning.

[24] Internal operational or analytical databases that organisations need/use for their operations.

These initiatives are not intended to reduce or compromise the rights and freedoms of individuals in relation to the protection of their personal data. In this regard, it should be borne in mind that the access to data that arises in a Data Space is defined as any use of data in accordance with specific requirements, but, and it is of great importance to qualify this, without necessarily implying the transmission or downloading of data, nor displacing the principles and rights of the GDPR. In this respect, there are different technological resources that allow access to personal data with data protection guarantees, offering more options than just anonymisation or resorting to the communication of personal data.

This document constitutes a first approach to GDPR compliance for Data Spaces. It could not be otherwise when, in addition to the intrinsic complexity of the organisational and technical solutions proposed, there is the complexity of a regulatory package which, moreover, is in the midst of development. Although the obligation to apply the principles of proactive accountability and data protection by design is clearly stated in the GDPR, it would not be the first time that innovative processing has been deployed without taking these principles into account. Therefore, although the GDPR is applicable to all processing of personal data regardless of the technical means to implement it, and controllers should seek the advice of Data Protection Officers (DPOs) and data protection experts from the conception of the processing operations, early action by the Supervisory Authority is essential to ensure that certain aspects are correctly defined. Without seeking to transfer the text of the GDPR to this document, nor to be exhaustive, an association of the concepts and terms used will be made, guidelines will be provided to apply data protection measures and guarantees from the design in the processing in the framework of a Data Space, both in the definition of its architecture and in its governance mechanisms, and specific aspects on the principles, rights and obligations in relation to Data Spaces will be addressed.

This document is not a mandatory guideline and its interpretation should be without prejudice to the applicable sectoral regulations[25]. It is addressed to controllers and processors involved in the Data Space, as well as to DPOs, data protection advisors and all parties involved in a data sharing model who process personal data, or who authorise, supervise or facilitate, technically or organisationally, the processing of personal data.

In its preparation, special attention has been paid to the opinions on this subject issued by the European Data Protection Board (EDPB), European Data Protection Supervisor (EDPS) and the European Union Agency for Cybersecurity (ENISA) guidelines on privacy engineering[26].

The document is structured in the following chapters:

- Firstly, a chapter on definitions, in which definitions from the GDPR, from the different rules of the European digital package, from specific rules and vocabulary in the field of Data Spaces are grouped and related.

[25] If there is no processing of personal data, e.g., Regulation (EU) 2018/1807 on a framework for the free movement of non-personal data in the European Union would apply.
[26] References to these documents can be found in the references chapter at the end of the text.

- A chapter on the basic regulatory framework affecting Data Spaces when they involve the processing of personal data.
- A first approach to the processing of personal data in the framework of Data Spaces.
- The applicability of data protection by design in Data Spaces.
- And finally, a chapter with issues related to data protection in Data Spaces, which is not intended to be exhaustive, but a first approach.

II. DEFINITIONS

This chapter contains definitions of the most relevant terms used both in this document and in the regulations and technical references related to Data Spaces.

- Access: Any use of data in accordance with specific technical, legal or organisational requirements, without necessarily involving the transmission or downloading of data.[27]
- Data altruism: Any voluntary exchange of data based on the consent of data subjects to the processing of their personal data, or on the permission of data subjects to the use of their non-personal data, without the purpose of obtaining or receiving a reward that exceeds a compensation related to the costs incurred by them in providing their data, for purposes of general interest as provided for by national law, where applicable, such as, for example, healthcare, combating climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the provision of public services, public policy making or scientific research in the general interest.[28]
- Anonymisation: Processing on a set of personal data that generates a new set of data that disables the ability to relate these data to any identified or identifiable person.[29]
- Competent body: a public sector body that assists other public sector bodies in the re-use of data using state-of-the-art techniques, provides best practices on processing and on secure processing environments to preserve the privacy of information. Its tasks may include granting access to data, where required by Union or national sectoral legislation.[30]
- Compute-to-data: a strategy whereby, instead of sending the data to the computing resources, the computing resources are brought the origin of the data. In this way, the privacy of the data is preserved and the controller (Data Holder) retains greater control over the processing of the data. One way to implement compute-to-data is federated learning, but not the only way.
- Consent of the data subject: any freely given, specific, informed and unambiguous indication of the data subject's agreement, either by a statement or by a clear affirmative action, to the processing of personal data relating to him or her.[31]
- Controller: the natural or legal person, public authority, service or other body which alone or jointly with others determines the purposes and means of the processing; if Union or Member State law determines the purposes and means of the processing, the controller or the specific criteria for its appointment may also be established by Union or Member State law.[32] The concept of GDPR

[27] Article 2(13) of the DGA
[28] Article 2(16) of the DGA
[29] Paragraph 2.2 of the document "WP 216 Opinion 05/2014 on anonymisation techniques (Article 29 Data Protection Working Party) [10 April 2014]"
[30] Article 7 and Recital 26 of the DGA
[31] Articles 4, 6 and 7, recitals 32, 42 and 43 of the GDPR
[32] Article 4 of the GDPR

controller should not be confused with the concept of a functional controller of a process or department in the organic attribution of duties in an entity. This will only be an authorised user within the organisation.

- Core platform service[33]: can apply to any of the following elements: online intermediation services; online search engines; online social networking services; video-sharing platform services; number-independent interpersonal communications services; operating systems; web browsers; virtual assistants; cloud computing services; online advertising services.
- Data: in the framework of Data Spaces, any representation of acts, facts or information, and any compilation of such acts, facts or information, including in the form of sound, visual or audio-visual recording.[34]
- Data aggregator: a service that allows data from different sources to be brought together in one place.
- Data intermediation service[35]: as defined in the DGA, any service the purpose of which is to establish commercial relationships for data sharing between an undetermined number of data subjects and data holders, on the one hand, and data users, on the other hand, by technical, legal or other means, including services aimed at the exercise of data subjects' rights in relation to personal data[36]. This definition excludes, at least, services that obtain data from Data Holders and process them for the purpose of adding substantial value and grant licences to Data Users, without establishing a commercial relationship between Data Holders and Data Users; services dedicated to the intermediation of copyrighted content; services used exclusively by a single Data Holder to enable the use of his or her data; those used by multiple legal entities in a closed group, including also those used in supplier or customer relationships or contractually established collaborations, in particular those whose main purpose is to ensure the functionalities of objects and devices connected to the internet of things[37]; data sharing services offered by public sector bodies without the intention of establishing commercial relationships[38].
- Data catalogue: a collection of dataset descriptions, organised in a systematic way and containing a public user-oriented part, where information on individual dataset parameters can be accessed electronically through an online portal[39].
- Data cataloguing: processing carried out on data or a set of data that allows the metadata necessary for their subsequent exploitation to be associated with

[33] Article 2(2) of the DMA
[34] Adapted from article 2(1) of the DGA
[35] Article 2(11) of the DGA
[36] Not to be confused with 'intermediary service', defined in Article 3(g) of the DSA, or 'online intermediation service', defined in the P2B, Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services.
[37] In such cases, reference should be made to the DA proposal.
[38] For example, the competent bodies established in Article 7 of the DGA
[39] Article 2(2)(ac) of the proposal EHDS

them. It generally involves the generation of resource (data) catalogues that can be made available to multiple interveners.[40]

- Data cooperatives: data intermediation services offered by an organisational structure made up of data subjects, sole proprietorships or SMEs belonging to such a structure, the main purposes of which are to assist their members in exercising their rights in relation to certain data, including assistance in making informed decisions before consenting to data processing, to exchange views on the purposes of data processing and the conditions which best represent the interests of its members in relation to their data, and to negotiate contractual conditions for the processing of data on behalf of its members before granting permission for the processing of non-personal data or before giving consent to the processing of personal data.[41]
- Data extraction: processing of a data set to generate a new data set more in line with the needs of a use case. In the new set, the extent of the data may be limited in its data categories (e.g., not all attributes), its granularity (e.g., not extracting the full address but the postcode), its frequency (e.g., only one location position per day), precision (e.g., instead of the age of the data subject, indicate only the qualification in minors or adults), etc.
- Data Holder: a legal person, including public sector bodies and international organisations, or natural person other than the data subject with regard to the specific data concerned, who, in accordance with applicable Union or national law, has the right to grant access to certain personal or non-personal data or to share them.[42]
- Data lifecycle: From a Data Space perspective, the lifecycle refers to the different stages a piece of data goes through from its birth to its end. Data is not a static asset during its life cycle, but goes through different phases. Without being exhaustive, and in no particular order, these could be: extraction, loading, transformation, maintenance, synthesis, use, publication, storage or disposal. The concept of data lifecycle in the framework of Data Space should not be confused with the data lifecycle in a processing.[43]
- Data quality: From the point of view of a Data Space[44], data quality is a subjective attribute[45] associated with a set of data about its usefulness for a specific processing[46]. This concept is distinct from the data quality of the accuracy principle[47] of the GDPR.

[40] Adapted from the 28/12/2020 publication of the Data Office
[41] Article 2(15) and Recital 31 of the DGA.
[42] Article 2(8) of the DGA
[43] Section 'V.C DESCRIPTION OF THE DATA LIFECYCLE' from the Guide 'Risk management and impact assessment in processing personal data' of AEPD and from the publication of 28/12/2020 of the Data Office on the importance of data cataloguing.
[44] Standards UNE 0079 and ISO 25012 can be consulted.
[45] Data on the Web Best Practices: Data Quality Vocabulary (w3.org)
[46] Article 2(ad) of the proposed EHDS defines it as: the degree to which the characteristics of electronic health data are suitable for secondary use.
[47] Article 5(1)(d) of the GDPR

- Data sharing: the provision of data by a Data Subject or Data Holder to a Data User, directly or through an intermediary and under a voluntary agreement or under Union or national law, for the purpose of joint or individual use of such data, for example, through open licences or through paid or free commercial licences.[48]
- Data sovereignty: a concept not defined in the European standard and generally interpreted as the idea that the place where data is collected determines the regulation and governance that applies to it, and also the ability of governments and companies to use of users' and companies' digital data.
- Data Space Mediator, Data Mediator: entities that establish the relationships in the Data Space between Data Subjects and/or Data Holders, on the one hand, and Data Users, on the other hand. In the framework of the DGA 'competent bodies[49] shall be considered as mediators, 'data intermediation services' (and their subtype 'data cooperatives') and 'data management organisations for altruistic purposes' shall be considered as mediators under the DGA. Under the EHDS proposal it will be, among others, the central platform for secondary use of electronic health data. In other areas they are referred to as 'data provider', 'data space operator', etc.
- Data Space: infrastructure based on common governance, organisational, regulatory and technical mechanisms, which facilitates access to data and thus the development of business models based on its exploration and exploitation.
- Data Subject: identified or identifiable natural person.[50]
- Data traceability: the ability to know the entire life cycle of the data.[51]
- Data user: a natural or legal person who has lawful access to certain personal or non-personal data and the right, including the one granted by the GDPR in the case of personal data, to use that data for commercial or non-commercial purposes.[52]
- Dynamic and static personal data: Data spaces could contain static personal data, such as name, address or date of birth, as well as dynamic data generated by an individual, for example, through the use of an online service or an object connected to the Internet of Things. They could also be used to store verified identity information, for example, passport number or social security information, and credentials (e.g., driving licence, diplomas or bank account information).[53]
- ELT, ETL, EtLT: acronyms referring to the processes of Extraction, Loading and Transformation of data. The lower case 't' refers to processes prior to the loading and transformation of data into formats suitable for a particular processing. For

[48] Article 2(10) del DGA
[49] Article 7 of the DGA
[50] Article 4 of the GDPR
[51] Adapted from the publication of 28/12/2020 of the Data Office on the importance of data cataloguing.
[52] Article 2(9) of the DGA
[53] Recital 30 of the DGA

example, the 't' could refer to anonymisation or pseudonymisation processing activities.

- Enabler: actor(s) providing services or tools that allow sharing or exploiting datasets and implementing governance measures.[54]
- Federated learning: A machine learning technique that trains an algorithm through a decentralised architecture of devices containing their own local and private data. Created by Google in 2017, this approach contrasts with techniques where all data is uploaded centrally to a server. This preserves the integrity of the information being used for learning without compromising privacy and security.
- Gatekeeper: is defined in the DMA as a company providing core platform services, for the purpose of this document a cloud computing service, with a strong influence on the internal market and an established and long-lasting position.
- High Value Data or HVDS: documents whose re-use is associated with considerable benefits for society, the environment and the economy, in particular due to their suitability for the creation of new decent and quality value-added services, applications and jobs, and the number of potential beneficiaries of value-added services and applications based on such datasets[55].
- Hyperscale: It is a core platform service[56] that specifically provides mass storage and processing services in the cloud that can scale a distributed computing environment to thousands of servers.
- Metadata: in the framework of Data Spaces, are data about data and serve to provide information about the data we want to use. Metadata consist of information that characterises data, describes its content and structure, conditions of use, its quality for a context, its origin and transformation, among other relevant information. They can be of a technical, operational or business nature.[57]
- Mixed dataset: Mixed dataset consists of personal and non-personal data. Mixed datasets represent the majority of datasets used in the data economy and are common due to technological developments such as the Internet of Things (digitally connected objects), artificial intelligence and technologies that enable big data analytics[58].
- Non- personal data: data that do not fall within the scope of the definition of 'personal data'.
- Personal data: any information relating to an identified or identifiable natural person (data subject); an identifiable natural person is one who can be

[54] Tool to elaborate use cases in data spaces of the Spanish Data Office [only available in Spanish]
[55] Defined in Directive 2019/1024 transposed in Law 37/2007 on the re-use of information in the public sector.
[56] Article 2(2) of the DMA
[57] Adapted from the publication of 28/12/2020 of the Spanish Data Office on the importance of data cataloguing.
[58] Paragraph 2.2 of the Communication 'Guidance on the regulation on a framework for the free flow of non-personal data in the European Union (COM (2019) 250 final) [29 May 2019]'.

identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Encrypted or pseudonymised data are personal data.[59]

- Processing of personal data: any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or interconnection, restriction, erasure or destruction.[60]
- Processor: a natural or legal person, public authority, service or other body processing personal data on behalf of the controller, with an established link to the controller by means of a contract or other legal act and which complies with Article 28 of the GDPR. Within a single processing operation, there may be several processors, and these in turn have recourse to sub-processors (processors of processors).[61] A processor is never a person or department of the controller itself, but external to the controller itself.
- Protected data: Data held by public sector bodies that are protected for reasons of commercial confidentiality, protection of intellectual property rights of third parties, or protection of personal data, in so far as the latter are excluded from the scope of Directive (EU) 2019/1024[62].
- Pseudonymisation: the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.[63]
- Reuser: although not explicitly defined[64], it is inferred from the definition of re-use, that it is the natural or legal person who re-uses data held by sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the data were produced, except for the exchange of data between public sector bodies purely in pursuit of their public tasks[65].
- Secure processing environment: the physical or virtual environment and organisational means to ensure compliance with Union law, such as, for example, the GDPR, in particular with regard to data subjects' rights, intellectual

[59] Extended article 4 of the GDPR and its Recital 24
[60] Article 4 of the GDPR
[61] Article 4(8) of the GDPR
[62] Definition adapted from Article 3(1) of the DGA
[63] Article 4 of the GDPR
[64] Proposals of the DGA defined Re-user as the natural or legal person who re-uses data held by public sector bodies for commercial or non-commercial purposes other than the initial purpose encompassed by the public service mission for which the data were produced. This definition is not found in the current wording.
[65] Article 2(2) of the DGA

property rights and commercial and statistical confidentiality, integrity and accessibility, as well as to ensure compliance with applicable national law and to allow for the operator of the secure processing environment to determine and supervise all data processing actions, including the display, storage, download and export of data, as well as the calculation of derivative data through computational algorithms.[66]

- Supervisory body: shall be the entity with the obligation to assess each of the requests submitted by a Data User and to grant, or not, the requested request for processing, in particular, taking into account compliance with the provisions of the GDPR.
- Trusted execution environment: an inviolable processing environment that takes place on the main processor of a device with hardware and software designed in such a way as to guarantee the integrity and confidentiality of the data and processing carried out on that processor against any type of attack. Not to be confused with Secure Processing Environment where, in addition to the aspects of confidentiality, integrity and availability of the data, the legal obligations laid down in national and EU law are guaranteed[67].

[66] Article 2(20) of the DGA

[67] Paragraph 4.3 of the document 'DATA PROTECTION ENGINEERING. From Theory to Practice. European Union Agency for Cybersecurity (ENISA) [January 2022]'

III. THE REGULATORY FRAMEWORK FOR DATA SPACES

To the extent that personal data are processed in a Data Space, the regulatory framework begins to be defined by the General Data Protection Regulation and the Organic Law 3/2018, of 5 December, on the Protection of Personal Data and the guarantee of digital rights (LOPDGDD).

In relation to the use of data in the digital world, a package of European and national rules is being developed which do not change the personal data processing regime for any of the regulated activities or the information requirements laid down in the GDPR[68], and in case of conflict with Union law on the protection of personal data or national law adopted in the field of personal data protection, the latter should prevail[69].

A. EUROPEAN DATA FRAMEWORK

Without being exhaustive, the following basic rules are listed below:

- The Data Governance Act (DGA)[70] which regulates the conditions for the re-use of certain categories of data held by public sector bodies and also defines categories of interveners in a Data Space for the use of data from both the private and public sector. The DGA defines conditions and guarantees for new data business models, such as data intermediation services, as well as altruistic data transfers, among others[71]. The DGA complements 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 inter alia by defining conditions for the re-use of certain categories of data held by the public sector that are protected for reasons of commercial and statistical confidentiality, intellectual property rights or personal data.
- The Digital Markets Act (DMA)[72] regulating core platform services provided or offered by Gatekeepers, in particular, those relating to cloud computing services.
- The Digital Services Act (DSA)[73] laying down harmonised rules on the provision of intermediary services in the internal market.
- The Regulation on the free flow of non-personal data[74] and the Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union , in relation to the processing of mixed data sets.

[68] Article 1(3) and Recital 4 of the DGA, article 1.3 and Recital 7 and 24 of the DA proposal.
[69] Article 1(3) and Recital 4 of the DGA, article 2(4) of the DSA, Recital 24 and article 1(3) of the DA proposal.
[70] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (UE) 2018/1724 (Data Governance Regulation or DGA),
[71] Data Governance Law explained | Shaping Europe's digital future.
[72] 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 or DMA)
[73] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/CE (Digital Services Act or DSA)
[74] 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

- The Commission Implementing Regulation (UE) 2023/138 of 21 December laying down a list of specific high-value datasets and the arrangements for their publication and re-use .

B. PROPOSALS FOR EUROPEAN REGULATION

The digital regulation package is complemented by the European Commission's proposals still in process, of which the following are worth highlighting, without wishing to be exhaustive:

- The proposal for a Data Regulation (DA) which, among other aspects, extends the rights of access to non-personal data and devotes VIII to interoperability obligations in Data Spaces.
- The proposal for an European Health Data Space Regulation (EHDS).
- The proposal for an Artificial Intelligence Regulation (AIA).

It is also important to highlight all existing initiatives on European data spaces within the framework of the European Strategy for Data as well as the Digital Single Market, published in the Commission' "Staff working document on data spaces" on common European data spaces on 23 February 2022, in the framework of its Data Spaces work stream, which, among other things, sets out the areas in which to create these data spaces. The European Data Spaces, in addition to health, include the sectors of Manufacturing, Green Deal, Mobility, Financial, Energy, Agriculture, Legal, Procurement, Security, Skills, Open Science, Media, Cultural heritage, Tourism, Construction y Smart communities.

C. NATIONAL REGULATION

- Spanish Law 37/2007, of 16 November, on the re-use of public sector information.
- Spanish Law 34/2002, of 11 July, on information society services and electronic commerce.
- Spanish Law 40/2015, of 1 October, on the Legal Regime of the Public Sector.
- Spanish Royal Decree 311/2022, of 3 May, which regulates the National Security Scheme.
- Spanish Royal Decree 4/2010, of 8 January, which regulates the National Interoperability Scheme in the field of Electronic Administration.

D. PROPOSALS FOR NATIONAL REGULATION

At the national level, at the time of drafting these guidelines, the creation of an Integrated Mobility Data Space (IMDS) has been foreseen in the future Sustainable Mobility Law (preliminary draft law approved by the Council of Ministers on 12 December 2022).

The regulation will establish its creation, definition and governance[75]. In particular, in its article 104 on serious infringements, it identifies with regard to the provision of data to EDIM: *'the use for purposes other than the provision of data to EDIM of personal data*

[75] Articles 6, 14 and 8 of the Preliminary Draft Law on Sustainable Mobility.

obtained directly by transport operators, infrastructure managers and activity centres. In this case, the sanctioning procedure will be that established in Organic Law 3/2018, of 5 December, on the Protection of Personal Data and guarantee of digital rights by the competent body in matters of personal data protection.'

On the other hand, it is also worth mentioning at national level some lines of work in Data Spaces promoted by the General State Administration:

- Development of the national hub of Gaia-X (Gaia-X National Hub) for the development of an open and secure data infrastructure, for which several working groups are being created focusing on specific sectors: health, industry 4.0, tourism, mobility, agri-food, engineering and construction, enabling technologies, finance and public administration, as well as four other transversal working groups oriented towards legal, technical, project and ethical aspects.
- Project carried out by the Oficina del Dato (Spanish Data Office) in the tourism sector for the implementation of work sessions aimed at collecting use cases and the complementarity of data spaces with the Smart Tourist Destination SEGITTUR of the Ministry of Industry, Trade and Tourism.
- Creation of a health data lake in the framework of the Estrategia de Salud Digital (Digital Health Strategy) through the Digital Health Commission of the Interterritorial Council of the National Health System, according to component 18.I6 of the Recovery, Transformation and Resilience Plan.
- Creation of the NSA Data Platform around which future public sector data spaces will be deployed.
- Development of the Territorial Networks of Technological Specialization (RETECH) initiatives with data space components, financed by the Secretary of State for Digitalisation and Artificial Intelligence and developed by different Autonomous Regions in coordination.
- The future 'Strategic plan for the transformation and digitalisation of the agri-food system and the logistics chain for the promotion of high-value big data spaces that support the digital transformation of the productive sectors', through component 11.I2 of the Recovery, Transformation and Resilience Plan as part of one of the initiatives of the 'Tractor Project Launcher'.
- In addition, there are projects such as the Digitalisation of the Water Cycle, within the objective of improving the efficiency of the urban water cycle, which already contemplate the need for digital development in the sector in order to meet the requirements that are being established at European level for the sector's Data Space: '*The different information systems mentioned will have to guarantee a fluid sharing of data between them and with the appropriate external systems, following for this purpose the recommendations and guidelines set by the Spanish Government's Data Office, thus ensuring compliance wherever necessary with the National Interoperability Scheme (RD 4/2010) and the conditions and requirements derived from the European sectoral environmental data space (Common European GreenDeal dataspace).*'

IV. PROCESSING ON DATA SPACES

The material scope of the GDPR is processing of personal data [76]. In mixed data sets, where the processing of non-personal data is inextricably linked to personal data, the processing is also subject to the GDPR[77].

The GDPR does not have as its material scope technologies or technological infrastructures, as these are means to implement data processing. A Data Space is an infrastructure that allows multiple processing operations to be implemented. To the extent that a Data Space involves processing of personal data, it will be subject to data protection regulations, without prejudice to the applicable sectoral regulations[78].

Any re-use of personal data must always respect the principles of lawfulness, fairness and transparency, as well as purpose limitation, data minimisation, accuracy, retention period limitation, integrity and confidentiality, in accordance with Article 5 of the GDPR[79]. In order to fully comply with the GDPR, the processing operations must be well defined, and to this end, the purposes, data controllers and their legitimacy must be precisely determined.

A. CATEGORIES OF DATA SPACE INTERVENERS FROM A GDPR PERSPECTIVE

From a data protection point of view, the following interveners or roles in the Data Space could be identified:

1. Data subjects
2. Data holder
3. Data user
4. Data Space Mediators
5. Technical and legal enablers
6. Supervisor of access requests
7. Others, such as supervisory authorities.

This division of interveners has a didactic character. In practical application it will be possible to find entities performing various roles to a greater or lesser extent. Already in the DGA one type of Data Space is defined, Data Cooperatives, in which a natural person could act as Data Subject, Data Holder, Data Space Mediator and Data User[80]. On the other hand, the DGA also limits the roles that can be adopted in the case of Data Mediation Services[81].

[76] Article 2 of the GDPR

[77] Section 2.2 of the Communication 'Guidelines for a Regulation on a framework for the free flow of non-personal data in the European Union (COM (2019) 250 final) [29 May 2019]" and Recital 30 of the DA proposal.

[78] If there is no processing of personal data, e.g., Regulation (EU) 2018/1807 on a framework for the free movement of non-personal data in the European Union would apply.

[79] Paragraph 73 of the document 'Joint EDPB-EDPS Opinion 3/2021 on the Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) [10 March 2021]'.

[80] Recital 31 of the DGA

[81] Article 12(a) of the DGA '*the data intermediation services provider shall not use the data for which it provides data intermediation services for purposes other than to put them at the disposal of data users and shall provide data intermediation services through a separate legal person*'.

By way of illustration, a diagram showing the different relationships between the different interveners is included to support the description of the process below.

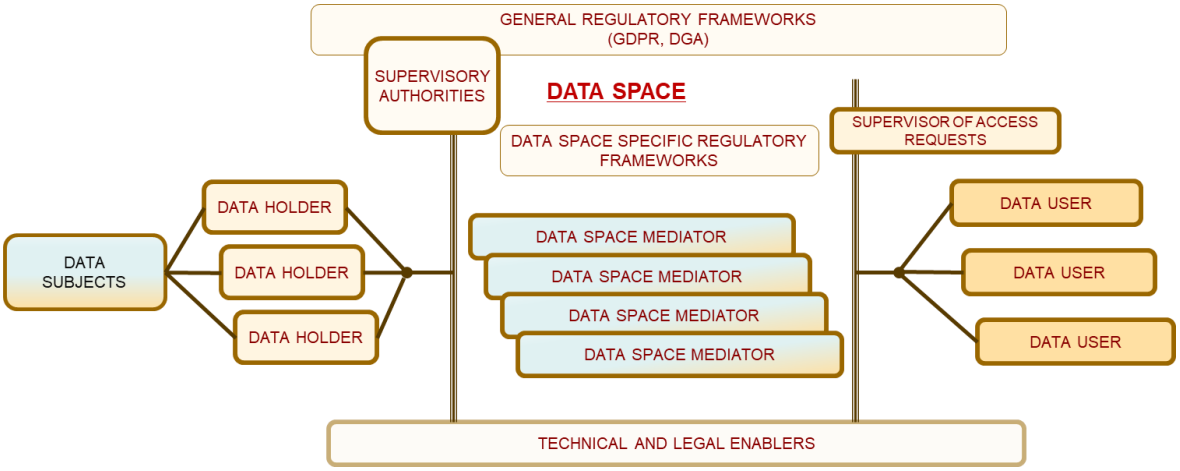

Figure 1: Diagram of interveners relationships from a GDPR perspective

In this section, a brief description of the interveners will be given, and in the last section of this chapter, the determination of the controller's/processor's role will be discussed.

Data Subject

In Data Space frameworks, the Data Subject, i.e., the identified or identifiable natural person whose personal data are intended to be processed, could be associated with the definition of 'data producer' used in some Data Space schemes. When the 'data producer' is associated with systems or services that collect or generate personal data of natural persons (e.g., IoT systems), to the extent that such data are linked to identified or identifiable individuals, we are still talking about data of a data subject.

However, when we talk about personal data, a data subject is not a mere 'data producer' but a natural person whose data can only be processed in compliance with the principles, legitimacy, respect for rights and other obligations set out in the GDPR.

Data Holder

In the chapter 'Definitions' of this document, the definition of Data Holder has been transcribed by transferring the definition established in the DGA, as a person who has the right to grant access to certain personal or non-personal data. Data Holders, within the framework of a Data Space, may perform data communication operations, implement mechanisms to enable on-premises processing, perform pseudonymisation and anonymisation processing, other processing such as extracting synthetic data, provide access with differential privacy implementation, perform data communications to other controllers or others.

In general Data Space frameworks one can find this figure labelled as the 'data owner' or 'data custodian'. This designation is misleading when referring to personal data, because a controller does not own the data of Data Subjects, but has a legal basis that legitimises him/her to process them in accordance with the obligations set out in data protection law.

The definition of Data Holder may change for certain specific Data Spaces established by law, defining entities of a particular sector or qualifying its definition[82].

Data user

A Data User is any natural or legal person who has a legitimate access to certain personal or non-personal data and the right, including the right under the GDPR in the case of personal data, to use it for commercial or non-commercial purposes[83]. In other areas it may be called by other names, such as 'data consumer'. In Directive 2019/1024[84] it uses the term 'end-user' for those re-users of data from public sector bodies who act as Data Users[85].

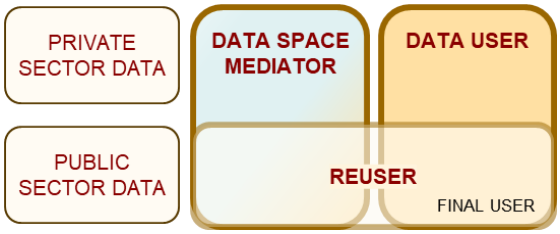

Figure 2: Correspondence between the terms Reuser, Data User (used in the DGA) and Data Space Mediator (e.g., the DGA's 'data mediation services' or 'data altruism organisations')

In the framework of a specific processing within a Data Space, one entity may act as Data User, in another processing it may act as Data Holder, while there may be processing where the same entity will be Data User of some data and Data Holder of others.

Data Space Mediator

Data Space Mediators are the entities that establish the relationships in the Data Space between Data Subjects and/or Data Holders, on the one hand, and Data Users, on the other hand. Those that implement the technical, legal, organisational, or other means that enable the operation of the Data Space between multiple Data Holders and multiple Data Users. Depending on the context in which the Data Mediator operates, it may have a different legal definition, e.g., 'data intermediation service', 'competent body', 'data management organisations for altruistic purposes', 'data cooperatives', etc. Similarly, technical references to Data Spaces may include names such as 'data provider', 'data space operator', or others.

Mediators may also be referred to as 're-users' in relation to the DGA when they process data from public sector bodies[86]. Mediators may also be public sector bodies.

[82] In the case of the EHDS proposal, 'data holder' is defined as any natural or legal person who is an entity or body in the health or care sector, or who carries out research in relation to these sectors, as well as Union institutions, bodies, offices and agencies which have a right or obligation under this Regulation, applicable Union law or national law implementing Union law, or, in the case of non-personal data, through the control of the technical design of a product and related services, to make available, as well as to record or hand over certain data, to restrict access to or to exchange such data.
[83] Article 2(9) of the DGA
[84] 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 version)
[85] Note that not all Data Users will be re-users of data from public sector bodies, nor will all re-users be Data Users.
[86] See definition of re-user.

Within the framework of a generic Data Space, one or more Mediator entities could carry out processing for the creation of data catalogues, creation of centralised databases, data transformation, creation of platforms for data sharing or exploitation, consent management, etc. In addition, the Mediator would keep track of all data sources and processing, evaluate and update data use policies throughout the data processing lifecycle. The Data Space Mediator would record data communications for each Data User with whom it interacts, and also for most Data Subjects, in addition to other functions[87]. It follows that these entities are key to the implementation of data protection measures by design and by default.

In the event that the Mediator is defined as a DGA data intermediation service, these entities will have to comply with the conditions for the provision of data mediation services set out in the DGA. The Data Space Mediator may include the offer of additional specific tools and services to Data Holders or Data Subjects for the specific purpose of facilitating data sharing (e.g., temporary storage, organisation, conversion, anonymisation and pseudonymisation), provided that such tools are only used upon the express request or approval of the Data Holder or the Data Subject, and that the third-party tools offered in that context are not used for other purposes[88]. The DGA[89] provides examples of 'data intermediation services' such as data marketplaces where companies could make data available to third parties, facilitators of data sharing ecosystems open to all interested parties, for example, in the context of common european data spaces, as well as datasets created in common by several natural or legal persons with the intention of licensing the use of such datasets, so that all participants contributing to their sharing receive a reward for their contribution.

In the case of Data Space Mediators managing data for altruistic purposes, where they have voluntarily decided to apply for registration in the national registers of data management organisations for altruistic purposes recognised in the Union, they will have to comply with the conditions and requirements of the DGA.

In the framework of the EHDS[90] proposal, the central platform for digital health, MyHealth@EU, the national contact point for secondary use of electronic health data and HealthData@EU could be considered as Data Space Mediators.

Supervisor of access requests

Supervisory bodies shall be those responsible for assessing requests submitted by a Data User for the processing of personal data.

Depending on the purpose of the Data Space, the granting of the request may be subject to different regulations and ethical principles. One of the regulations to be taken into account will be the specific and sectoral regulations in relation to data protection. These

[87] Section 4.3 of the document 'ENGINEERING PERSONAL DATA SHARING, Emerging Use Cases and Technologies. European Union Agency for Cybersecurity (ENISA) [January 2023]'.
[88] Article 12(e) of the DGA
[89] Recital 28 de la DGA
[90] Article 2 of the EHDS proposal.

regulations may require that such a function be performed by a body that is independent, or at least external to any other function in the Data Space.

In the event that requests include access to personal data, this Supervisor should set out the conditions under which access to data will be granted, depending, inter alia, on the legal basis on which the request is based and the measures that are presented to ensure and be able to demonstrate compliance. It should also set out in certain cases the conditions for access to personal data, independently of the risk management measures, including, inter alia, pseudonymisation and/or anonymisation mechanisms envisaged, the secure Processing Environment provided for the processing of personal data and the time limitation of access to personal data. Data Protection Impact Assessments (DPIAs) in relation to the requested processing should be part of the elements to be included in the requests.

Supervisory bodies cannot absolve controllers from their obligations to meet their obligations under EU data protection law, to be supervised by the independent supervisory authorities set up for that purpose[91].

The Supervisor could be part of the entity of a Data Space Mediator and exercise certain functions of the Data Protection Enabler. For example, this could be the case with the Competent Bodies defined in the DGA when required by Union or national sectoral rules[92]. If this is not the case, it is advisable to coordinate with the necessary Data Space Mediators and with the Data User to learn how to achieve data protection by design, with the support of the Data Protection Enabler who will guide the Data User on the actions to be taken to comply with the GDPR.

Enabler

The Enablers in a Data Space environment would be those that will support all the interveners described above in order to guarantee that the implementation is carried out in an efficient, coherent process, implementing governing and management mechanisms among multiple interveners, avoiding duplication and repetition of tasks, facilitating procedures and requests.

Enabler functions could include[93] providing components for accessing the Data Space, for mediation, for identity management and secure data communication or for managing the data space. Also providing applications to work with the data, such as machine learning models, visualisers, data cleansing or data quality analysis tools, etc., providing vocabularies and ontologies and orchestration services to automate various activities, among others.

The Single Point of Information[94] defined in the DGA, in its role of making available a searchable asset list containing a summary of all available data resources, would act as an Enabler. Also, as an example, a cloud storage service provider could be a technological Enabler contributing to the creation of such an infrastructure.

[91] Paragraph 29 of the document *"Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]"*
[92] Recital 26 of the DGA
[93] Article published by the Data Office on the 20/10/2022 on the main elements in a data space.
[94] Article 8 and Recital 26 of the DGA

An Enabler would not only be of a technical nature, but would also implement organisational measures (governance or control measures, for example, traceability or monetisation), as well as legal assistance. It is considered appropriate to emphasise the position of this intervener in the Data Space model from the point of view of the GDPR.

In the case of the processing of personal data, the possible figure of the Data Protection Enabler is of particular significance, providing support to the Data Holder, the Data Space Mediators and the Data User in order to guarantee compliance with the GDPR[95]. It should be borne in mind that when designing a personal data processing operation in a Data Space, it will be necessary to determine many aspects that are transversal to the different interveners, involving different possible architectures, different technological tools and different requirements and regulatory guarantees, in particular for the proper preparation of a DPIA of the processing operation from each of the perspectives of the interveners involved.

Supervisory authorities

In data protection matters, the competent authorities will be those indicated in the GDPR, which in the case of Spain will be the AEPD, or the Autonomous Authorities according to their competence. Where other authorities act as competent authorities, for example, under the DGA, they should do so without prejudice to the supervisory powers and competences of the data protection authorities under the GDPR[96].

Chapter III 'Requirements applicable to data intermediation services' of the DGA sets out the administrative requirements for the management of data intermediation services and their supervision by a competent authority for data intermediation services in relation to the notification and supervision requirements set out in Articles 11 and 12 of the DGA. In principle, and as in any other regulated sector subject to supervision by sectoral authorities, this does not imply, in the case of processing of personal data, a lack of competence of the data protection Supervisory Authority. Recital 44 of the DGA states that '*for any question requiring an assessment of compliance with Regulation (EU) 2016/679, the competent authority for data intermediation services should seek, where relevant, an opinion or decision of the competent supervisory authority established pursuant to that Regulation'*, however, this does not imply either that the Data Protection Authority delegates the exercise of its powers.

In any case, the AEPD emphasises the advisability of defining the most effective possible cooperation mechanisms between these sectoral authorities and the Data Protection Supervisory Authorities.

[95] The importance of the DPO in a Data Space is discussed below.
[96] Recital 4 and article 1(3) of the DGA

B. PROCESSING AND PURPOSES WITHIN THE FRAMEWORK OF A DATA SPACE

A processing operation is any operation or set of operations[97] which is performed upon personal data or sets of personal data, whether or not by automated means[98]. The most important aspect defining a processing operation is its purpose.

The purpose principle involves defining why certain personal data are processed. This means being as specific as possible about the purposes for which a processing operation justifies the collection and processing of personal data[99]. In order to define the purpose, an expression of will on the necessity of the processing, especially when it is based on generic purposes of social benefit, is not sufficient, but must be objectively grounded on the fact that these purposes will be achieved by the proposed processing, that is, the appropriateness analysis of the concrete implementation of the processing.

The purpose is what differentiates an isolated operation from the set of operations making up a processing operation. An operation in a processing shall be justified to the extent that it contributes with other operations to achieving the purpose of the processing. A technology (e.g., artificial intelligence, biometric, use of the cloud, etc.) is a means to implement one or more operations in the processing. It is the ultimate purpose on which the legitimacy of the processing will be considered.

On the other hand, in the framework of the Data Space, the data life cycle is defined as the set of different operations that could be executed on the data from its gestation to its elimination, without the intention of presenting an exhaustive list, we can enumerate the following:

- Data collection.
- Extraction[100] of data from datasets for the creation of new datasets.
- Transformations of data in relation to the nature of the dataset as relating to an identified or identifiable person (anonymisation or pseudonymisation).
- Access without dissemination of data.
- Communication by transmission or dissemination.
- Recording and storage of personal data.
- Syntactic or semantic transformations of the data set by organisation, structuring, adaptation or modification.
- Analysis of data for cataloguing and metadata generation.
- Anonymization and pseudonymisation of data.
- Generation of synthetic data.
- Analysis of risks of reidentification, quality of the resulting data, etc.
- Use or exploitation of data.

[97] Article 4(2) of the GDPR '…collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;"
[98] Article 4(2) of the GDPR
[99] Paragraph 5.7 of the document 'WP 211 Opinion 01/2014 on the application of the necessity and proportionality concepts and data protection within the law enforcement sector (Article 29 Working Party) [27 February 2014].
[100] Extraction can be a physical separation between a dataset and its copy, for processing without affecting the original data or for security, or because a subset of the original data is generated by reducing the extent of fields, records, temporal extent of the recorded data.

- Establishment of limitations on the data.
- Deletion or destruction.
- Other possible operations.

Each operation described above on the data does not necessarily imply a processing. The set of all possible operations on a piece of data during its entire life cycle will generally not be a single processing operation. The life cycle of the data may involve different processing and different controllers for each processing, which is the purpose of the infrastructure of a Data Space.

Regardless of its legitimacy, a processing operation could have the purpose of creating common repositories of anonymised data from personal data from different sources. This processing could involve operations of extraction, transformation of data for anonymisation, loading and storage under the control of the controller providing the common repository service.

Another possible processing would result from the analysis by a third party of sets of personal data held by multiple interveners for the purpose of creating a catalogue of content and location of data sources[101].

Processing would also include access by a Data User through the Data Space infrastructure to information from various personal data sources for the purpose of market research or any other ultimate purpose legitimately defined by a Data User. The set of involved operations and interveners could vary greatly depending on the capabilities of the Data Space.

In short, a Data Space may involve different processing operations, which could involve one or more controllers depending on how the architecture of the Data Space and the privacy safeguards by design are implemented for each use case.

C. LEGITIMISATION OF PROCESSING

The processing of personal data in the framework of a Data Space lacks a per se legitimation, as established by the DGA[102], and needs a specific legitimation based on Article 6 of the GDPR. For the re-use of data held by public sector bodies, Union or Member State law must provide for an appropriate legal basis under the GDPR, and public sector bodies must define it in a conscientious manner[103].

The legitimisation of the processing of personal data in the framework of a Data Space can be based on any of the legal bases in Article 6 of the GDPR, including compliance with

[101] It could be the case that the single information point should have a list of assets containing a summary of all available data resources and including, where appropriate, those data resources that are available at sectoral, regional or local information points, together with relevant information describing the available data. (Recital 26 of the DGA) This processing could be intended to be implemented actively by the single information point exploring the datasets, or passively by only receiving descriptions in which case we would not be in the framework of this example.

[102] Article 1(3) and recital 4 of the DGA.

[103] Paragraph 83 of the document 'Joint EDPB-EDPS Opinion 3/2021 on the Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) [10 March 2021]'.

legal obligations [104] [105] [106], or legitimate interest where processing is not carried out by public administrations in the exercise of their functions. The point is that the legitimisation is clearly and correctly defined in the processing.

Closely linked to the lawfulness of processing is the purpose limitation principle. The boundaries of what constitutes lawful processing and further compatible processing of data should be very clear to all parties concerned. In the case of compatible processing, processing in the context of the Data Space must meet the requirements of Article 5(1)(b) of the GDPR (purpose limitation) and Article 6(4), of the GDPR (compatibility test). Where further processing is to be carried out, the controller must first ensure that such processing is compatible with the original purpose and design it accordingly. The compatibility or incompatibility of a new purpose shall be assessed in accordance with the criteria set out in Article 6(4)[107].

In addition, where the purposes are archiving in the public interest, scientific and historical research purposes or statistical purposes, it must be in accordance with the Article 89(1) of the GDPR (safeguards and exceptions relating to processing for scientific purposes), read in the light of Article 50 of the GDPR. Opinion 3/2013 of the Article 29 Working Party provides a useful guidance on the implementation of the purpose limitation principle as well as on the appropriate use of the various legal bases for processing personal data and remains largely relevant also under GDPR[108].

Where the lawfulness of the processing has been based on consent, if further processing on the basis of a compatibility test pursuant to Article 6(4) of the GDPR is considered possible, the very principle of consent requirements would be circumvented[109]. Therefore, where processing has been based on consent, further processing may only take place if the controller requests a specific consent for that other distinct purpose or if the controller can demonstrate that it relies on a Union or member State law to safeguard the purposes referred to in Article 23 of the GDPR.

Article 5.1.b of the GDPR provides that further processing of personal data for archiving purposes in the public interest, scientific and historical research purposes or statistical purposes shall not be considered incompatible with the initial purposes. This does not mean that these purposes are always considered compatible, but rather that the starting point of the analysis is the possibility of compatibility. '*In order to ascertain whether a purpose of further processing is compatible with the purpose for which the personal data are initially collected, the controller, after having met all the requirements for the lawfulness of the original processing, should take into account, inter alia: any link between those purposes*

[104] Such as anonymisation for sharing high-value data or HVDs, see Recital 8 of Implementing Regulation (EU) 2023/13.
[105] In the case of the EHDS proposal, Recital 37 states that 'More specifically: for processing of electronic health data held by the data holder pursuant to this Regulation, this Regulation creates the legal obligation in the sense of Article 6(1)-point c) of Regulation (EU) 2016/679 for disclosing the data by the data holder to health data access bodies … This Regulation also meets the conditions for such processing pursuant to Articles 9, (2), (h), (i), (j) of the Regulation (EU) 2016/679'.
[106] Also in the EHDS proposal it is stated in the Article 33 'Minimum categories with Regard to the Secondary Use of Electronic Health Data' the obligation for Data Holders to make certain categories of electronic data available for a secondary use.
[107] Paragraph 71 of the document 'Guidelines 4/2019 on Article 25 Data protection by design and by default (EDPB) [20 October 2020]"'
[108] Paragraph 18 of the document 'Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]"'
[109] Paragraph 53 of the document Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications (EDPB) [9 March de 2021]'

and the purposes of the intended further processing; the context in which the personal data have been collected, in particular the reasonable expectations of data subjects based on their relationship with the controller as to their further use, the nature of the personal data; the consequences of the intended further processing for data subjects; and the existence of appropriate safeguards in both the original and intended further processing operations'[110]. Controllers should also be careful not to extend the boundaries of 'compatible purposes' of Article 6(4), and have in mind what processing will be within the reasonable expectations of Data subjects[111].

The existence of a legal basis does not excuse compliance with all the principles, rights and obligations established in the GDPR. In particular, the proactive accountability-based compliance model set out in the GDPR requires more than the choice of a legal basis for processing on the basis of the categories in Article 6:

- On the one hand, in the case of processing of special categories of data, it is necessary to demonstrate that the conditions for lifting the prohibition of such processing set out in Article 9(2) of the GDPR and the provisions of Article 9 of the LOPDGDD[112] are met.
- Article 8 of the LOPDGDD establishes that the processing of personal data due to legal obligation (Article 6(1)© GDPR), public interest or exercise of public powers (Article 6(1)€ GDPR), as well as the specialities of processing subject to Organic Law 7/2021, may only be considered justified when so provided for or deriving from a competence conferred by a rule of European Union law or a rule with the status of law and the appropriate safeguards are established.
- The lifting of the prohibition[113] on processing special categories of data referred to in Article 9(2)(g) (essential public interest), (h) (purposes of preventive or occupational medicine, assessment of the worker's capacity to work, medical diagnosis, provision of health or social care or treatment, or management of health and social care systems and services) and (i) (public interest in the field of public health) of the GDPR must be covered by a regulation having the force of law and appropriate safeguards must be put in place.
- The GDPR requires the explicit assessment of necessity, which also involves an analysis of the appropriateness, of the processing on those entitled under Article 6(1)(b) to (f), and on the lifting of the prohibitions based on Article 9(2)(b), (c), (f) and (g).
- The GDPR requires an assessment of the proportionality of the processing in those authenticated by a legal obligation (Article 6(1)(c) GDPR), public interest or exercise of public authority (Article 6(1)(e) GDPR), for the lifting of the prohibitions to process special categories of data by Articles 9(2)(g) (essential

[110] Recital 50 of the GDPR
[111] Paragraph 51 of the 'Guideline 4/2019 on Article 25 Data Protection by Design and Default (EDPB) [20 October 2020]'
[112] The EHDS proposal, in Recital 37 states: 'This Regulation provides the legal basis in accordance with Articles 9(2) (g), (h), (i) and (j) of Regulation (EU) 2016/679 fort the secondary use of health data, establishing the safeguards for processing, in terms of lawful purposes, trusted governance for providing access to health data (through health data access bodies) and processing in a secure environment, as well as modalities for data processing, set out in the data permit'.
[113] The lifting of the prohibition does not imply the existence of a legitimate basis, but it will be necessary to address the analysis required by Article 6 and, where appropriate, Article 7 to determine the legitimate basis.

public interest) and 9(2)(j) (archiving purposes in the public interest, scientific or historical research purposes or statistical purposes).

- For any high-risk processing, a DPIA is necessary to manage that risk and to pass the assessment of appropriateness, necessity and strict proportionality.
- Where the processing involves decisions based solely on automated processing, including profiling, which produces legal effects on him or her or significantly affects him or her in a similar way, the conditions enabling the processing pursuant to Article 22 of the GDPR must be met.

With regard to anonymisation, it should be noted that it is a processing of personal data, and like all processing, it must comply with the same requirements as outlined above[114].

Finally, there are other limitations to the processing of personal data that do not arise from the GDPR. For example, a Data Intermediation Service, to be considered as such under the DGA[115], may not use the data in relation to those providing its services for purposes other than making it available to Data Users and shall provide the mediation services through a legal entity that is independent from the other activities of the provider of such services. Such services may also not perform format conversions of personal data[116] unless a number of conditions are met and Data Subjects are offered an exclusion possibility[117]. The DGA also limits the possibilities of data processing by data altruism organisations that have voluntarily decided to apply for registration in the relevant national register, in the sense that they may not use the data for purposes other than those of general interest for which the Data Subject or Data Holder permits the processing.[118]

D. DETERMINATION OF PROCESSING RESPONSIBILITIES

From a data protection perspective, the most important part of the governance structure of a Data Space is the clear establishment of the roles of controllers and processors/sub-processors when processing personal data[119], which should be defined from the design. In addition, these roles must be established in accordance with the regulations and guidelines set by the supervisory authorities[120].

The status of controller or processor shall be attributed to an entity in relation to a processing operation, depending on the decision-making on purposes and means, so that the same entity may be a controller for some processing operations and a processor for others. An entity shall be a processor when it is processing data on behalf of a controller. If it is processing without a controller, or has not been chosen by the controller in accordance with Article 28.1 of the GDPR, or there is no contract or legal act binding it to the controller, it infringes the GDPR by setting purposes and means[121], or whoever in its own name and without being shown to be acting on behalf of another, establishes

[114] WP 216 Opinion 05/2014 on anonymisation techniques (Article 29 Data Protection Working Party) [10 April 2014]
[115] Article 12(a) and Recital 33 of the DGA
[116] The conversion of the format of personal data is or may also be part of a processing operation.
[117] Article 12(d) of the DGA
[118] Article 21(2) of the DGA
[119] Paragraph 39 of the document 'Joint EDPS-ECDC Opinion 3/2021 on the Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) [10 March 2021]'.
[120] WP 169 Opinion 1/2010 on the concepts of 'controller' and 'processor' (Article 29 Working Party) [16 February 2010].
[121] Article 28(10) of the GDPR.

relations with the data subjects[122], or whoever, appearing as a processor, uses the data for its own purposes[123], then it is a data controller.

There may be more than one controller for a processing operation, and the communication of data between two controllers should not be confused with the existence of controller-processor relationship or co-responsibility.

When you are a controller, you have the obligation to ensure and be able to demonstrate compliance with data protection law. When you have the status of a processor/sub-processor, you will have to comply with the obligations set out for the processor in data protection law, in particular, Article 28(3) of the GDPR. For example, it is the controller that has the obligation to ensure that the data subject obtains the exercise of his or her rights, but not the processor, although the processor has the obligation to assist the controller in accordance with the nature of the processing.

Where roles are assigned by regulation, this assignment should also be done in compliance with the provisions of the GDPR, in particular, that the controller can exercise over the processor the obligations set out in Article 28 of the GDPR, such as the duty of care in Article 28(1), control over sub-processors in Article 28(2), and the stipulations in Article 28(3)[124]. In case this is not possible, it is recommended that the relationship should be shaped as communications of data between controllers.

Next, an analysis of the responsibilities for processing in the Data Space will be made without prejudice to the applicable special rules and the legal basis for the processing.

Data Holder

Where there is processing of personal data, those who are considered data controllers may be considered Data Holders. Processors may only act as Data Holders when they have been specifically instructed to do so by the controllers[125]. In this case, the processor shall be limited to complying with the instructions documented by the data controller in relation to the processing of personal data contained in the contract or legal act binding them[126].

Processing carried out by Data Holders within the framework of a Data Space, whether for their own purposes, e.g., for altruistic purposes, to participate in an initiative, or because of a legal obligation, would be considered data controllers.

It should be noted that the GDPR does not apply if the information about Data Subjects that is communicated to the Data Space is non-personal data. However, if mechanisms had to be used to process personal data, e.g., to generate anonymised data sets, this in itself constitutes processing of personal data with regard to the Data Holder.

[122] Article 33(2) of the LOPDGDD which shall not apply to processing orders carried out within the framework of public sector procurement legislation.
[123] Article 33(2) second paragraph of the LOPDGDD.
[124] In this regard, it is noteworthy that in Article 12.7 of the current draft of the EHDS, the Commission is given the role of being processor.
[125] The DA Proposal expresses itself in this sense in Recital 21 'Data processors as defined in Regulation (EU) 2016/679 are by default not considered to act as data holders, unless specifically tasked by the data controller.'
[126] Article 28(3).a of the GDPR

In addition, the data under the responsibility of the Data Holder is likely to be stored in a way that requires transformation of the data both in its format, content, metadata and even file format, and may be entrusted to a Data Space Mediator[127]. The transformation of the data also involves a processing operation.

Data Space Mediator

Data Space Mediators, where they do not act on behalf of a controller but, by defining purposes and means, process personal data, for example, for storage, transformation into anonymised information, generation of data catalogues processing personal data, providing identity services to natural persons, transferring responsible data outside the framework of a controller-processor relationship or others, shall be data controllers.

Mediators may be processors in the context of a processing operation where they act on behalf of a controller, for example, when a Data User enters into a contract, or other legal act, with the Mediator to process personal data for which those entities are responsible.

In the case of re-use of data from public sector bodies, the DGA states that the Competent Bodies must act in accordance with the instructions received from the public sector body, so that any processing it may carry out must be completed under the responsibility of the public sector body in charge of the register containing the data, which remains the data controller as defined in the GDPR to the extent that personal data are concerned[128].

Data User

To the extent that the Data User defines the purposes of the processing and the means (which in this case will be to use the Data Space infrastructure) for the processing of personal data, the Data User will be responsible for such processing and must therefore comply with all obligations arising from the data protection regulation.

As soon as the Data User determines to implement all or part of the processing through the Data Space, the measures taken to ensure compliance with the GDPR must be coordinated, where appropriate, with other interveners such as Data Space Mediators, Data Holders or Enablers.

Enablers

Where they do not process personal data, they may not have a role under the GDPR.

Where they process personal data for their own purposes, they are controllers. They will be processors (or sub-processors) where they process personal data on behalf of controllers or other processors in compliance with the requirements of Article 28 of the GDPR.

[127] Corresponds to article 12(d) of the DGA in cases where private data are processed in a non-altruistic transfer.
[128] Recital 26 of the DGA

V. PRIVACY BY DESIGN IN A DATA SPACE

The implementation of the principles, rights and obligations of the GDPR requires data controllers to take appropriate measures to achieve this[129]. These measures could be legal, organisational and also technical tools. With regard to the latter, and with respect to the free movement of personal data, the GDPR states in its initial recitals that technology must facilitate solutions that implement a high level of protection of personal data[130]. In any case, the adequacy of the measures will be established according to the context, nature, scope, purposes and risks to the rights of the Data Subjects involved in the processing.

The principles of minimisation and data protection by design and by default are essential where processing involves significant risks to the fundamental rights of individuals. Taking into account the latest technical developments, all parties involved in data sharing should implement technical and organisational measures to protect these rights. Such measures include not only anonymisation, pseudonymisation and encryption. These techniques are neither the only ones, nor in many cases may be the most appropriate. There are other technologies that are increasingly used that allow algorithms to be introduced into data and valuable information to be obtained without the need for transmission between parties or superfluous copying of raw or structured data[131]. Examples of such techniques are differential privacy, generalisation, suppression and randomisation[132], the use of synthetic data, federated learning, Secure Processing Environments and other privacy enhancing tools and technologies (PET[133]). Member States should provide support to public sector bodies I order to make optimal use of these techniques and thus make as much data as possible available for sharing[134].

In the introduction we have highlighted that Data Spaces, as technical and governance infrastructure, should be characterised by allowing the above listed measures, among others, to be adopted in order to implement the necessary data protection guarantees that allow for the free flow of personal data within the Union.

A. POSSIBLE CONFIGURATIONS OF A DATA SPACE

A Data Space may have different configurations[135]. This section provides a set of examples of Data Space configurations, without being exhaustive. These configurations could be combined with each other and, for the purpose of didactics, Data Holders and Data Users are identified as distinct entities, although as noted in section IV.A, both roles could be shared.

On the one hand, a Data Space could be configured in such a way that a single entity establishes all the mediation and supervision functions of the Data Space, with the possible use of processors/sub-processors. This Mediator could centralise all operations

[129] Article 24 of the GDPR
[130] Recital 6 of the GDPR
[131] Recital 8 of the DA proposal
[132] Recital 7 of the DGA
[133] Privacy Enhancing Technologies
[134] Recital 7 of the DGA
[135] Escenarios de compartición de datos. Francisco Javier Esteve Pradera June 2022 – Bulletin nº 91 [only available in Spanish]

of access to metadata, their data and the resources for their exploitation between Data Holders and Users:

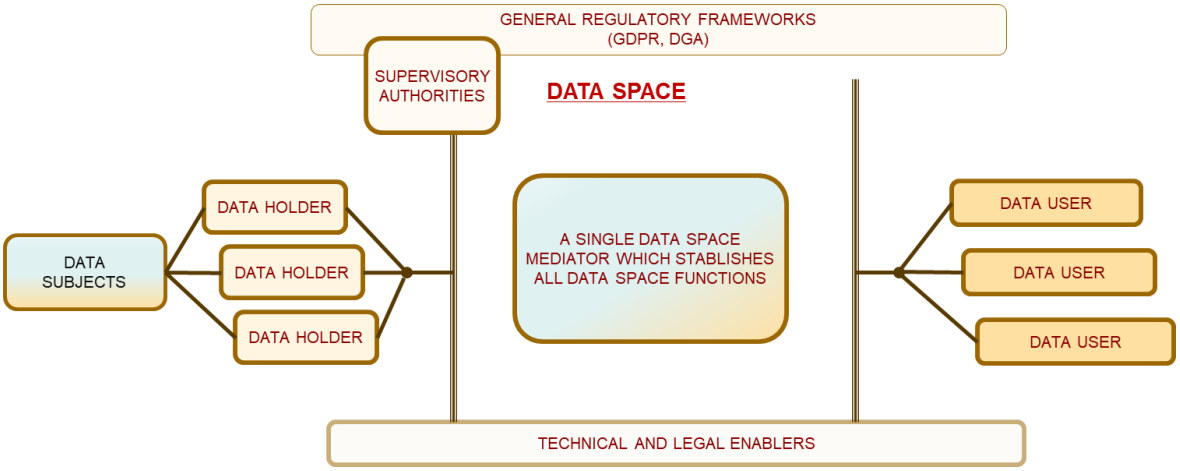

Figure 3: Configuration of a Data Space based on sharing via a central node

In a Data Space such a central Data Space Mediator could be limited to managing the participants, the data catalogue and the security mechanisms, among other services, while Holders and Users can access peer-to-peer data:

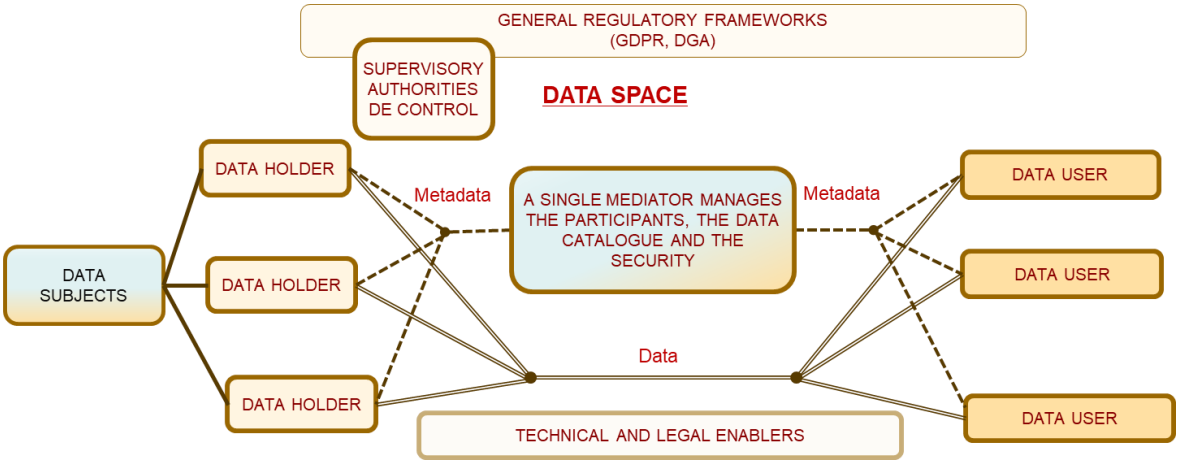

Figure 4: Configuration on the basis of a Data Space Mediator as a central hub or data marketplace

A Data Space could have a more complex configuration, involving multiple entities that will set up multiple functions, some of them repeatedly, implementing different offerings of the same service:

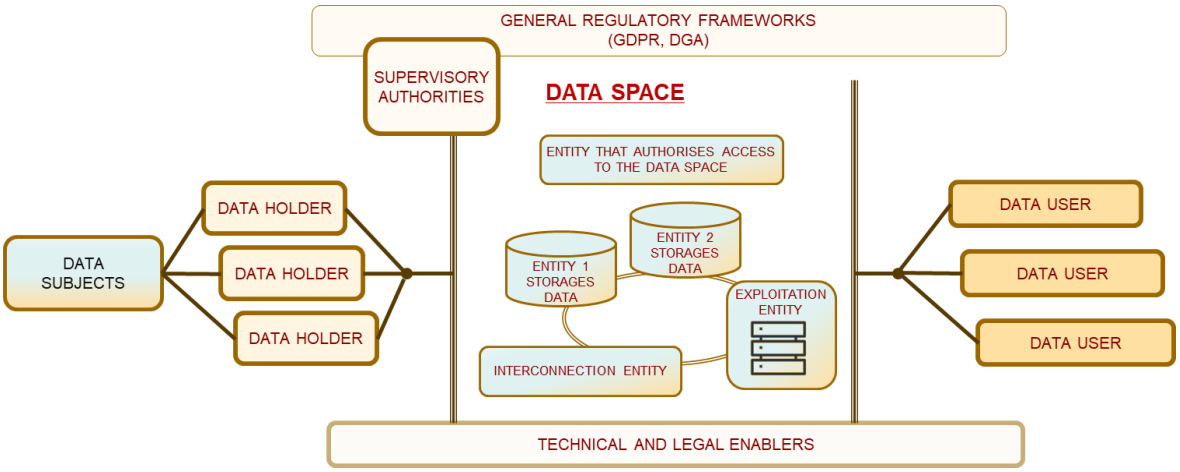

Figure 5: Complex configuration in the definition of interveners in a Data Space

On the other hand, the Data Space could be established to support direct interaction between Data Subjects and Data Users[136], such as data altruism organisations[137]:

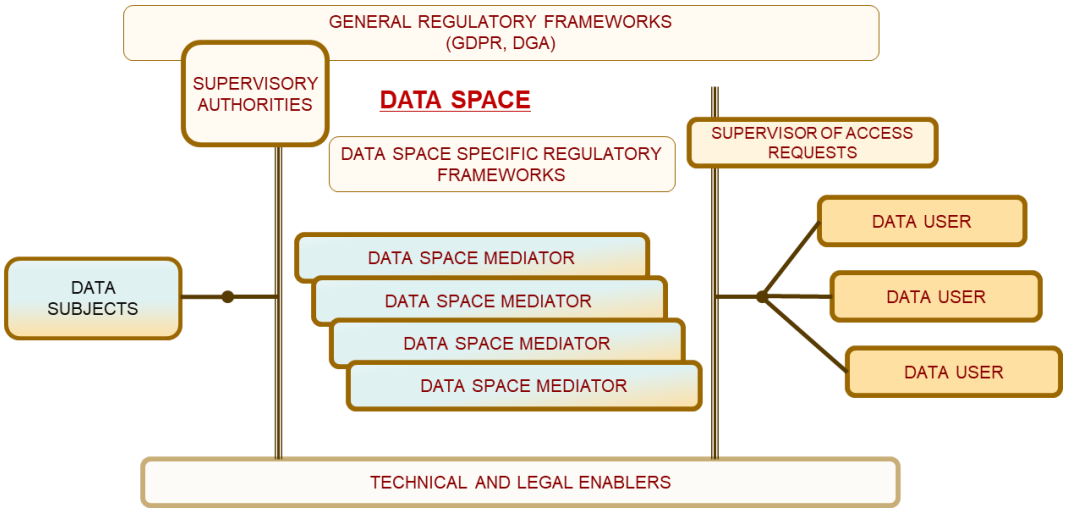

Figure 6: Complex configuration in the definition of interveners in a Data Space

The Data Space may also be configured to allow access to Data Holders data by Data User(s) without the use of Mediators.

[136] Article 10(b) of the DGA
[137] Chapter IV of the DGA

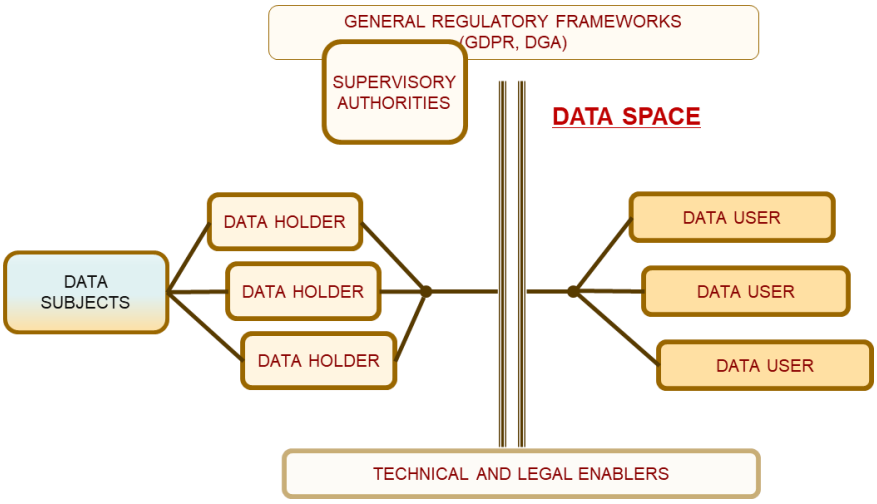

Figure 7: Data Space configuration without the use of data mediaton services

Finally, a Data Space could also be established as an infrastructure to facilitate data access agreements between Data Subjects and Data Holders as follows:

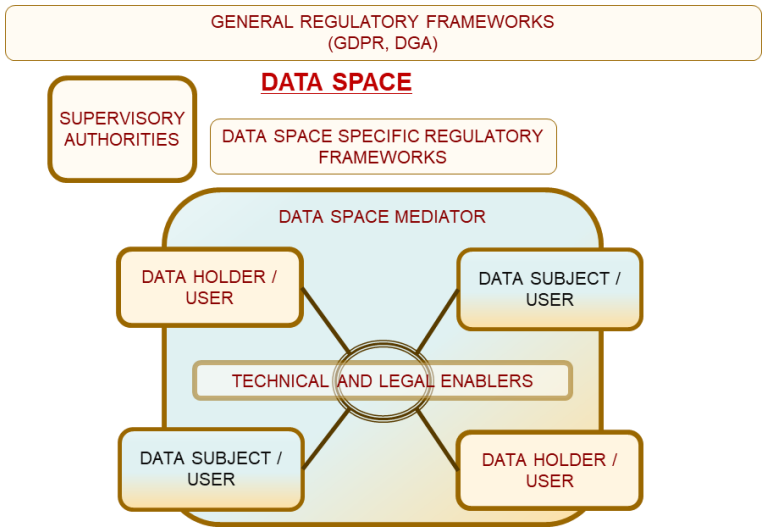

Figure 8: Configuration of a Data Space with data access agreements between Data Subjects and Data Holders

If the above approach is not limited to a closed group of interveners, but to an open group, it could fall under the legal definition of 'data cooperatives' [138].

Finally, there would be the participation of any of the above options in federations of Data Spaces. In the federation, a Holder/User would be able to consult the catalogues of its data Space, as well as those of other connected interoperable Data Spaces, and access or offer resources to all other interveners:

[138] Recital 31 of the DGA

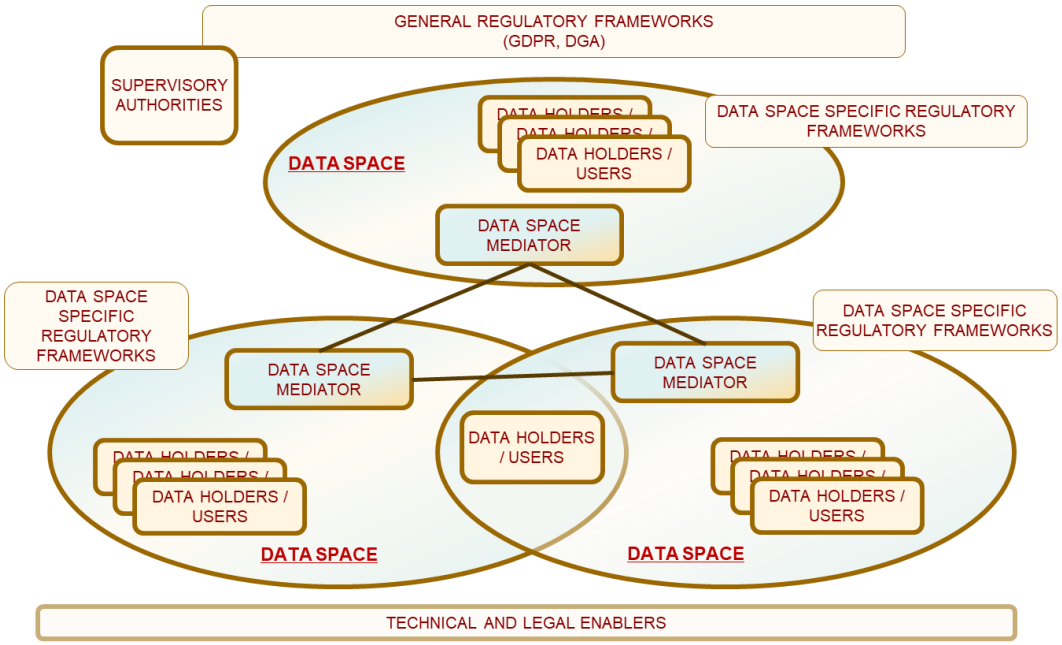

Figure 9: Federation of Data Spaces

In the following sections we will analyse a series of aspects to be taken into account in order to reach the subsection 'V.E Uses cases and architectures for privacy response where we will include an analysis of architectures and data protection considerations.

B. ACCESS TO DATA AND INFORMATION

The DGA defines 'access' as any use of data in accordance with specific technical, legal or organisational requirements, without necessarily involving the transmission or downloading of data.[139] In other words, in the framework of Data Spaces, a distinction is made between two concepts that may seem similar but are very different:

- Access to data by transmission or downloading.
- Access to information generated by the processing of data by means of access that does not involve either transmission or downloading of the data. Information is that which increases knowledge in a given context and is necessary and relevant to meet the Data User's objectives.

The data may not contain the necessary information for a given context and, in any case, prior processing of the data will be necessary to obtain the required information. A Data Space must enable the Data User to obtain the necessary information for each of the processing operations, and this does not necessarily imply the communication or dissemination of personal data. Therefore, a Data Space could (and from a data protection point of view is the most advisable) grant access to personal data, but without disseminating them, that is to say, not necessarily communicating the data to third parties. In fact, a Data Space architecture, which is built by applying data protection principles by design, will minimise the exposure of personal data (minimisation principle)

[139] Article 2(13) of the DGA

and preserve the ability to make available the information necessary to fulfil the purposes of the Data Space.

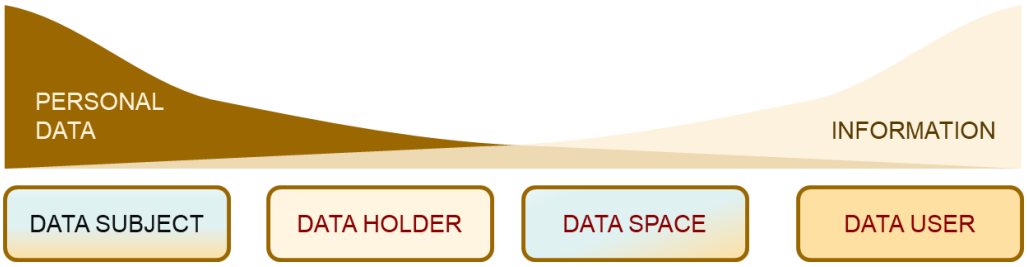

Figure 10: Evolution of data processing in a Data Space

The infrastructure of a Data Space must provide access to data, understood as the possibility of exploiting data to obtain value (information) without necessarily involving the communication of data, in this case personal data, between interveners.

Data Spaces that make information available but do not communicate or disseminate personal data, e.g., by leaving the actual control of the data and purposes in the hands of the Data Holders will increase the Data Holders' confidence[140] to participate in the Data Space and, furthermore, the Data Holders' willingness to engage in the development of the digital economy.

As an additional note, the european development of systems, solutions and maintenance of services that implement data protection by design is a driver for the digital economy, which are the general purposes justifying the creation of a Data Space. Moreover, if the luck of trust between the interveners involved does not allow the processing operations proposed in the framework of the Data Space to fulfil the stated purposes, such processing operations would not meet the criteria of adequacy and necessity.

The approach to data access set out in this section may in some cases be desirable, and in others mandatory, either because of the result of the risk management for rights and freedoms, or because it is required by law. For example, in the EHDS proposal, each mediator will have to set up a service to prevent[141] the Data User from making local copy or communicating personal data outside a Secure Processing Environment set up in the mediator, unless this is supported by a legal basis and is assessed and authorised by the Supervisor of access request on the basis of risk management.

C. TYPES OF DATA SETS

On the other hand, it should be borne in mind that the different use cases that arise in providing access to data, and therefore in a Data Space, may require different types of data to be handled:

- Structured (databases and others) and/or unstructured (documents, voice, image, etc.).

[140] The reluctance of entities to share information that may harm their commercial interests or reveal their business strategy has to be taken into account, apart from other regulation for the protection of intellectual and industrial property.
[141] In line with Article 5 and Recital 15 of the DGA for data held by public sector bodies and Article 50 and Recital 54 of the EHDS.

- In real time (IoT and similar in transport, health, supplies, smart cities, etc.) or in deferred time (consolidated databases that are subsequently processed in batch).

And depending on the processing mode:

- Automated processing.
- Manual processing. It must be taken into account that in many processing activities for the development and evolution of machine learning systems it is necessary to use massive manual processing equipment for the labelling of samples, especially when processing unstructured data[142].

In turn, depending on the flow of information and knowledge, there could be two cases:

- In the direction only from Data Holders to Data Users.
- Also, in the opposite direction: from Data Users to Data Subjects.

The latter could be the case in clinical research[143], where a result has to be communicated to a specific patient. It would not be the case when general feedback could be given to a group.

D. DATA SPACE ARCHITECTURES AND USE CASES

The basic and most immediate architecture of a Data Space involves collecting data from multiple Data Holders, concentrating the data at a single point and giving Data Users access to the data.

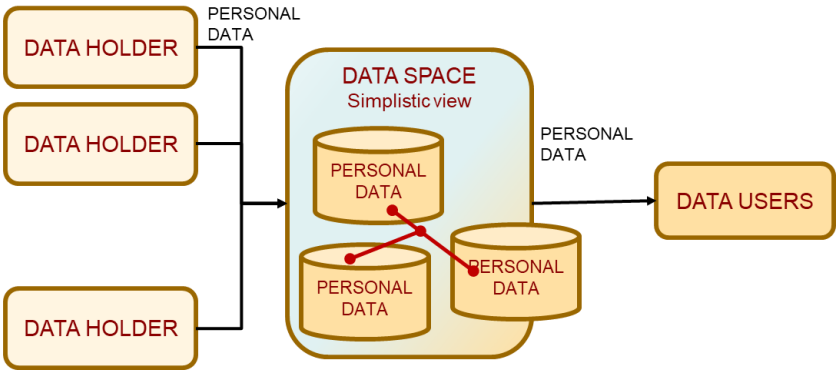

Figure 11: Basic architecture diagram of a Data Space

To the extent that the implementation of processing in a Data Space implies the need to access massive storage and processing resources, Data Users might lack them (such as SMEs and some research groups), would in many cases have to make use of cloud computing services.

[142] Mechanical Turk at Amazon, or voice tagging at Sigma.
[143] Recital 44 of the EHDS.

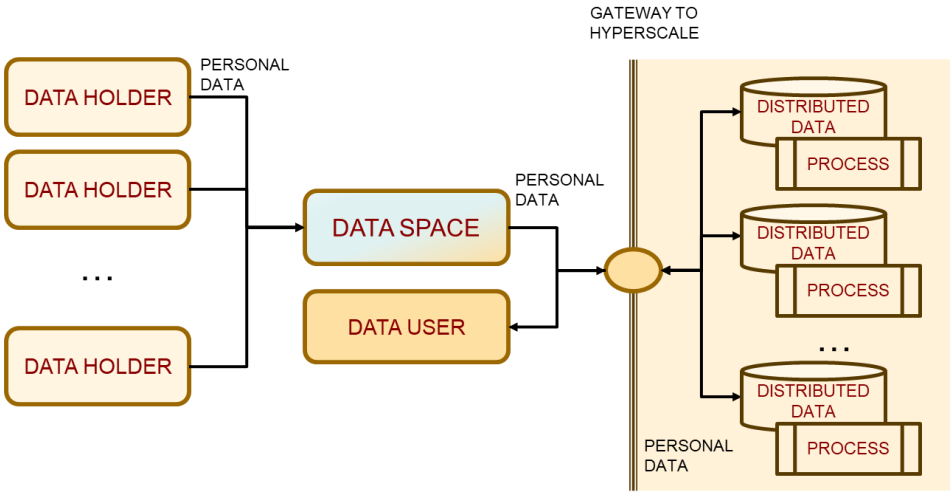

Figure 12: Basic architecture diagram of a Data Space using hyperscale

These architectures would involve shifting regulatory compliance issues, preservation of principles on rights and freedoms and other collateral effects, to those less-resourced Data Users. This could make it difficult to achieve the purposes of the processing, such as lack of trust of Data Subjects, reluctance to share data by the Data Holders themselves to protect interests or business secrets, etc.

There are different approaches to a Data Space that could solve these issues in advance. One of these possible solutions could be achieved by employing the same technologies used by hyperscales for distributed processing but, in this case, to implement compute-to-data techniques[144], i.e., distributed data processing would be executed at the data source. This would avoid the communication of data to third parties, the processing would be performed at the data source, and the exposure of personal data in communication networks and the accumulation of data in large repositories would be reduced.

[144] As it appears in the definitions, it implies that the processing is done at the source of data, rather than the data being communicated to a 'cloud' where the processing is done. It is also related to Edge-Computing strategies, which is one of the characteristics of 5G, and which involves bringing data processing closer to the end users' own systems, with the advantage that the network traffic is offloaded, and service providers need fewer servers, as they use the capacity of the users' terminals.

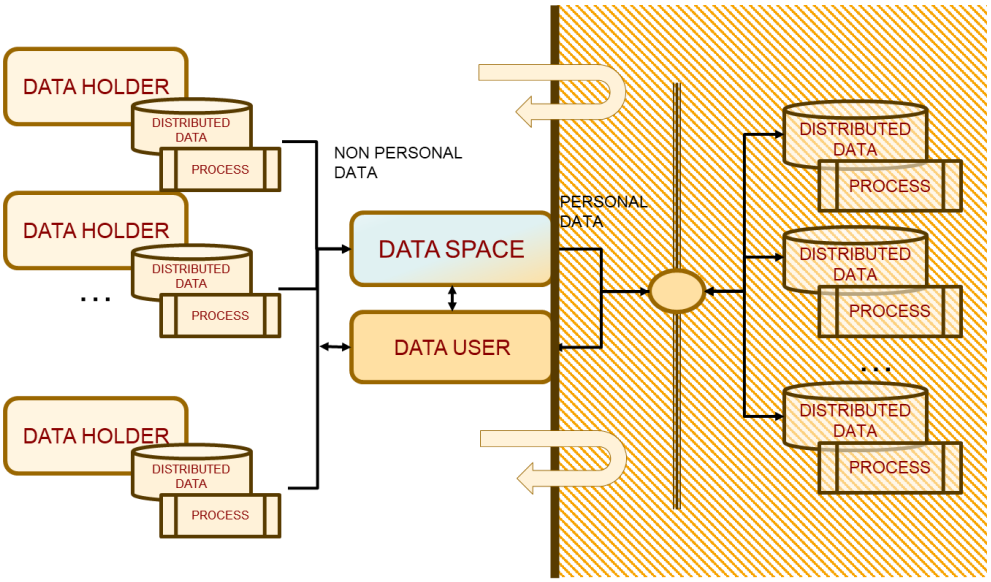

Figure 13: Basic architecture diagram using the compute-to-data strategy

Like any data protection strategy by design, such as anonymisation or differential privacy among others, this would not be the only solution for the regulatory and purpose compliance of a Data Space in case of personal data processing. Nor is it the most appropriate solution for all possible use cases, nor is it suitable for all data processing scenarios. Different strategies should always be considered depending on the specific use case and the design of the Data Space should allow for them. For this to be possible, it is necessary to take data protection into account from the design of the Data Space.

E. USES CASES AND ARCHITECTURES FOR PRIVACY RESPONSE

Within the framework of a Data Space, different types of personal data processing can be set out, i.e., different use cases[145] in which specific strategies can be adopted to implement data protection by design for accesses between Data Holders and Data Space Mediators:

1. The use case requiring the processing of non-personal information in cases other than the anonymisation of personal data.
2. The use case where the processing can be executed without transferring personal data by the Data Holder, but by processing them in its own systems and providing information that does not constitute personal data to the Data Mediator (aggregated, processed or other). In other words, by implementing compute-to-data strategies.
3. The use case that can be fulfilled by the Data Holder communicating anonymised information to the Data Space Mediator.
4. The use case that can be fulfilled by the Data Holder communicating pseudonymised information to the Data Space Mediator.

[145] In Data Spaces where there is no processing of personal data, e.g., only data from industrial environments, this use case development would not apply.

5. The use case that can only be fulfilled by making communication of personal data from Data Holders to a Data Space Mediator.

The above cases describe the relationships between Data Holders and Data Mediators. They could also apply where a Data User wants to access information from multiple Data Holders[146].

Only in cases 4 and 5 personal information is communicated to the Mediator, in case 4 pseudonymised and in case 5 without being pseudonymised. It could be envisaged that, in these last two cases, the following sub-cases could be deployed between the Data Space Mediator and the Data User:

1. The subcase of use where the processing can be executed without transferring personal data by the Data Space Mediator, but by processing them in its own systems and providing information that does not constitute personal data to the Data User (aggregated, processed or other). In other words, by implementing secure data processing environments.
2. The subcase of use that it can be fulfilled by the Data Space Mediator communicating anonymised information to the Data User.
3. The subcase of use that it can be fulfilled by the Data Space Mediator communicating pseudonymised information to the Data User.
4. The subcase of use that it can only be fulfilled by communicating personal data from the Data Mediator to a Data User.

In order to implement the listed cases/subcases, different data processing architectures can be proposed in the Data Space. At this point, the factor to be highlighted is that the Data Space must be defined from the design stage to allow the implementation of those architectures that are considered to have an acceptable risk for the rights and freedoms of the data subjects and society in accordance with the specific processing operations.

Next, architectures that could be used to respond to the use cases from a data protection point of view. will be developed. The purpose of this section is not to provide a complete and exhaustive list, nor a full analysis of the obligations and implications involved, but rather to provide a number of examples. The aim of the following examples is to exemplify some of the possibilities that exist to implement access, as defined in the DGA, minimising the transmission or dissemination of personal data.

Processing of non-personal data

Where the processing in the Data Space does not involve any processing of personal data whether by the Data Holder, by Data Mediators, by the Data User or any other, this architecture guidance would not apply.

There will be two sources of non-personal data: data sets generated by anonymisation processes and data sets that are non-personal at source, such as, in principle, the geolocation database of mobile phone masts. However, it would require diligence on the

[146] By way of clarification, and with reference to the provisions of Recital 28 and Article 2 of the DGA, it is noted that a direct request by a Data User to a single Data Holder is not part of what it is considered here as Data Space and will not be considered as a use case.

part of the Data Mediator or, as the case may be, the Data User and Data Holders to make an assessment of whether a reidentification of natural persons is possible due to the massive accumulation of data from various sources and the use of novel technologies.

With regard to anonymised data, it should be borne in mind that the larger the volume of data received, the greater the chances of reidentification despite the fact that the data are non-personal, especially when data are received from a variety of sources[147]. These sources include both anonymised data and originally non-personal data. Therefore, among the functions of Data Space Mediators, it is of particular relevance to perform a first review of the strength of the anonymisation processing of a data set before making it available to the Data User.

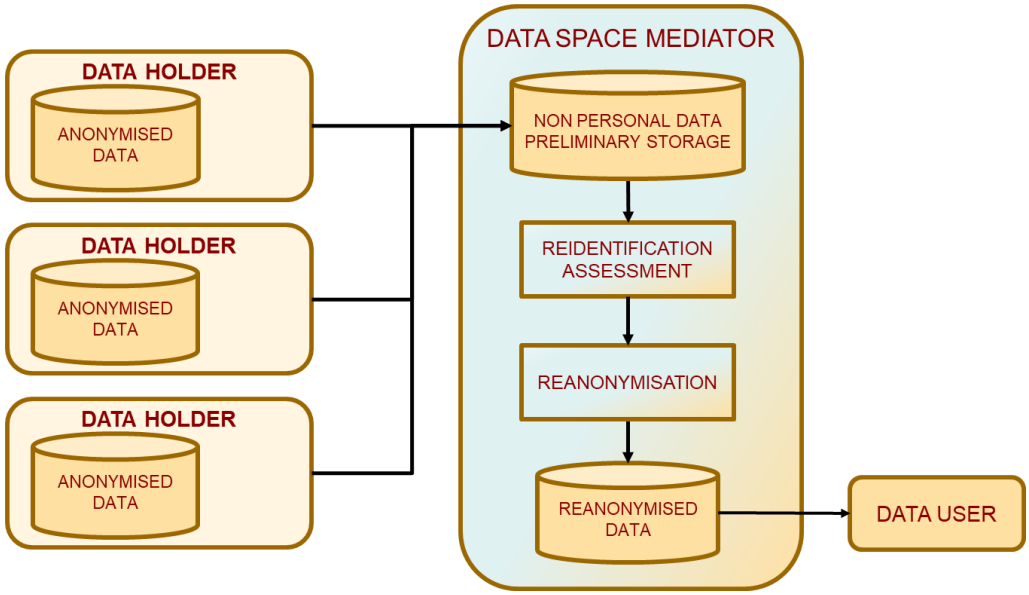

Figure 14: Diagram of the architecture for the use case for the consolidation of non-personal data from different Data Holders

In cases where the possibility of reidentification is detected, anonymisation (or reanonymisation) mechanisms will have to be re-implemented in the Data Space Mediator, in a secure environment, where all data that allowed for reidentification will be deleted.

In the case of non-personal data sets, the DPO of the Data Space Mediator should know which data sets are available at any given time.

The measures that could be implemented to ensure and be able to demonstrate compliance with data protection regulation in this use case could be legal, organisational and technical. For example, legal measures could include in the data communication agreements the obligation of the Data Mediator to carry out the re-anonymisation processing.

Organisational and data protection policy measures could include that the assessment of anonymisation is coordinated between the Data Space Mediators and, especially, the Supervisor of the requests. Other measures could be that the DPO is informed of the

[147] Recital 15 of the DGA and Recital 64 of the EDHS proposal

anonymised and non-personal data sets collected in their systems, that there is a reidentification risk assessment when a new data set is added to the entity, that there is an internal process to check the state of the art and reidentification events and developments, and also that there is a process for deletion of data that is re-identifiable once the need for it has been extinguished, among others.

Technical measures could include, for example, a systematic assessment in a controlled and secure environment of the anonymity of the overall results stored in the Data Space on a temporary basis and prior to making them available to the Data User.

Compute to data strategies and federated learning

An example of what compute-to-data strategies are was shown at the beginning of this chapter. These strategies involve taking the data processing to the original source of the data, in this case the Data Holder, in order to extract the (no longer personal) information from the Data Holder. For example, compute-to-data strategies can be used for the implementation of federated learning in machine learning-based artificial intelligence training.

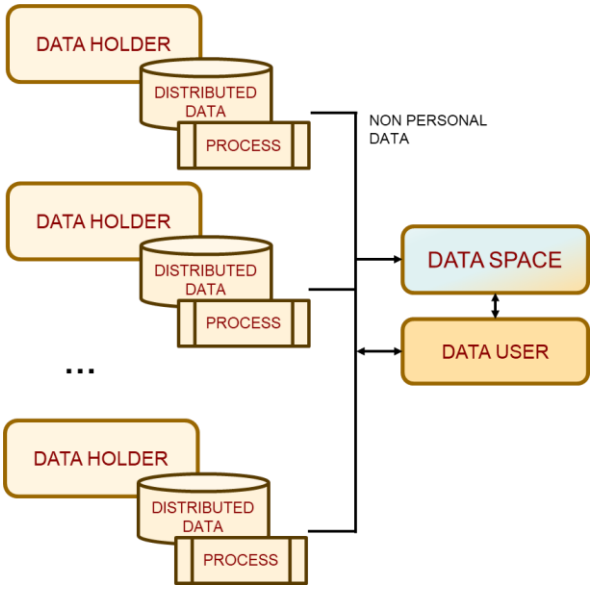

Figure 15: Diagram of the architecture using compute-to-data strategies

The implementation of these strategies implies that the Data Space has defined governance and information management obligations in a distributed environment. In some cases, the processes that the Data User intends to run on the Data Holders' premises will have to be audited or certified before being distributed to the latter.

In these cases, Data Holders should have ad-hoc spaces in their infrastructures, with complete separation from their operating systems (similar to a DMZ zone[148]).

[148] DMZ refers to demilitarised zone or secure area not connected to the entity's operating systems.

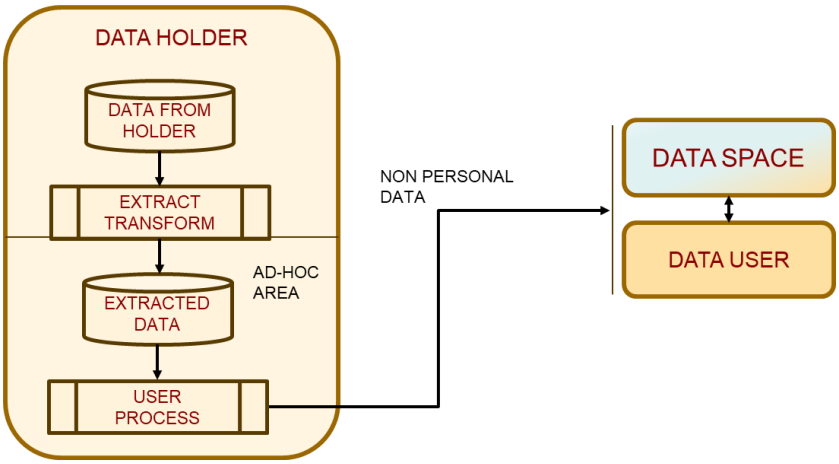

Figure 16: Diagram of specific spaces in the Data Holders to enable the compute-to-data infrastructure

This use case could be developed in multiple cases depending on the specific processing involved.

The measures that could be implemented to ensure and be able to demonstrate compliance with data protection law in this use case could be legal, organisational and technical. For example, legal measures could include obligations for prior auditing or certification of Data Users' processes.

Organisational and data protection policies measures, could include human supervision of the download and execution of processes at the Data Holder, non-execution of User processes on the data and operating systems of the Data Holder, extraction of a working data set that does not contain the entire data set, assessment that the User process does not generate and communicate personal data, or assessment of the results obtained, among others.

Technical measures could include, for example, the establishment of ad-hoc areas for the execution of the processes with physical isolation from the systems of the Data Holder, ad-hoc areas that form Secure Processing Environments, etc.

A case of compute-to-data: Cataloguing

Cataloguing is a processing of data or a set of data that allows the metadata necessary for further exploitation to be associated with the data. The metadata will at least include descriptions of the types of data and where the data are located, but may extend to assessing the quality of the data, which implies a thorough processing of the data, or even determining whether personal data exist in the dataset. In this way, resource (data) catalogues could be generated that can be made available to multiple interveners in a virtual, intermediated or real way.

The cataloguing of data sets is the first task to be performed in the framework of a Data Space. Cataloguing could be done in different ways: by Mediators and/or Users accessing the Holders' systems, by communicating the data to the Mediators and/or Users, or by a process at the Holders themselves, as a particular case of Compute-to-data.

In the latter case, cataloguing does not require the data sets held by the Data Holder to be transferred to third parties, but only the metadata resulting from the analysis needs to be communicated.

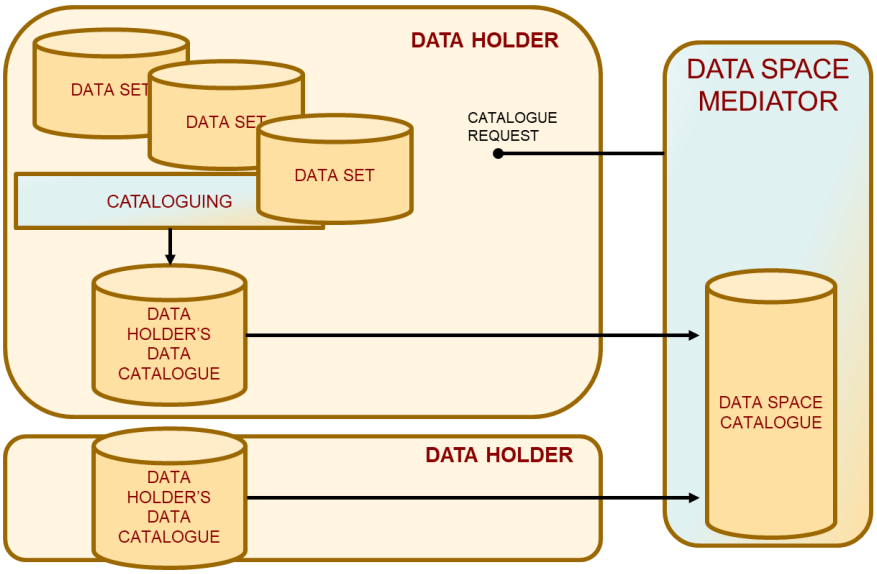

Figure 17: Diagram of the architecture for cataloguing

Measures that could be applied to ensure and be able to demonstrate compliance with data protection regulations will be of the same order as compute-to-data strategies.

Cataloguing and metadata generation processing can also be part of the implementation of data protection techniques by design in a Data Space, such as through tagging and tag hierarchies to be used for access privilege management.

Anonymisation: Processing that requires anonymised aggregated data of Data Holders with dissociation of data from different Data Holders

In this case, the Data Space Mediators or Data Users, either on their own initiative or in response to a User's request, require anonymised, or non-personal, information from Data Holders, and in such a way that the information to be extracted does not need to link data of the same subject that is stored between different Data Holders.

An example is mobility studies based on geolocation data from telecommunications operators, since a user is normally linked to a single telecommunications operator, and the case of users whose mobility profile depends on information from two operators is residual in order to fulfil the purpose of the processing.

In this case, the following architecture can be designed:

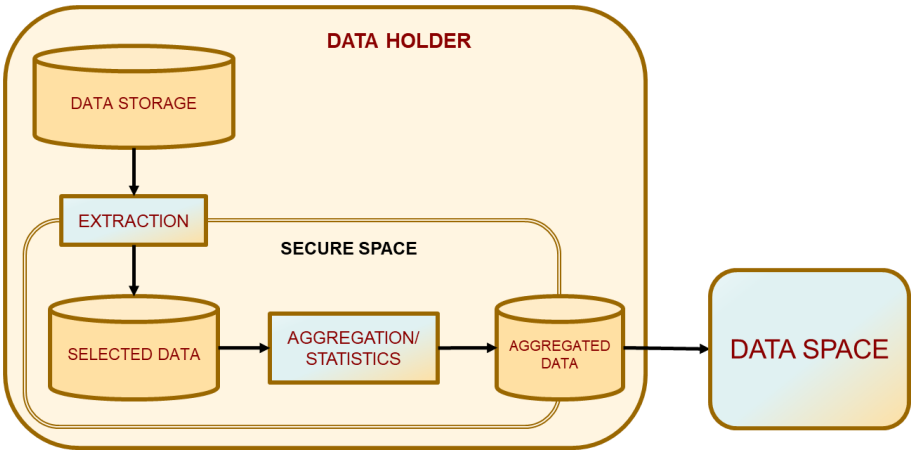

Figure 18: Diagram of the architecture for the use case of anonymised data without linkage between the data from different Data Holders

Measures that could be implemented to ensure and be able to demonstrate compliance with data protection regulations in this use case could be, for example, legal measures such as verifying that reidentification is not possible by means of a risk analysis.

Possible organisational and data protection policy measures could include, in relation to the aggregation process or other types of data analysis, among others, not carrying it out on the storage of operational data but on the already extracted data, following minimisation criteria, and carrying it out in an ad-hoc environment.

Among the technical measures that could be applied could be an assessment of the impossibility of reidentification in the resulting data.

Anonymisation: Processing that involves the consolidation of anonymised data from different Data Holders

This use case can occur in combination with other use cases shown in this guide, in particular with the use case of processing non-personal data.

The Data Space should take into account that the higher the volume of data received, the higher the chances of identification, despite the fact that the data is non-personal data, especially when data is received from different sources. Therefore, among the functions of the Data Space Mediators, it is of particular relevance to carry out a first review of the strength of the anonymisation processing of a dataset before making it available to the Data User.

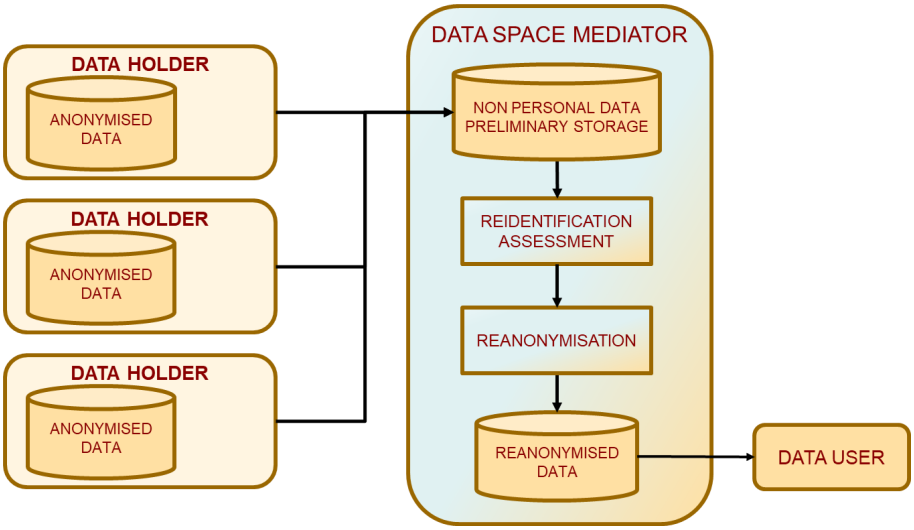

Figure 19: Diagram of the architecture for the use case for the consolidation of anonymised data from different Data Holders

In cases where the possibility of reidentification is detected, anonymisation mechanisms will have to be applied again in the Data Space, in a secure environment, where any reidentification keys will be removed, as well as all data from this first review.

Measures that could be applied to ensure and demonstrate compliance with data protection regulations in this use case could be, for example, legal, such as applying safeguards to limit the dissemination of anonymised data until they have been properly assessed for anonymisation or re-anonymised, and limiting the dissemination or retention of anonymised data by legal agreements beyond the provisions of the GDPR, among others.

Organisational measures and data protection policies could include, among others, the deletion of data sets that lead to reidentification and informing Data Holders of this possible vulnerability.

Technical measures could include the implementation of a controlled and secure environment for the temporary storage of anonymised data from the different sources to be linked and subsequently anonymised.

Anonymisation: Generation and use of synthetic data

Another data minimisation strategy is the use of synthetic data. Synthetic data are not random data, but data that meet the same requirements as real data for a specific purpose. The requirements will depend on the specific use case: a certain statistical distribution, fitting a certain type of patterns, etc. These patterns should be extracted from personal data by processing such personal data and generating non-personal information[149]. As soon as personal data are being used, the process of generating synthetic data is or will be part of a processing operation subject to compliance with the GDPR.

[149] An example of this use case is the pilot project for data sharing that is carried out at European level with data from the central banks of each country, where prior to making the data available, a synthetic database is generated that has the same characteristics as the original. It is a project led by the European Commission's DG for Financial Stability, Financial Services and Capital Markets Union as part of its project to create a Data Hub in the EU Digital Finance Platform.

The Data Holder may process the personal data in an ad-hoc environment for the analysis of the data, then generate the patterns derived from its stored data and, in this way, allow the Data Holder itself to develop the synthetic data, or simply release the patterns for a Data Space ecosystem intervener (maybe the Data User itself or an Enabler) to generate the synthetic datasets.

The following architecture is proposed for this case:

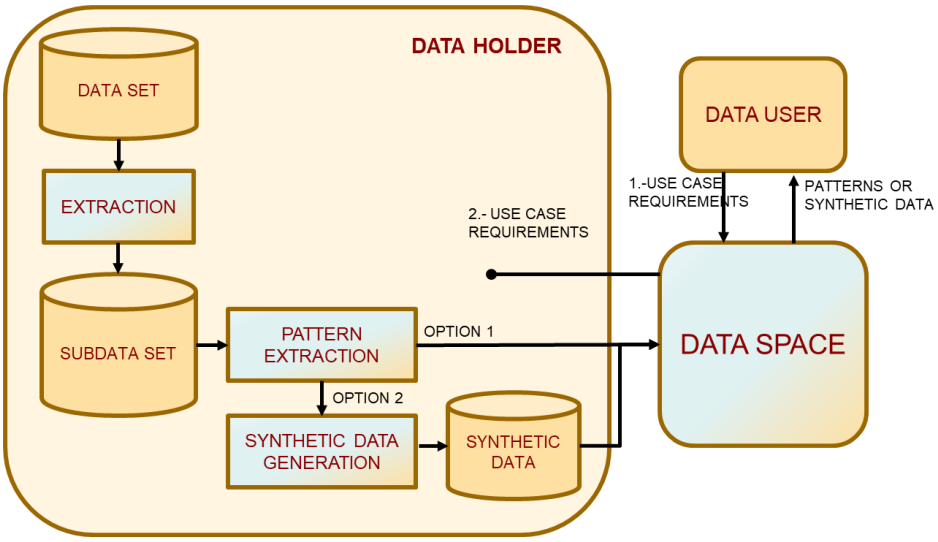

Figure 20: Diagram of the architecture for the synthetic data provision use case

Measures that could be implemented to ensure and be able to demonstrate compliance with data protection regulations in this use case could be legal, organisational and technical. For example, legal measures could include the obligation to create the synthetic data set at the Data Holder or the requirement of audits or certifications of the synthetic data generation tools.

Organisational and data protection policy measures could include the process of data analysis in a secure environment, with extraction of a subset of data from the operating systems, among others.

Technical measures could include, for example, assessing the anonymity of synthetic results.

Anonymisation: Secure Multiparty Computing

Secure Multiparty Computation[150] or SMPC. This is a cryptographic protocol that, by means of Additive Secret Sharing, allows a secret data to be segmented into different parts, so that, when the information is shared, the original data cannot be revealed by any of the sources. In the protocol, the desired result is obtained without the need to reveal any sensitive data, and the result obtained does not suffer any type of deviation.

This strategy is useful in certain scenarios and requires technological assistance to implement it.

[150] AEPD blog article entitled 'Privacy by Design: Secure Multi-Part Computation: Additive Sharing of Secrets | AEPD [May 2022]'

Anonymisation: Differential privacy

Differential privacy[151] guarantees, by incorporating random noise to the original information, that in the result of the analysis process of the data to which this technique has been applied, there is no loss in the utility of the results obtained. It is based on the Law of Large Numbers, a statistical principle that states that when the sample size grows, the average values derived from it approach the real mean value of the information. Thus, the addition of random noise to all the data compensates for these effects and produces an 'essentially equivalent' value.

One example of use can be found in the US Census Bureau[152], which applies differential privacy to ensure the accuracy of its statistics and prevent personal information from being disclosed even through the statistics, and thus increase citizens' confidence in the security of the data they provide.

Anonymisation: Anonymisation-oriented documents

Recital 9 of the DGA, in the case of re-use of data, states the need to develop data processing in which anonymisation is built into the concept of the data and in which data formats allow for efficient anonymisation 'by design': *'In order to facilitate the protection of personal data and confidential data and to speed up the process of making such data available for re-use under this Regulation, Member States should encourage public sector bodies to create and make available data in accordance with the principle of 'open by design and by default' referred to in Article 5(2) of Directive (EU) 2019/1024 and to promote the creation and the procurement of data in formats and structures that facilitate anonymisation in that regard.*'

Other techniques for safeguarding data protection

Without aiming to be exhaustive, there are other techniques used to safeguard data protection when sharing data. For example, homomorphic encryption, the recovery of private information, or the federated learning techniques in machine learning. The following is a brief overview of each of these techniques.

Homomorphic encryption[153] is a privacy-by-default technique that is suitable for cases where a controller outsources a part of an activity to a processor, and wants to technically ensure that the processor will not access the data.

In a traditional scheme, the data controller transmits the information to the processor in encrypted form, to protect confidentiality during transit. Once the processor has received it, it is decrypted and processed. However, this scheme presents both legal and technical risks, so ideally, to minimise the risks, the processor should not have the possibility to decrypt the information, and all processing should be carried out on the data encrypted by the data controller. This would prevent a disloyal processor or a third party from accessing the data and using it for different purposes. One way to achieve this protection is through the so-called homomorphic encryption.

[151] AEPD blog article entitled 'Anonymisation and pseudonymisation (II): Differential privacy | AEPD [October 2021]'
[152] Differential Privacy and the 2020 Census (census.gov)
[153] AEPD blog article entitled 'Encryption and Privacy III: Homomorphic encryption [June 2020]'

Homomorphic encryption therefore makes it possible to perform operations on encrypted data and obtain results, also encrypted, equivalent to the operations performed directly on the original information.

On the other hand, Private Information Retrieval (PIR) [154] is a cryptographic technique that allows the user to retrieve an entry from a database without revealing to the data custodian the item that has been retrieved and unlink the information that could be inferred regarding who is performing the access[155].

This can be taken to the example of a company that wants its customers to be able to access a database. In a default environment, each time a customer accesses the database, the data custodian knows which entry has been accessed. Over time, the controller will be able to know which entries in the database are of interest to customers. By implementing the private information retrieval technique, the data controller minimises the amount of information disclosed about the accessed data, as the PIR technique prevents the data controller from knowing which entries have been accessed.

Lastly, we can also highlight federated learning techniques[156], both horizontal and vertical, for artificial intelligence applications based on Machine Learning. Federated learning techniques are a category of PET (Privacy-Enhancing Technology) that allow the development of machine learning systems without the need to communicate personal data between participants. These techniques can be both horizontal and vertical and are key in new scenarios for the improvement and development of society, such as Data Spaces.

Federated learning enables the creation of Machine Learning models where, instead of centralising the data in a large repository for analysis, models are sent to the place where the data is located. This strategy, of the 'compute-to data' type, allows local processing of the data to subsequently aggregate the results of the partial models developed and consolidate the information obtained from the learning into a complete model. In this way, it enables the creation of federated data spaces in which each participant maintains control, sovereignty and preserves data protection, choosing at all times who can make use of the data and for which particular use case.

Pseudonymisation of data

Pseudonymisation is implemented through a set of operations within a processing operation (in some very specific cases it could be a processing operation in itself) and is intended as a security measure when it is not possible to fulfil the purposes of processing through anonymisation. One such purpose could be the need to link the data of the same subject between different Data Holders, another is where data is not received in batch mode, but on an ongoing basis (e.g. received from mobile devices or IoT), and another is where a Data Subject needs to be informed specifically about an outcome of the

[154] DATA PROTECTION ENGINEERING, From theory to practice. European Union Agency for Cybersecurity (ENISA) [January 2022]
[155] For example, in the case of a health, financial or police investigation, the Data Holder or the Data Space Mediator would not be informed that the data of a certain person is consulted.
[156] AEPD blog article entitled 'Federated Learning: Artificial Intelligence without compromising privacy | AEPD [April 2023]'

processing of their data (e.g. clinical research) and, therefore, sporadic and selective re-identification is necessary in order to ensure the vital interests of the Data Subjects.

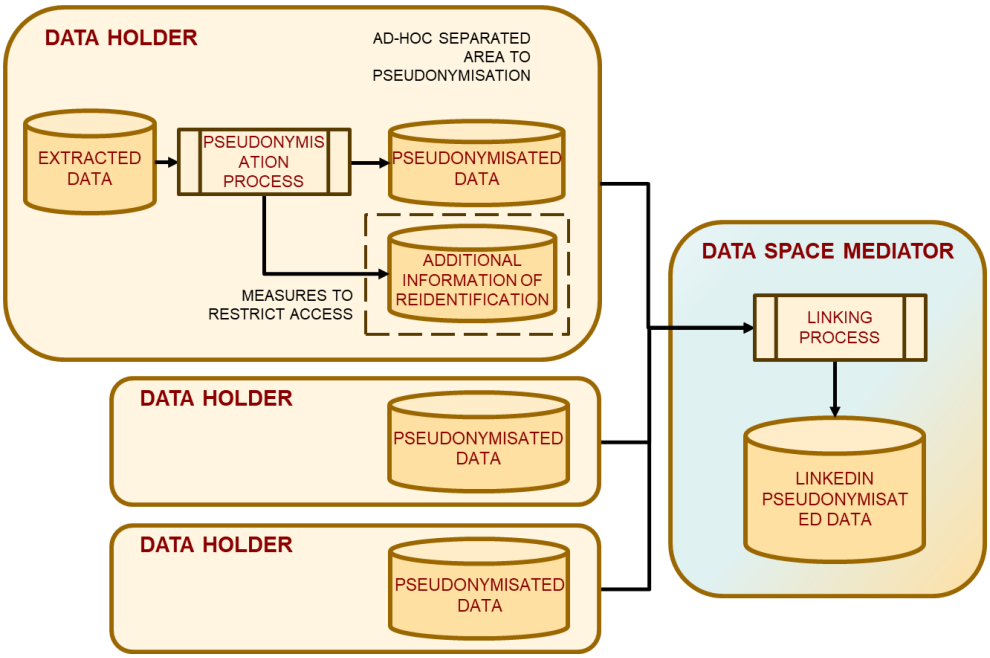

Figure 21: Diagram of the architecture for the pseudonymisation use case

Furthermore, in certain sectors, such as clinical research, there are specific requirements and regulations[157], so that what is developed here will be without prejudice to these regulations. In some sectors, the figure of an Enabler is regulated as a trusted entity that will carry out the pseudonymisation process and is responsible for the custody of the additional re-identification information, for example, the monitor in the case of clinical research.

Polymorphic Encryption and Pseudonymisation techniques (PEP) [158] can also be used in the pseudonymisation process. Each individual is assigned different pseudonyms for each Data User requesting access to the Data Subject's data, thus avoiding the linking of pseudonyms across multiple third parties.

In the specific case of Health Data Spaces, each patient could have a unique identifier. This identifier could be transformed by the Mediator into different pseudonyms depending on the recipient and the context or purpose of data sharing. Each pseudonym is communicated to each recipient together with the polymorphic encrypted data. As a new pseudonym is being generated for each recipient, pseudonyms used for the same patient cannot be linked and are, therefore, considered unlinkable and preserve the confidentiality of the patient's data.

[157] Additional Provision 17 of the LOPDGDD, or the Code of Conduct regulating the processing of personal data in the field of clinical trials and other clinical research and pharmaco-surveillance.
[158] Section 2.2.1 of the document 'ENGINEERING PERSONAL DATA SHARING, Emerging Use Cases and Technologies. European Union Agency for Cybersecurity (ENISA). [January 2023]'.

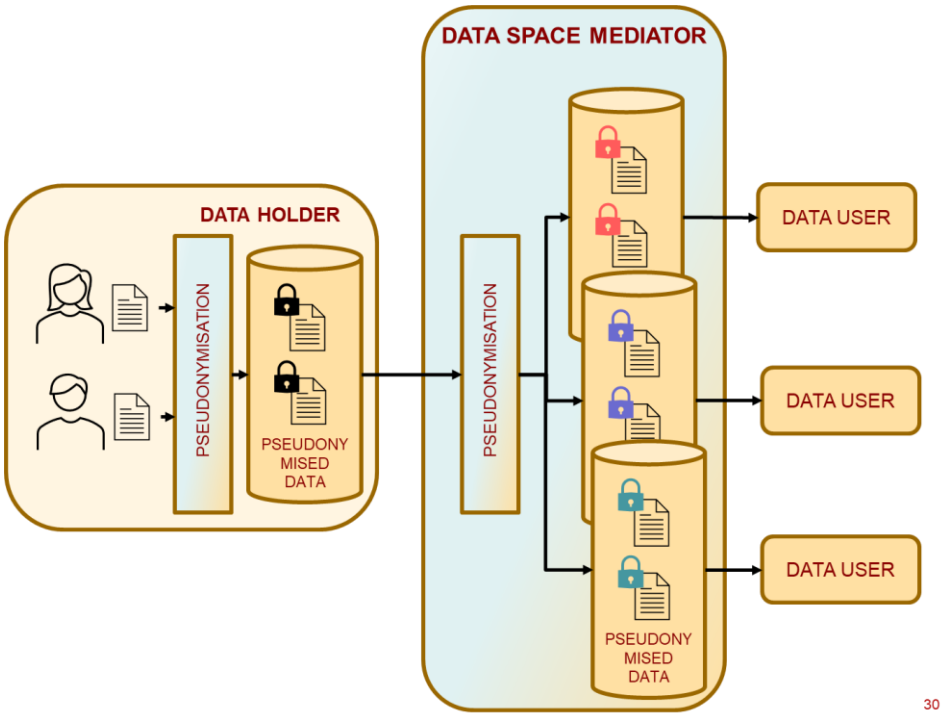

Figure 22: Diagram of the architecture for pseudonymisation of the same data set for different Data Users

Processing requiring anonymised data where it is relevant to link personal information processed by different Data Holders

This is a use case that could arise where strategies such as Secure Multiparty Computing or Differential Privacy cannot work. An example could occur when you want to analyse products or services that the same Data Subject is carrying out on different Data Holders, for which it is necessary to initially link them, but which will eventually be displayed in an anonymised way.

In this case, a prior process of replacement of identifiers and pseudo-identifiers with new pseudo-identifiers not linked to personal data could be carried out. This should be done through a mechanism previously agreed by all Data Holders in the framework of the Data Space governance, so that when records are communicated to the Data Space it is possible to link those corresponding to the same user.

Once received in the Data Space, a consolidation of the anonymised data would be created and the information used to link the records would be discarded.

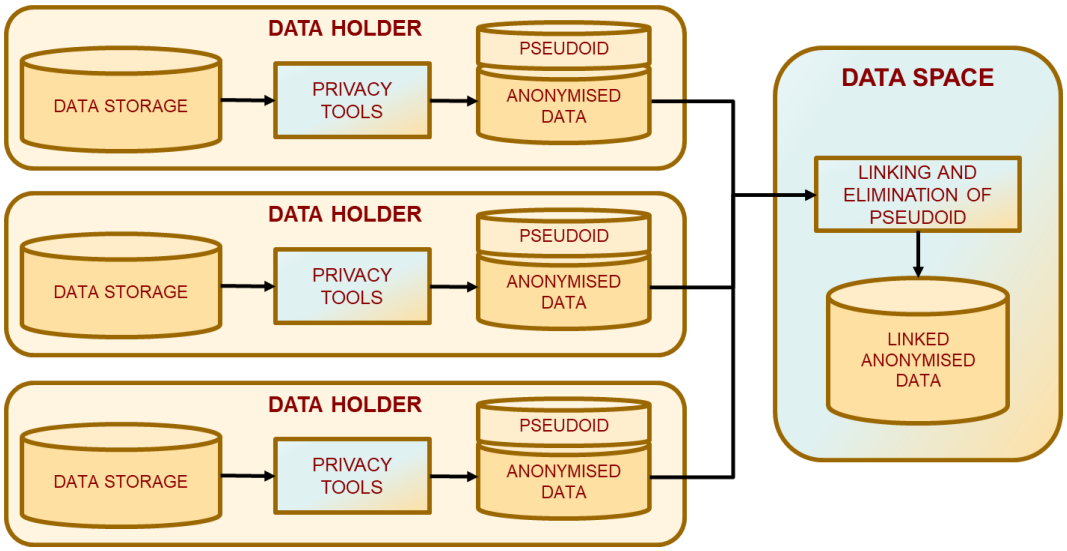

Figure 23: Diagram of the architecture for the use case of anonymised data with linkage between the data of different Data Holders

This approach might have to make use of ad-hoc spaces to implement extraction and anonymisation processes on Data Holders, and in turn a reidentification analysis in the Data Space of the consolidated dataset.

The safeguards that could be put in place could be derived from those already mentioned for the processing of personal, pseudonymised and anonymised data.

Processing where it is not possible to anonymise data

There may be processing operations which by their very nature cannot achieve an adequate level of quality with anonymised data. This circumstance must be correctly assessed in the DPIA and have passed the analysis of suitability, necessity and strict proportionality by data controllers.

In the framework of risk assessment, in particular, the assessment of the proportionality of processing, there are three options, ranked in order of lowest to highest risk, for implementing processing:

1. Transfer the desired processing by the Data User to the Data Holders (e.g., using federated learning techniques).
2. Transfer or communicate personal data to the Data Space, to a Mediator, and move the processing to a Secure Processing Environment provided by the Mediator (applying where appropriate pseudonymisation or anonymisation in the Data Space or other strategies).
3. Transfer or communicate personal data to the Data Space and from the Data Space to the Data User.

This last case will be the last one to be considered, after discarding all the previous ones based on a strict necessity analysis. Moreover, it will be the one that requires a more restrictive assessment of the necessity, already mentioned, and proportionality of the processing, i.e., by providing for stricter data protection safeguards.

Any processing involving the extraction of personal data from the scope of the Data Holder must apply a prior data minimisation analysis. In particular, one or more methods such as the following should be applied:

- Removal of unnecessary fields and metadata (e.g., in case of images).
- Decrease the granularity of the transmitted field information.
- Decrease the frequency of collected events.
- Noise with statistical characteristics that does not degrade the required quality.
- Obfuscation.
- Clustering (e.g., between 40 and 45 years).
- Scrambling.
- Tokenisation.
- Application of encryption techniques.

Among the encryption techniques to be used, it is necessary to consider the use of modern strategies such as homomorphic encryption as described above, as well as attribute-based encryption, proxy re-encryption, polymorphic encryption[159] and others.

In addition, in the framework of a Data Space, it should be noted that personal data collected by a Data Holder may be subject to the following additional processing operations that should be recorded and included in the DPIA:

1. Processing at the Data Holder itself to fulfil the purposes of the Data Space or the Data Users.
2. Transfer of personal data to the Data Space.
3. Transfer of personal data from the Data Space to the Data User.

Secure processing environments

As mentioned above, it may be the case that in order to implement specific processing operations, it is strictly necessary to give access to non-anonymised personal data of Data Holders to the Data Space, as otherwise the purposes of the processing could not be fulfilled.

Some of the previous architectures have dealt with the use of an ad-hoc area to store the extraction and processing of personal data for pre-processing prior to the communication of information to the Data Space. A Secure Processing Environment is related to such ad-hoc areas, and could be defined as areas or services provided by the one that physically stores the data and that allow previously authorised personnel to directly access and analyse data whose free access would pose an unacceptable risk, even with legal safeguards[160]. In the case of access by Data Users to personal data stored in the Data Space, they represent an organisational and technical measure to minimise data processing and data retention (almost bringing retention to zero) in the hands of the Data Users.

[159] ENGINEERING PERSONAL DATA SHARING, Emerging Use Cases and Technologies. European Union Agency for Cybersecurity (ENISA). [January 2023].
[160] SOMA_D2.1.pdf Evaluating Safe space solution including data management and processing setups (disinfobservatory.org)

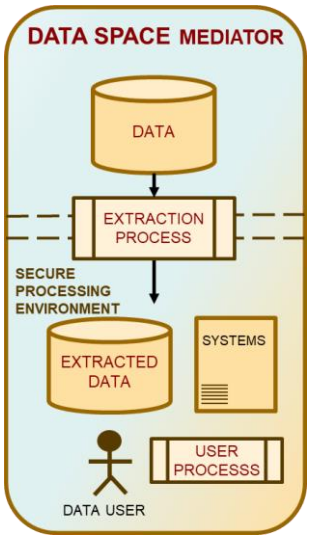

Figure 24: Diagram of a Secure Space in the Data Space Mediator

Access and re-use of data in a Secure Processing Environment cannot be considered an alternative to the legal bases exhaustively listed in article 6 of the GDPR[161].

In the case of protected data held by public sector bodies, Recital 15 of the DGA clarifies that re-use on-site or remotely in a Secure Processing Environment could be allowed provided that the possible requirements to perform a DPIA and to consult the supervisory authority under articles 35 and 36 of the GDPR have been fulfilled, and the risks to the rights and interests of data subjects have been found to be minimal. Public sector bodies shall impose conditions that preserve the integrity of the functioning of the technical systems of the secure processing environment used, reserve the right to verify the process, means and results of the data processing carried out by the Reuser to preserve the integrity of the data protection, as well as the right to prohibit the use of results that contain information that jeopardises the rights and interests of third parties[162].

On the other hand, the EHDS proposal states in its Article 50 '(1) *The health data access bodies shall provide access to electronic health data only through a secure processing environment, with technical and organisational measures and security and interoperability requirements*' and '(2) *The health access data bodies shall ensure that electronic health data can be uploaded by data holders and can be accessed by the data user in a secure processing environment. The data users shall only be able to download non-personal electronic health data from the secure processing environment*'.

The same article in the EHDS proposal lists the security measures that such Environments are required to have and can serve as a guide for other Data Spaces:

a) *'restrict access to the secure processing environment to authorised persons listed in the respective data permit;*
b) *minimise the risk of the unauthorised reading, copying, modification or removal of electronic health data hosted in the secure processing environment through state-of-the-art technological means;*

[161] Paragraph 81 of the document "Joint EDPB-EDPS Opinion 3/2021 on the Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) [10 May 2021]".
[162] Article 5(4) of the DGA

c) *limit the input of electronic health data and the inspection, modification or deletion of electronic health data hosted in the secure processing environment to a limited number of authorised identifiable individuals;*

d) *ensure that data users have access only to the electronic health data covered by their data permit, by means of individual and unique user identities and confidential access modes only;*

e) *keep identifiable logs of access to the secure processing environment for the period of time necessary to verify and audit all processing operations in that environment;*

f) *ensure compliance and monitor the security measures referred to in this Article to mitigate potential security threats.*'

Secure Processing Environments could be defined at two levels:

- Secure On-site/Physical Processing Environment, which implies that the Data User's processing and operator physically moves to the premises where the data is located, and there it is subject to access and processing controls. The resulting information can then be extracted from the secure space.
- Remote/virtual Secure Processing Environment, where although the processing is executed at the premises where the data is stored, the operator can manipulate the processing remotely, so that the operator does not have access to the data, but does have access to the resulting information. Manipulation would be done through secure virtual networks or even physical private networks.

In both cases, the original data does not leave the physical storage location of the information.

The second case, the Secure Processing Environment with virtual access, has demonstrated in its effective application its vulnerability, even when access is allowed only through physical private networks, resulting in data breaches of great social impact. Its use must therefore be complemented by other safeguards.

In addition, such Secure Processing Environments should be implemented over Trusted Execution Environments[163]. While such environments should be implemented for all processing of personal data, it is in Secure Spaces where it would be more critical. A Trusted Execution Environment (TEE), as defined by ENISA, is an inviolable processing environment on the main processor of a device. Running in parallel to the operating system and using both hardware and software, TEEs are designed to be more secure than traditional processing environments. It is also called a rich operating system execution environment (REE), in which the device's OS and applications run.

F. STORAGE OF PERSONAL AND NON-PERSONAL DATA IN DATA SPACE

In the event that personal data are stored, even temporarily, in addition to non-personal data (other than mixed datasets) at a Data Space Mediator, it should be analysed whether a high risk is incurred (see section VI.B. High Risk).

[163] Section 4.3 of the document 'DATA PROTECTION ENGINEERING, From Theory to Practice. European Union Agency for Cybersecurity (ENISA) [January 2022]'

Measures that could be implemented to ensure and be able to demonstrate compliance with data protection law in this use case could be, for example, legal measures such as agreements for the limitation of storage and processing beyond the obligations of the GDPR, confidentiality commitments of personnel, clauses on cancellation periods or limitation of processing, among others.

Organisational and data protection policies could include human supervision of access to data sets, functional separation of personal and non-personal data, among others.

Technical measures could include physical separation of the two data sets, preventing access directly from the Internet to personal data sets, incorporating traceability of third-party access by means of a corresponding report (Traceability report) including person (natural, not legal), date of access and access session time, automatic audits, technical limitation in terms of data accessed, among others[164].

[164] Irrespective of the measures that would be recommended or obligatory in the regulation for non-personal data. For example, the DGA sets out requirements for the security of non-personal data and data intermediation services in Article 12(l), or for altruism organisations in the Union in Article 21(4).

VI. PERSONAL DATA PROTECTION ISSUES IN A DATA SPACE

To strengthen and implement individuals' control over their own personal data[165], the GDPR establishes a set of principles and rights, as well as obligations for those who process personal data, which run throughout the text of the regulation.

The principles[166] of lawfulness, fairness, transparency, purpose, minimisation, accuracy, retention, security and proactive responsibility are all mandatory, autonomous and complementary to each other, and cannot be reduced by each other, such as, for example, considering that security for the protection of personal data subsumes any of the above. Therefore, processing operations on a Data Space must comply with all of them.

Failure to comply would impede the free movement of personal data within the European Union and thus constitute an obstacle to the exercise of economic activities[167]. The proper functioning of the internal market requires that the free movement of personal data within the EU is not restricted or prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data[168] and, in order for this not to be the case, processing operations must be implemented with appropriate measures to ensure and demonstrate compliance with the GDPR, and that those measures are reviewed and updated where necessary[169].

Some of the issues arising from data protection law for the specific framework of Data Spaces have already been developed in the previous sections. In this chapter, some other relevant issues will be developed, without claiming to be exhaustive and without prejudice to the applicable sectoral rules.

A. DATA PROTECTION OFFICER

Data Spaces are proposing processing of a so far unknown dimension, which means that minimum or formal compliance cannot be considered, but that it will be necessary to apply a level of compliance equivalent to the dimension of the processing. As established in article 24(1) of the GDPR, the measures that, from the conception and design of the Data Spaces, guarantee and enable compliance to be demonstrated, must be sized according to 'nature, scope, context and purposes of the processing'. Hence the importance of the involvement of Data Protection Officers and data protection advisors in the earliest stages of the definition of a Data Space and from the design stage in its processing.

Data Space interveners should determine whether they are obliged to appoint a Data protection Officer (DPO), in relation to the scale of the processing they carry out, as this may fall within the scope of Article 37(1)(b) or (c). Public authorities, as established in Article 37(1)(a), must have a DPO.

In the event of not having the obligation, it is considered advisable that all Mediators have a DPO and/or a data protection advisor. This is especially advisable for entities that

[165] Recital 7 of the GDPR
[166] Article 5 of the GDPR
[167] Recital 9 of the GDPR
[168] Recital 13 of the GDPR
[169] Article 24(1) of the GDPR

exercise the functions of supervision and authorisation of access to the Data Space. In this sense, and insofar as these supervisory actions are carried out by the research ethics committees, the legislator itself requires these committees, in the health, biomedical or medication field, to include among their members a data protection officer or, failing this, an expert with sufficient knowledge of the GDPR when dealing with research activities that involve the processing of personal data or pseudonymised or anonymised data[170].

The DPO must be involved in the definition of Data Space governance models and policies, analysis and assessment of use cases, selection of data protection measures from design, management of personal data breaches and in advising on and monitoring impact assessments.

B. RISK MANAGEMENT AND DATA PROTECTION IMPACT ASSESSMENT

Under Articles 24 and 35 of the GDPR, a controller of personal data is obliged to carry out a management of risks of varying likelihood and severity to the rights and freedoms of natural persons and, where appropriate, a data protection impact assessment (DPIA). This obligation has to be exercised by controllers who are considering processing operations on the Data Space infrastructure, and also by the legislator when it is the legislator who establishes a Data Space by law[171]. It is therefore of vital importance to identify the different controllers behind the data access authorisation and the role of the Access Supervisor supported, where appropriate by a Data Protection Enabler, is fundamental.

Proper management of risk to rights and freedoms includes the assessment of risk and the selection, implementation, review and updating of appropriate measures to ensure compliance. A processing model in the Data Space that does not manage risk by complying with data protection principles by design and by default could mean stretching all data protection principles, in particular minimisation, retention and limitation of processing, to the limit.

Risk management is not a formal requirement, but a tool for making decisions on how the use cases of processing operations are to be implemented by design to ensure compliance with data protection law and minimisation of the impact on Data Subjects. As an example of a flawed approach to risk management for the rights and freedoms of Data Subjects in the framework of Data Spaces, the '*EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space*', in paragraph 49, states that the EHDS draft has failed to manage the risks to rights and freedoms by not having conducted a DPIA since, according to footnote 18 accompanying the paragraph, both institutions interpret the accompanying document '*Commission Staff Working Document Impact Assessment Report*' as not being a DPIA, despite its title, as it does not carry out a risk assessment or provide the necessary measures to mitigate the risk.

Risks for fundamental rights

There is a clear need to manage by design the risks to Data Subjects that may arise from materialisation of personal data breaches in Data Spaces. However, the risks to

[170] Additional provision 17 of the LOPDGDD.
[171] Consult the Guidelines for conducting a data protection impact assessment in regulatory development from AEPD.

fundamental rights go beyond personal data breaches, as the data processing itself may pose a risk or limitation to these fundamental rights. The WP248 Guidelines[172] state that protection should extend to other fundamental rights, such as freedom of expression, freedom of thought, freedom of movement, prohibition of discrimination, freedom of conscience and religion, inviolability of the home or communications, or effective judicial protection.

It should be borne in mind that many of the processing operations in Data Spaces will involve both public authorities and private parties, and whose legal basis can be found in the regulation.

Both the case law of the Court of Justice of the European Union (CJEU[173]) and the opinions of EDPS[174] state that the data protection impact assessment of a regulation should be carried out in cases where the proposed legislative measure involves the processing of personal data. Any data processing operation envisaged by the legislation entails a limitation of the right to the protection of personal data, irrespective of whether such a limitation may be justified. In turn, the European Court of Human Rights (ECHR) has held that the storage by a public authority of data relating to the private life of an individual amounts to a limitation of the right to respect his or her private life[175].

It is settled case law of the CJEU that in establishing 'the existence of an interference with the fundamental right to privacy, it does not matter whether the information on the private lives concerned have been inconvenienced in any way'[176]. The assessment of the respect to the essence of the right may, in some cases, require an in-depth legal analysis, hence the importance of assessments of appropriateness, necessity and strict proportionality of processing operations and the implementation of measures, beyond cybersecurity measures, derived from Data Protection Impact Assessments.

High risk

Processing operations in the framework of a Data Space entail at least the access, which could include communication and storage in third parties, of massive amounts of personal data from different sources with a novel technical/organisational solution, on a large scale, in a comprehensive, systematic way, involving association and combination of data, automated, and oriented towards the application of novel technologies[177]. It should be noted, in particular for European Data Spaces, that it opens up the possibility of processing operations involving all Data Subjects, with massive collection of information in terms of

[172] WP248 Guidelines on data protection impact assessment (DPIA) and for determining whether processing is 'likely to result in a high risk' for the purposes of Regulation 2016/679 (Article 29 Data Protection Working Party) [4 October 2017]

[173] TJUE, joined cases C-293/12 y C-594/12, Digital Rights Ireland Ltd, paragraphs 34 - 36; see also the joined cases C-92/09 y C-93/09 Volker und Markus Schecke, paragraph 58.

[174] Section II.5 of the document 'Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit (EDPS) [11 April 2017]'.

[175] TEDH, Leander c. Sweden, paragraph 48.

[176] CJEU, cases C-465/00, C-138/01 and C-139/01 Österreichischer Rundfunk and others, paragraph 75 and Digital Rights Ireland, paragraph 33.

[177] Section II.B.a. of the document "WP248 Guidelines on data protection impact assessment (DPIA) and for determining whether processing is 'likely to result in a high risk' for the purposes of Regulation 2016/679 (Article 29 Data Protection Working Party) [4 October 2017]".

data categories, granularity and frequency, correlation of different sources, access by multiple interveners, etc., and for multiple purposes.

Although risk management for the rights and freedoms is required for all processing operations under Article 24(1), high-risk processing operations [178] are obliged to carry out a data protection impact assessment (DPIA). In order to determine whether there is a high risk, it is necessary to start by looking at the cases that are already assessed, such as:

- Cases under Article 35(3) of the GDPR.
- The special regulation requiring a DPIA for processing or identifying risk factors.
- Cases and examples from the WP248 Guidelines.
- The cases on the list approved by the AEPD on the basis of article 35(4) of the GDPR.
- Cases under article 28(2) of Spanish Organic Law 3/2018, of 5 December, on the Protection of personal Data and Guarantee of Digital Rights (LOPDGDD, by its initials in Spanish).
- The cases of article 32(2) of the GDPR.
- The risks identified in Recital 75 of the GDPR[179].
- The specific cases and conditions described in the guidelines issued by the EDPB for specific processing operations.
- The specific cases and conditions described in the codes of conduct pursuant to Article 40 and certification mechanisms pursuant to Article 42 of the GDPR.

Even if a processing is not in the high-risk data set, which is a list of minimums, it must be determined for each processing that the impact of the processing is not high risk. It is important to note that the existence of a high risk to the rights and freedoms of Data Subjects is not exclusively linked to the processing of special categories of data, although the processing of this type of data will increase the possibility of a higher impact for Data Subjects. To facilitate the analysis of the existence of a high risk, the AEPD has made available guides and tools[180] that simplify its determination.

If the process of managing a high risk fails to mitigate the risk, the data protection authority must be consulted. If the criteria of appropriateness, necessity and proportionality are not met, or the high risk could not be mitigated, the process cannot be carried out.

[178] Article 35(1) of the GDPR

[179] *The risk to the rights and freedoms of natural persons, of varying likelihood and severity, may result from personal data processing which could lead to physical, material or non-material damage, in particular: where the processing may give rise to discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data protected by professional secrecy, unauthorised reversal of pseudonymisation, or any other significant economic or social disadvantage; where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data; where personal data are processed which reveal racial or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, and the processing of genetic data, data concerning health or data concerning sex life or criminal convictions and offences or related security measures; where personal aspects are evaluated, in particular analysing or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles; where personal data of vulnerable natural persons, in particular of children, are processed; or where processing involves a large amount of personal data and affects a large number of data subjects.*

[180] Available on the web site: Innovation and Technology | AEPD

Social risk

Data protection measures, in particular data minimisation measures or security measures for access control management, are normally aimed at minimising the individual impact on the right to data protection of excessive processing of personal data by a controller or the compromise of the personal data. However, when we are in the framework of Data Spaces where the extent of the processing, whether in the categories of data, in the categories of Data Subjects or even in the retention periods is very long, it is necessary to establish specific data protection measures from the design to diminish a possible impact both at the individual level and on society as a whole.

There is also the perspective of the risk to society of the impact of a massive amount of Data Subjects having their personal data compromised. Intrusion of the right to data protection arises from the mere accumulation of personal information in certain organisations. In these cases, not only must each individual impact on fundamental rights be considered as a sum of these, but it has a multiplier effect that affect the fundamentals of our society: lack of trust in institutions, manipulation of large sectors of the population or particularly vulnerable groups, putting at risk massive parts of the population making mitigation measures unfeasible, etc.

Means accountability

The principle of accountability of processing operations can hardly be fulfilled if the technical means used to implement it (its nature) are not themselves accountable.

Therefore, the controller shall have the obligation to demand the information and collaboration necessary for compliance with the standard to be guaranteed and to be able to demonstrate it, both from data processors and from Enablers that do not have such a nature, but provide tools used by controllers/processors/sub-processors to implement the means of processing. The services and tools they provide must have their quality audited, accredited or certified when they can influence the processing of personal data carried out by controllers or processors, so that the controller can comply with its obligations.

Application of the precautionary principle by design

The European Commission stated in its Communication on the precautionary principle[181]: '*Although the precautionary principle is not explicitly mentioned in the Treaty except in the environmental field, its scope is far wider and covers those specific circumstances where scientific evidence is insufficient, inconclusive or uncertain and there are indications through preliminary objective scientific evaluation that there are reasonable grounds for concern that the potentially dangerous effects on the environment, human, animal or plant health may be inconsistent with the chosen level of protection'*.

In this respect, the EDPS has pointed out the timeliness of the application of the precautionary principle as a preventive measure in the case of processing with a high

[181] Communication on the Precautionary Principle (COM(2000)1 final) [2 February 2000]

impact or where there is uncertainty as to what the impact might be[182]. That is to say, as a tool for risk management.

With regard to Data Spaces, the application of the precautionary principle can be adopted with several strategies: the implementation of sandboxes[183], the adoption of an 'incremental approach' in the deployment of the processing (limitation in geography, in categories of data subjects, in categories of data, number of interveners, processors/sub-processors, etc.) with milestones of evaluation supervised by independent authorities, etc.

Safeguards in data communications

In the case of communications of personal data, depending on the risks of Data Subjects, a thorough assessment should be made of the appropriateness and necessity of such communication, and of the proportionality of such communication with respect to the processing in the framework of risk management. In addition, the following safeguards could be obtained in contractual commitments, licences[184] of use limiting their processing, including penalty clauses, irrespective of criminal, civil or administrative liabilities incurred, in relation to:

- No re-use of information.
- Non-transmission to third parties.
- No anonymisation of the information to be made available to third parties.
- No storage or processing of information in data processors, especially in the Cloud.
- No storage or processing of information in Data User's systems outside the European Economic Area.
- Implementation of data protection solutions by design and by default in the design of the processing.
- Availability of an effective personal data breach management system.
- Involvement of the Data Protection Officer in such processing.
- Demonstration, by means of certification from an independent third party, of compliance with data protection regulations for the specific processing it intends to carry out, the qualification of the employees in relation to such processing and compliance with the conditions listed above.
- Certification, by means of a request to the Supervisory Authority, of the non-existence of sanctions for non-compliance with data protection regulations.

Security measures

Data Spaces must ensure a high level of cybersecurity[185]. Security measures of the utmost importance, however, it should be remembered that many of the limitations and

[182] EDPS Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the protection of personal data [19 December 2019].
[183] Controlled testing environments.
[184] Recital 28 and the definition in Article 2(10) of the DGA refer to the use of licences in data sharing.
[185] Recital 2 of the DGA

risks to the rights and freedoms of individuals in data processing are not solved by security measures. In particular, those risks related to the principles of minimisation, limitation of processing and limitation of retention, among others, need to be managed with measures on the concept of processing, governance, data protection policies, data protection measures by design and by default (such as those outlined in the previous chapter), personal data breach management, etc. Not all security measures at aimed at protecting the rights and freedoms of Data Subjects, some will be processing for other purposes and their own legitimisation[186].

It should be remembered that security measures, according to experience and the doctrine of the Supreme Court[187], are an obligation of means, but not of ends. The reality of personal data breaches makes it evident that the materialisation of threats to datasets is a matter of time, and that the only unknown is the dimension it will have.

In the case of public authorities, or other obliged entities, they have to implement the ENS measures[188] corresponding to the assessment of the level of risk to the rights and freedoms of the natural persons involved in the processing. These measures are a catalogue of minimum measures, and depending on the specificities of the processing (see Article 32 of the GDPR) and the assessment of the risk from other perspectives (e.g., in the case of critical infrastructures) they will have to be extended. Furthermore, depending on their role in the Data Space and the impact this may have on the security of the Data Space as a whole, it is highly recommended that they be certified.

Availability and resilience

Article 32(1)(b) of the GDPR requires, inter alia, ensuring the continued availability and resilience of services and processing of personal data.

The DGA also expresses itself in this sense, but in a general manner, when it requires that Data Mediators acting as data intermediation services shall ensure in the event of insolvency, the reasonable continuity of the provision of their data intermediation services and, where such data intermediation services include the storage of data, they shall have in place the necessary safeguards to enable Data Holders and Data Subjects to access, transfer or retrieve their data and, where they provide such intermediation services between Data Subjects and Data Users, to enable Data Subjects to exercise their rights[189].

Personal data breach scenarios

When considering a Data Space, those who process personal data should consider different data breach scenarios. This analysis exercise should seek an answer to at least the following questions:

- What personal and societal impact a personal data breach can have.

[186] Recital 49 of the GDPR
[187] C.G.P.J - Noticias Judiciales (poderjudicial.es) [only available in spanish]
[188] National Security Scheme established by Royal Decree 311/2022 of 3 May, which regulates the National Security Scheme.
[189] Article 12(h) of the DGA

- What data protection measures have been implemented or should be implemented to minimise the impact for Data Subjects and society in the event of a personal data breach.
- What contingency measures should be in place when the breach occurs to also minimise such impact, to cope with notification obligations to Supervisory Authorities and communication to Data Subjects.

In such an analysis, it has to be taken into account that the Data Space implies, at the very least, processing that may not only be at national level, but also at European level, and with implications beyond the EU framework. This implies that, when considering breach scenarios, the following should be taken into account:

- Failure of the rule of law.
- Situations of national or international emergency.
- Crises in international relations and agreements.

Due to the relationship with Data Spaces, it is recommended to consult the Guidelines to manage data breach risk in public sector bodies massive data communications.

Reidentification

Finally, although it is dealt with in the section of anonymisation, it should be noted that reidentification is considered a personal data breach.

In particular, the Public Sector Reuser of Protected Data shall notify the public sector body that gave access to it of any breaches that allow the reidentification of Data Subjects in non-personal data sets[190].

Cooperation between interveners

The controller may receive assistance from third parties to carry out its obligations. Already Article 28(3)(f) of the GDPR establishes the obligation of processors to provide such assistance where appropriate. In particular, the GDPR provides that the processor *'shall assist he controller in ensuring compliance with obligations pursuant to Articles 32 to 36, taking into account the nature of the processing and the information available to the processor*'. Furthermore, Recital 78 of the GDPR states that *'When developing, designing, selecting and using applications, services and products that are based on the processing of personal data or process personal data to fulfil their task, producers of the products, services and applications should be encouraged to take into account the right to data protection when developing and designing such products, services and applications and, with due regard to state of the art, to make sure that controllers and processors are able to fulfil their data protection obligations'*.

The complexity of the Data Space environment, in which there may be several areas of responsibility, make collaboration between all interveners managing the risks involved in the processing of personal data, whether they are controllers, co-responsible parties, processors, sub-processors or technology providers, essential. Effective and efficient data

[190] Article 5(5) of the DGA

protection will require a concerted effort to plan and select data protection-oriented use case implementation scenarios from design and a combined approach to the GDPR-compliant solution. The DPIA and the solutions that mange the limitations and risks to rights and freedoms must emerge from a common effort and the result must be unique.

Such cooperation must take place in the implementation and management of governance measures, policies, data protection strategies by design and by default, security measures, breaches management, etc., aimed at managing the risk to rights and freedoms.

As mentioned above, it is essential to have data protection Enablers (a task that could be carried out by a Data Mediator) to coordinate and provide legal, organisational and technical support to those involved in a processing operation within the framework of the Data Space. This figure will be vital to fulfil the purpose that justifies the objective of the processing of personal data in the framework of a Data Space, which is that the use of such data is available to society, paying special attention to the position of SMEs, entrepreneurs and small research groups.

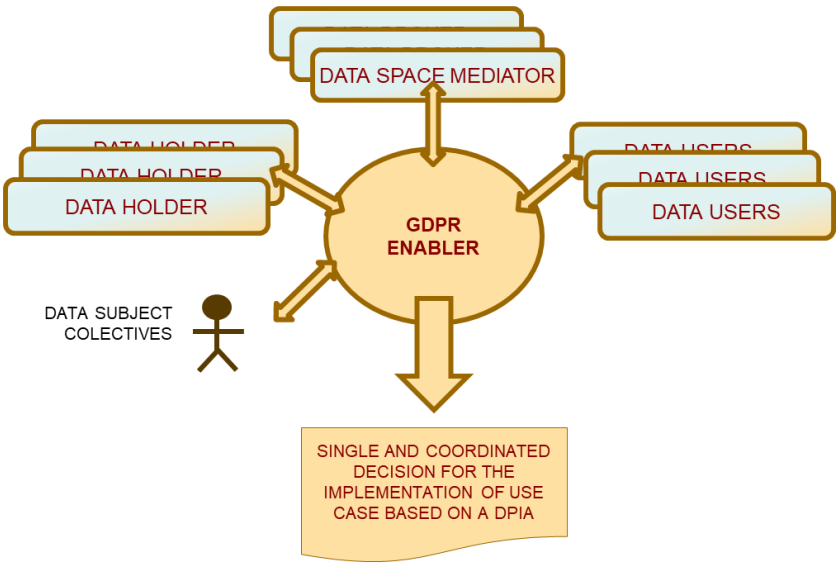

Figure 25: Role of a Data Protection Enabler for the coordination and legal, organisational and technical support to the different interveners involved in a processing operation in a Data Space

Scenarios in relation to the implementation of the DPIA

The set of scenarios in which the implementation of a DPIA with collaboration between the interveners may be as wide as the data processing scenarios envisaged within the Data Space and will have to take into account specifically the roles of controller, or joint controller, or processor, or sub-processor, or technology provider without access to data adopted by each of the interveners in the processing.

Therefore, it is not proposed to provide an exhaustive list of these, nor an assessment of other requirements such as the legitimacy to carry out the processing, but rather to show some examples of how such collaboration could be approached when carrying out the DPIA:

Example 1

A Data Space Mediator is envisaged to build a repository with anonymised information collected from multiple Data Holders. In that case, each Data Holder should perform a risk analysis of the anonymisation process and, where appropriate, a DPIA. The Data Mediator, in turn, shall perform a risk analysis of the consolidation of several anonymised sources based on the characteristics of the anonymised data, their diversity, their volume, and the possible reidentification processing, and, where appropriate, a DPIA. In the execution of this DPIA, from which recommendations on how to execute the anonymisation should be concluded, and to conclude that there is a strategy that allows it to be carried out with sufficient guarantees and quality, the Data Space Mediator and the Data Holders should work in coordination to draw conclusions.

Example 2

A Data User wishes to carry out a processing operation on data held by a Data Space Mediator acting as data controller, which can be carried out by anonymisation. This Data Mediator will carry out anonymisation processing prior to transferring the data to the Data User. In this case, the Data Mediator has to carry out the risk management and, where appropriate, the DPIA anonymisation process. This assessment has to be carried out in coordination with the Data User, both to determine his/her data quality requirements or that of the possibility of reidentification that might exist from other sources accessed by the Data User.

Example 3

The Data User requests processing on non-anonymised data that a Data Space Mediator already has available, for which the Data Mediator acts as data processing controller and on which the Data Mediator is going to enable execution in a Secure Processing Environment for the Data User to extract only non-personal information.

The Data Mediator has to perform the risk management and, where appropriate, the DPIA of the processing within the scope of its obligations in the processing to enable a non-anonymised data set. The Data User must also perform the risk management and the DPIA of the processing in his/her area of responsibility because, although the processing is carried out in a Secure Processing Environment of the Data Mediator, the User has his/her share of responsibility for the processing. Both have to be coordinated with regard to the safeguards established in the Secure Processing Environment, the possibility of remote access to the Secure Processing Environment, the processing carried out in the Secure Processing Environment and the procedure for extracting results.

Example 4

The Data User requests to process non-anonymised data which the Data Space Mediator does not physically hold in its systems, but which is catalogued as being held by several Data Holders. The Data Mediator can manage access to Secure Processing Environments at Data Holders in order to implement the requested processing using data protection strategies by design and allowing the extraction using anonymised information.

Data Holders, as controllers, have to perform risk management and, where appropriate, DPIA of the processing within the scope of their processing obligations. The Data Mediator would take the role of Enabler, at least in the task of serving as a bridge between Data Holders and Data Users, and the Data User has to perform the risk management and the DPIA of the processing in his/her area of responsibility. The risk assessment and the measures taken cannot be carried out efficiently without a collaboration between Holders, Mediators and User, therefore, they have to act in a coordinated manner in the execution of risk management, possibly assisted by a third party or with the leadership of the Mediator in his/her role as Enabler.

Review and update of measures

Article 24 of the GDPR establishes the need to review and update the measures[191] when these conditions change, as the measures adopted must be adequate to ensure and be able to demonstrate compliance. In a Data Space, the elements that make up the processing operations carried out (nature, extension, context and purposes) will change very dynamically, as will the risks to the rights and freedoms that arise from them. Therefore, the governance of the Data Space should provide for mechanisms to implement the review and updating of measures, in particular, in relation to the risks of reidentification and the need for re-anonymisation of datasets.

Article 32(1)(d) of the GDPR goes beyond this obligation to review security measures, requiring regular reviews. Due to the risks in these environments, it is recommended that reviews be conducted on an annual basis, with the frequency of reviews being increased after the completion of the DPIA for the different use cases in the Data Space, if this reduces the risk.

Resources and transparency

The AEPD has published a number of resources for the management of risk to rights and freedoms, and for the implementation of the data protection impact assessment process, which can be found on the AEPD's website.

Making the results of the DPIA public is a good transparency measure to increase confidence in the processing carried out in the framework of the Data Spaces[192].

[191] AEPD blog article entitled 'When to review data protection measures | AEPD [February 2023]'
[192] Section 27 of the 'Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]'

C. RELATIONSHIPS BETWEEN INTERVENERS IN THE DATA SPACE

Chapter IV of this document dealt with the categories of interveners in the Data Space, as well as with the responsibilities in the processing operations. Further aspects of the relationships between Data Space interveners will be developed in this section.

Formalisation of processing between interveners

In addition to the obligations established in the GDPR for the formalisation by means of a contract or other legal act of the controller-processor-subprocessor in compliance with all the requirements of article 28 of the GDPR, the data communication relationship between controllers and, where applicable, the situations of co-responsibility[193] must also be formalised.

This formalisation should be part of the governance arrangements in the Data Space and it is advisable to set out in such arrangements:

- The provision of anonymisation/pseudonymisation tools to Data Holders, either through Enablers, or with own resources.
- The provision of traceability and logging tools for the use of data.
- The provision of notification tools to Data Subjects in cases where their personal data are used either through Enablers, or with the Data Space Mediator's or Data Users' own resources.
- Clauses to establish in a coordinated manner diligent security measures for access and transfer of data.
- And other aspects implementing common data protection policies[194].

Because of their importance, it is worth highlighting the importance of legal measures that go beyond the minimum required by data protection law for risk management in the dissemination of personal and anonymised data that may be part of the formalisation of the relationships between interveners. These include contractually limiting the scope of dissemination of anonymised data (e.g., only among a group of researchers) or establishing retention requirements and limitations, which are the kind of safeguards common to reduce other kinds of risks in non-personal data.

These safeguards can be represented in licences for use in Data Spaces.

Procedure for access to the Data Space when personal data are processed

The processing of personal data in the framework of a Data Space must be authorised, processed and supervised by those who are responsible for the data to be processed under the GDPR. The responsibility itself cannot be delegated, but it can be supported by a competent judgement, which in this case could be that of the Supervisor of the Data Space access requests.

The procedure to be considered would have to be reflected in the governance of the Data Space and should have sufficient safeguards such as being properly documented,

[193] Article 26 of the GDPR
[194] Article 24(2) of the GDPR, not to be confused with privacy policies.

reasoned, with references to the legal basis, the information that needs to be obtained to justify access to certain data and all the requirements for compliance with the GDPR. The procedure should include the necessary steps to define, in cooperation between the User with Data Mediator and Data Holders if applicable, the data protection measures from design, processing limitations, risk management, likelihood of reidentification of personal data, management of possible breaches, compliance controls and audits, formalisation of the relationships between the interveners, etc., until a precise definition of the processing is obtained. All this would also support the drafting of an eventual DPIA.

Human oversight in the decision to access personal data

The DA proposal envisages the possibility of providing access to data in Data Spaces through the use of 'Smart contracts or contracts established in algorithms and of automatic execution[195] with certain guarantees: rigorous access control mechanisms, limitation of the period of validity of the contract and interruption capacity, transparency and the same effective judicial protection as any other contract.

On the other hand, the DGA, in the case of single information points allowing the re-use of specific categories of data held by public sector bodies, although it must be able to rely on automated means when transmitting queries or requests for re-use, sufficient human supervision in the transmission process must be ensured[196].

In relation to data protection, a decision to give generic access to, or communicate, personal data to third parties could have legal consequences for data subjects or affect them in a similar way. In that case, it would have to be considered whether a fully automated decision could contravene Article 22 of the GDPR. On the other hand, in certain cases and for certain types of Data Users, there should be procedures for granting automatic access after initial assessment, with the condition of establishing regular monitoring and control of the accesses made, for example, for the primary use of data in the framework of Health Data Spaces.

Independently of the above, it would also be necessary to determine the risk to rights and freedoms that could arise from giving access to personal data to third parties automatically, inter alia, due to possible errors in smart contracts[197].

Interoperability

Chapter VIII of the DA proposal is aimed at defining the minimum requirements for interoperability in Data Spaces. The interoperability obligation is also set out in the DGA in the Data Space Mediators under its competence[198].

Interoperability plays a key role in the proper implementation of GDPR compliance measures. Poorly defined interoperability or interoperability that does not address data protection requirements would not be able to properly implement data protection

[195] Article 30 of the DA proposal.
[196] Recital 26 of the DGA
[197] Like any algorithm, it has a probability of error or simply suffering from faulty design: AkuDreams dev team locks up $33M due to smart contract bug (cointelegraph.com)
[198] In the case of data intermediation services, Article 12(i) of the DGA applies and for altruistic purposes Article 21(1)(d) applies.

solutions by design or by default and data protection principles, in particular, consent management, exercise of rights, traceability of datasets and personal data, implementation of compute-to-data strategies, efficient application of Secure Processing Environments, etc.

The lack of interoperability in the implementation of a Data Space will affect the fulfilment of the purposes of the processing operations envisaged in the framework of the Data Space. In this case, it would be necessary to reconsider whether the requirements of adequacy and necessity of the processing operations are still met.

Interaction between Data Space Mediators

In some cases, there may be more than one Data Space Mediator involved in the processing that a Data User wants to raise. Even those Mediators may be acting in the framework of different Data Spaces. The interaction between Mediators, or between several Mediators and a Data User, has to consider other aspects beyond those of technical interoperability.

The governance mechanisms, starting from the figure of the Supervisor of access requests, must contemplate the implementation of privacy measures from the design and also risk management measures for rights and freedoms, as well as, where appropriate, the performance of the DPIA, in particular in relation to the likelihood of reidentification.

For example, the decision to include a certain dataset in a response may be taken by a Mediator, but the data itself may be controlled by a different Data Space Mediator. In such cases, Data Space Mediators have to cooperate within the framework of their responsibilities to respond to the data request, which involves governance, management (in particular of the risk to the rights and freedoms of Data Subjects) and technical mechanisms[199].

Selection of processors/sub-processors in the Data Space

The selection of processors/sub-processors is part of the nature of the processing, that is, the way in which the processing will be implemented. The controller shall be obliged to ensure that the principles of accountability, data protection by design and risk management for the rights and freedoms are followed in the selection process of processors and sub-processors. The selection of processors/sub-processors in the Data Space requires an analysis in the context of the special characteristics of the Data Space.

It should be noted that the use of a processor does not imply a diversion of obligations from the controller to a third party. On the contrary, very specific obligations are added for the controller, such as the obligation to use only processors providing sufficient guarantees, in particular in terms of expertise, reliability and resources, for the implementation of technical and organisational measures that meet the requirements of this Regulation, including the security of the processing[200], the necessary diligence in their selection and, as part of the nature of the processing, and the obligation to manage the

[199] Section 4.3.2 of the document 'ENGINEERING PERSONAL DATA SHARING, Emerging Use Cases and Technologies. European Union Agency for Cybersecurity (ENISA). [January 2023]'.
[200] Recital 81 of the GDPR.

specific risk that may arise from the selection of different processors. It should be recalled that the documentation of such diligence is part of the proactive accountability obligations.

In particular, compliance with the conditions listed in Article 28(3) of the GDPR is required, including 28(3)(h): '*makes available to the controller all information necessary to demonstrate compliance with the obligations laid down in this Article and allow for and contribute to audits, including inspections, conducted by the controller or another auditor mandated by the controller*'. In addition to the requirement of Article 28, the processor must effectively enable the exercise of the powers granted to the Supervisory Authorities in Article 58(1) as regards their powers of investigation, which cannot be effectively devoid of content.

The selection of processors/sub-processors should be oriented towards fulfilling the purposes of the Data Space and should not be a way of circumventing the safeguards and measures that will be required from the Data Space Mediator and Data Users to ensure effective compliance with data protection principles.

Gatekeepers

EU legislation, in particular in the DMA and the DA proposal, has identified a figure called Gatekeepers[201]. Gatekeepers, in relation to Data Spaces, are, among others, those that provide cloud computing services[202] employing extreme economies of scale, with very powerful network effects, high connectivity capacity between users, creating a significant degree of dependence that can undermine the contestability of services and, therefore, the fairness of the commercial relationship[203]. Some of the undertakings considered Gatekeepers control entire ecosystems of platforms in the digital economy and increase the possibility that underlying markets may not function well[204]. Gatekeepers have zero marginal costs of adding users and provide integrated solutions for the implementation and exploitation of data lakes ranging from tools for real-time data loading from IoT systems[205], through multiple solutions for extraction, transformation, cataloguing, presentation, etc., to processing with artificial intelligence tools and pre-defined biometrics.

The DMA regulates the performance of Gatekeepers from the point of view of how they can affect the detriment of prices, quality, fair competition, choice and innovation in the digital sector[206], establishing rules to ensure contestability and fairness of markets in the digital sector[207]. The DA proposal states in Recital 36 that '*the inclusion of such gatekeeper undertakings as beneficiaries of the data access right would not be necessary to achieve the purpose of this Regulation and thus would be disproportionate in relation to the data holders subject to such obligations. This means that an undertaking providing core platform services*

[201] Definition in the article 2(1) of the DMA
[202] The DMA applies the definition of Gatekeeper to other services beyond cloud computing, but in this text, we will focus on the latter.
[203] Recital 2 of the DMA
[204] Recital 3 of the DMA
[205] Internet of Things
[206] Recital 4 of the DMA
[207] Recital 7 of the DMA

that has been designated as a gatekeeper cannot request or be granted access to users' data generated by the use of a product or related service or by a virtual assistant based on the provisions of Chapter II of this Regulation. An undertaking providing core platform services designated as gatekeeper pursuant to Digital Markets Act should be understood to include all legal entities of a group of companies where one legal entity provides a core platform service. Furthermore, third parties to whom data are made available at the request of the user may not make the data available to a designated gatekeeper. For instance, the third party may not sub-contract the service provision to a gatekeeper. However, this does not prevent third parties from using data protection services offered by a designated gatekeeper. This exclusion of designated gatekeepers from the scope of the access right under this Regulation does not prevent these companies from obtaining data through other lawful means.

With regard to the development of a digital regulation governing the market balance between SMEs and Gatekeepers, it should also be asked what effect such Gatekeepers may have in the framework of the Data Spaces in relation to compliance with the principles, rights and obligations of the GDPR. In this respect, the EDPS has stated that *'processing of data for the public good should not create or reinforce situations of data oligopoly (dependency of the public sector, SMEs, etc., on few powerful IT companies, so-called Big Tech). This is also relevant from a data protection perspective, since monopolies and oligopolies create situations of users' lock-in and ultimately restrict the possibility for individuals to exercise effectively their rights.'*[208]

Impact of gatekeepers on data protection measures

For example, in a case where legal, organisational and technical safeguards are to be implemented for the processing of high-impact data (e.g., health data[209]) in a Data Space that includes pseudonymisation. To this end, Data Holders could be required to have an ad-hoc space for extracting the information to be shared, different from the one where the information is stored for exploitation. In that ad-hoc space, the selected information would be accessible to a Data Space Mediator who would perform the pseudonymisation and who would store the additional reidentification information. When a Data User needs the information for a specific use case, the Data Space Mediator will extract the data set and transfer it to the Data User in a Secure Processing Environment, as set out in Recital 54 of the EHDS in the framework of a secondary use of the data, for the processing.

[208] Paragraph 26 of the '*Opinion 3/2020 on the European strategy for data (EDPS) [16 June 2020]*'
[209] Be it the case of the implementation of the Código de Conducta regulador del tratamiento de datos personales en el ámbito de los ensayos clínicos y otras investigaciones clínicas y de la farmacovigilancia [only available in Spanish]

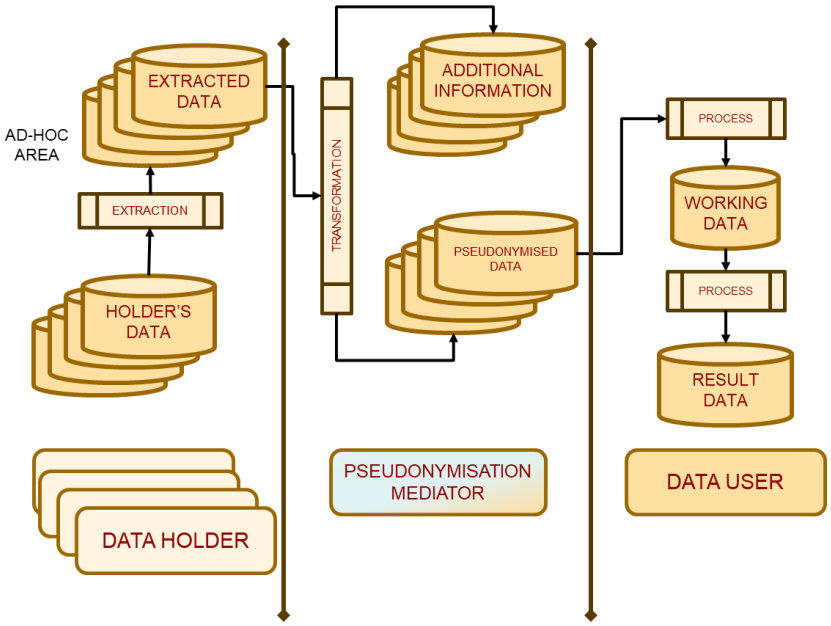

Figure 26: Diagram of implementation of pseudonymisation guarantees by physical separation of the interveners

In the above diagram, the physical separation between Data Holder and Data Space Mediator, and the independence with the Data Users' systems is established as an additional guarantee for the implementation of the GDPR safeguards.

However, let us assume that all interveners select the same Gatekeeper as data processor (or sub-processor depending on the role of the interveners). In that case, a large part of the organisational and technical measures could be compromised, and the legal measures binding the interveners would lose some of their effectiveness. For example, data separations between interveners would be diluted as they would effectively reside with the same processor, regardless of the security measures employed.

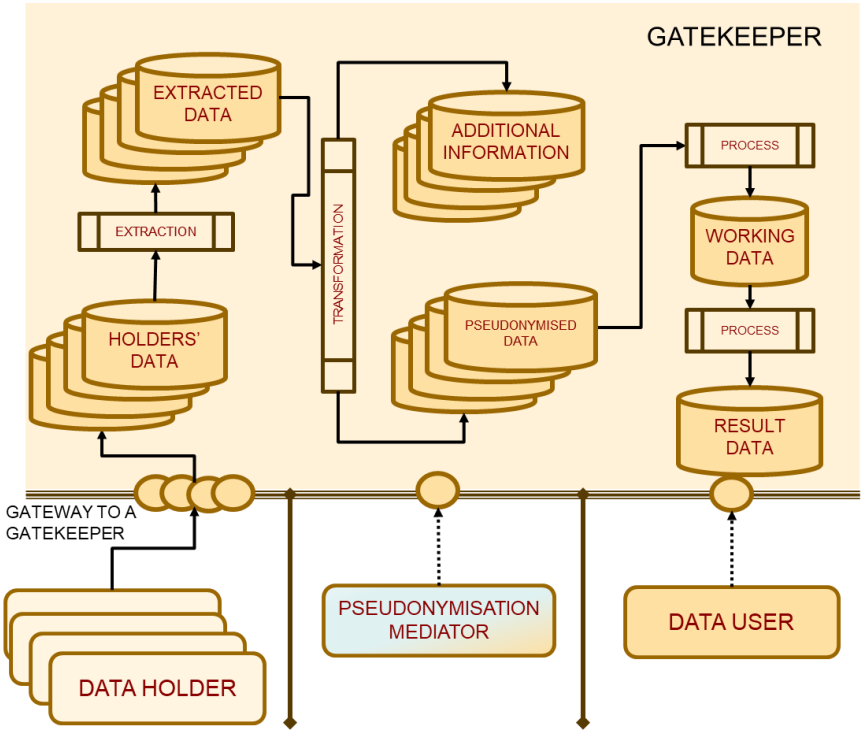

Figure 27: Diagram of implementation of pseudonymisation guarantees by physical separation of interveners when they share the same Gatekeeper.

In the figure above, safeguards based on physical separation of reidentification information could be compromised when implemented on Gatekeepers. Of course, these will be subject to legal safeguards[210], but these are the same safeguards that in the initial example have not been considered sufficient for Data Space Mediators or to Data Users.

This situation can arise even when several Data Spaces are distributed over the of few Gatekeepers.

[210] Set out in Article 5(2)(b) of the DMA 'combine personal data from the relevant core platform service with personal data from any further core platform services or from any other services provided by the gatekeeper or with personal data from third-party services'.

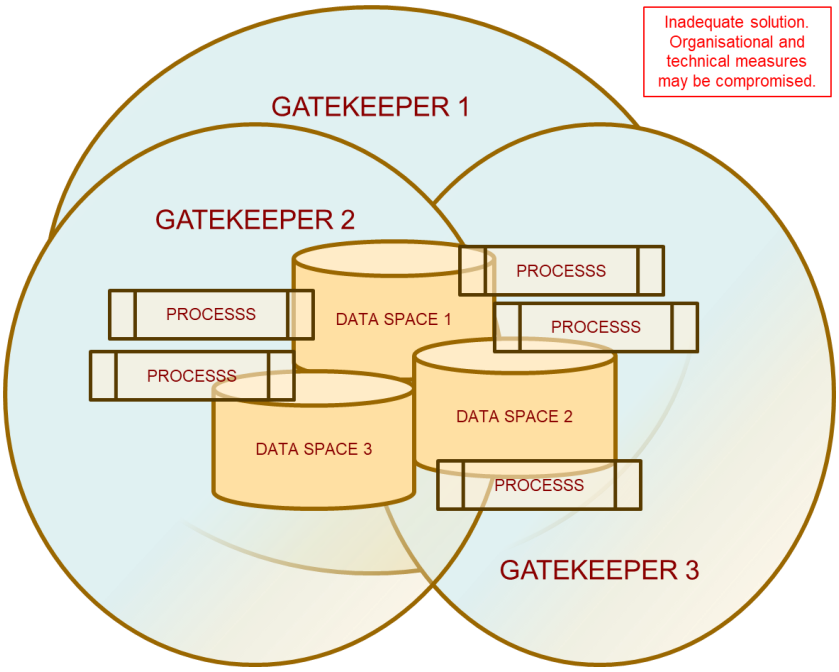

Figure 28: Distribution of various Data Spaces on the services of few Gatekeepers

On the other hand, such an implementation could lead to a federation of *data lakes* rather than a Data Space.

Risk management in the selection of processors/sub-processors

One of the main virtues of the GDPR is its flexibility when it comes to adapting to new technological contexts, as the principle of proactive responsibility requires data controllers to go beyond compliance with the minimum requirements directly demanded in the text of the regulation and to make processing conditional on adequate management of the risks to the rights and freedoms of Data Subjects.

With regard to the selection of processors/sub-processors, and in particular Gatekeepers, in addition to the possible loss of the effectiveness of certain safeguards, we may face new risks. Such risks may arise from a higher concentration of data, the massive impact of potential personal data breaches, the subjection to regulations of third parties and a reduced ability to enforce and control by controllers and Supervisory authorities (e.g., when accessing traceability information or the impossibility of accepting or rejecting sub-processors). To this must be added the risks to availability resilience that may be posed by unilateral changes in service conditions, best-effort[211] offers in service quality conditions, modification and discontinuity of services following policies that do not meet the needs of controllers, executive action by third country authorities[212], withdrawal of service due to geopolitical events [213], etc.

All these risks need to be assessed and, as a general recommendation to reduce the risk, it is advised to limit as much as possible the use of processors for the storage or bulk traffic

[211] The non-existence of real guarantees of availability levels.
[212] We recall the Megaupload case [only available in Spanish]
[213] Companies pulling back from Russia over the war in Ukraine | CNN Business

of non-anonymised personal data (see also the recommendation to perform a reidentification analysis when many anonymised data sets are accumulated in a Mediator). The level of risk of the use of processors is likely to increase when:

- A processor is used to provide services exclusively for that Data Space.
- A processor is used to provide services to several Data Spaces, but exclusively for Data Spaces.
- A processor is used to provide services to a single Data Space and to provide other ICT services on its own account or as a processor for third parties.
- A processor is used to provide services to multiple Data Spaces and to provide other ICT services on its own account or as a third-party processor.

In these last three cases, where the impact of the above risks would be higher, and which are common to public clouds or Gatekeepers, they are considered to be of a high level of risk (see section 'VI.B. Risk Management and Data Protection Impact Assessment - High Risk') that is very difficult to mitigate. In these cases it should be considered that prior consultation with the data protection authority will most likely be required[214].

D. TRACEABILITY, TRANSPARENCY AND THE EXERCISE OF RIGHTS

Date traceability is the ability to know the entire data lifecycle: the exact date and time of extraction, when, where and by whom it was transformed, and when, where, by whom and for what purpose and legitimacy it was uploaded, used or downloaded from one environment to another destination. This process is also known as Data Linage.

Traceability can serve purposes other than data protection, such as implementing the monetisation inherent to Data Spaces, also to support the principle of sovereignty of Data Holders, or other requirements for the control of intellectual and industrial property, contract completion, informing the Data Subject of the results of the data processing[215], etc.

Traceability for data protection

Transposing ENISA's conclusions, *"In the era of Big Data, 'traditional' information and consent mechanisms do not provide adequate transparency and control"*[216] for data users. It should be recalled that one of the four pillars of the European Data Strategy[217] is to support *"individuals in exercising their rights in relation to the use of the data they themselves generate. They can be empowered to have control over their data through tools and means that allow them to decide at a more detailed level what is done with their data ('personal data spaces')"*.

[214] Article 36(1) of the GDPR

[215] Recital 44 of the EHDS proposal cites '*Natural persons should be able to access the results of different research projects on the website of the health data access body, ideally in a easily searchable manner'.*

[216] Conclusions of the ENISA document 'Privacy by design in big data. An overview of privacy enhancing technologies in the era of big data analytics European Union Agency for Cybersecurity (ENISA) [17 December 2015]'.

[217] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European Data Strategy (COM (2020) 66 final) [19 February 2020]

In order to bring the control mechanisms in line with the reality of processing in the framework of Data Spaces, data traceability could fulfil the following objectives from a GDPR point of view:

- Comply with the transparency requirements to data subjects of the GDPR.
- Enable the effective exercise of data subjects' rights, in particular the management of consent.
- Enable to exercise the obligations of the controller (e.g., to ensure the principles of restriction of processing, purposes compliant with the legal bases or of processors/sub-processors).
- To allow Supervisory Authorities to exercise their powers in accordance with Article 58(1) of the GDPR.

These objectives are complementary but different, and the information that needs to be collected to enable the full exercise of these functions will vary, for example, between that which needs to be made available to the Data Subject and that which needs to be made available to the Supervisory Authorities.

Data traceability requires identifying roles and implementing access control and access logging policies. In the case of Data Spaces, due to their special nature, access authorisation must establish access control and logging policies that allow traceability at the level of the individual user, and not only at the level of organisations or departments[218]. The maintenance of a log of accesses as well as of actions performed during access to the Data Space is recommended both to meet the above objectives and to implement the obligations of Article 32 of the GDPR, to comply with the obligations of data intermediation services[219] or the transparency obligations of recognised data management organisations for altruistic purposes[220]. For example, the latter will be obliged to keep a complete and accurate record of:

- all natural or legal persons who have been allowed to process data held by that recognised data management organisation for altruistic purposes, and their contact details;
- the date or duration of the processing of personal data or the use of non-personal data;
- the purpose of the data processing as stated by the natural or legal persons to whom such processing is permitted;
- any fees paid by the natural or legal persons carrying out the data processing.

Traceability of data sets

The availability of traceability mechanisms in data sets is a data protection management measure related to the traceability of personal data themselves.

[218] For example, in the EHDS proposal, in the Article 37 '*Tasks of health data access bodies*' in letter k proposes '*(k) maintain a management system to record and process data access applications, data requests and the data permits issued and data requests answered, providing at least information on the name of the data applicant, the purpose of access, the date of issuance, duration of the data permit and a description of the data application or the data request*'.
[219] Article 12(o) of the DGA.
[220] Article 20(1) of the DGA

In copyright protection, various techniques, such as watermarking or fingerprinting, have been used to detect and prevent illicit dissemination of digital content. The use of these techniques precedes the digitisation of information and can be used at the level of the data set or at the level of elements within the data sets. Although in the digital world they have been used mainly to control the dissemination of audio, video, integrated circuits, etc., there are also specific developments for databases[221].

In the event that the Data Spaces envisages processing operations where the dissemination of personal data is indispensable, it could be of great interest to consider systems for managing the traceability of the dissemination of data, including watermarking or fingerprinting.

Transparency

With regard to transparency to Data Subjects, the GDPR establishes minimum obligations in Chapter III. The Data Holder must implement the obligations of transparency and exercise of rights to Data Subjects and implement the obligations of information to Data Subjects prior to carrying out any further processing[222]. However, these obligations can be extended in cases where transparency mechanisms are an appropriate measure to mitigate the high risk we face in the framework of a Data Space or as a guarantee to carry out the purpose compatibility analysis.

In the cases where data are disclosed to third parties, the third parties must inform the Data Subjects in compliance with Article 14 of the GDPR on '*Information to be provided where personal data have not been obtained from the data subject*'. Information obligations where personal data have not been obtained from the data subject may fall within the exceptions set out in Article 14(5) of the GDPR. However, the Data Space has to be designed so that it does not fall under Article 14(5)(b) of the GDPR as a matter of course simply because the need for compliance with these obligations has not been taken into account from the design stage as far as technically feasible.

In the case of data intermediation services, when offering services to Data Subjects, they shall inform and, where appropriate, advise Data Subjects in a concise, transparent, intelligible and easily accessible manner about the intended uses of the data by Data Users and the general conditions applicable to such uses before Data Subjects give their consent[223]. In the case of organisations recognised as data management organisations for altruistic purposes, they shall inform Data Subjects or Data Holders in a clear and easily understandable manner prior to any processing of their data under the conditions of Article 21 of the DGA.

In the framework of the Data Space, the conditions for allowing further processing of personal data should be made public[224]. Data Space Mediators and Enablers, in relation to Data Users, may provide tools to exercise this principle of transparency. It could be envisaged that, in the case of communications to Data Users, there should be a

[221] CiteSeerX (psu.edu)
[222] Article 13(3) of the GDPR
[223] Article 12(m) of the DGA
[224] Paragraph 19 of the 'Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]'

commissioning contract for this specific processing between the Data User as controller and the Data Space Mediator, who would act in this case as processor.

Inventory of processing activities

Likewise, in relation to processing operations within the framework of the Data Space, the parties listed in article 77(1) of the LOPDGDD, in particular Public Administrations, are obliged to make public an inventory of their processing activities, accessible by electronic means, containing the information established in article 30 of the GDPR and its legal basis.

Exercise of rights

In the framework of the Data Space, the Data Subject must have the possibility to object where the processing is based on legitimate interest[225] or public interest[226] and the other rights set out in the GDPR. In addition, in the case of data intermediation services, they may object to their data being converted into another format (which is itself a processing operation), and must be given the opportunity to do so, unless such conversion is required by Union law[227].

In the case that the Data Holders are Public Administrations, the use of the Citizen File as an Enabling instrument becomes especially relevant as it is an additional transparency tool. To this extent, the improvement of the possibilities of the Citizen File must be contemplated with regard to receiving sufficient information on the purpose of new processing operations, facilitating communication channels to resolve doubts or being able to obtain consent or refusal in some way in cases where it is based on consent.

Consent management

The Data Subject must also have the possibility to give, modify or withdraw his or her consent to the processing of his or her personal data for purposes other than those for which they were processed. When consent is the chosen legitimising basis for processing in a Data Space, this consent must be a freely given, specific, informed and unambiguous expression of will. In particular, the requirements of Article 7 of the GDPR, which is further developed in Recitals 32, 42 and 43 of the GDPR and in more detail in the Guidelines 5/2020 del EPDB. Consent should not be considered as freely given when the data subject does not have a genuine or free choice or cannot refuse or withdraw his or her consent without detriment[228]. The harm to the Data Subject includes the social or emotional pressure that may be exerted in certain situations and contexts.

In particular and among others, the DGA establishes the framework for the altruistic transfer of both personal and business data[229]. In the case of personal data, the altruistic transfer of data would be based on the Data Subjects' GDPR consent. Data Mediators offering services to Data Subjects should inform and, where appropriate, advise in a

[225] Article 21(1) of the GDPR
[226] Except, as provided for in Article 21(6) of the GDPR, where they are processed for scientific or historical research or statistical purposes in accordance with Article 89, section 1 of the GDPR
[227] Article 12(d) of the DGA.
[228] Recital 42 of the GDPR
[229] Chapter IV and Recitals 50, 51 and 52 of the DGA

concise, transparent, intelligible and easily accessible manner about the intended uses of the data by Data Users and the general conditions applicable to such uses before the data subjects provide their consent[230]. It would also be necessary to specify the territory of the third country in which the data are intended to be used[231].

The Data Space shall provide tools for obtaining the consent of Data Subjects as well as for withdrawing their consent. Withdrawal of consent shall not affect the lawfulness of the processing based on consent prior to its withdrawal.

In the provision of consent, it will be necessary to determine mechanisms to establish the granularity of such consent, in terms of categories of data, categories of processing and categories of recipients. The granularity should allow for expressing and respecting the wishes and rights of the data subjects[232]. In this respect, the adoption of both "white" and "black" lists[233] can be envisaged, allowing for a precise definition of preferences reflecting the moral or ethical values of the Data Subjects.

A granular consent requires the establishment of resources for its management in the Data Space and these resources must be defined from the design. Poorly designed granularity criteria can lead to future problems in its application, with doubts in its application for specific cases and lack of effectiveness, hence the importance of its good definition from the design of the Data Space. It is also important to consider that no matter how much the criteria are refined, there will be doubts about their application in specific processing activities[234]. In this circumstance the Data Mediator, or whoever acts as data controller, could seek the advice of the DPO, to include the data or exclude the data. Other strategies could include contacting the data subject to clarify whether the purposes are compatible or not, determining whether the data subject would consent to the particular purpose of the processing in question, or consulting relevant third parties, such as data protection authorities. In the case of re-use of personal data held by public sector bodies, no contact details should be provided to enable re-users to contact the Data Subjects directly[235].

Consent, and in particular the altruistic transfer of data, does not waive the fundamental rights of the data subject or exempt data controllers from complying with principles, rights and obligations, but merely provides a legitimate basis for the processing of such data. In particular, and as stated above, where processing has been based on consent, they may only be processed for a different purpose if the controller requests a specific consent for that other purpose or if the controller can demonstrate that it relies on a Union or Member State law to safeguard the purposes referred to in Article 23 of the GDPR[236].

---

[230] Article 21(1) of the DGA.
[231] Article 21(6) of the DGA
[232] Paragraph 19 de la 'Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]'
[233] Section 4.3.1 of the document 'ENGINEERING PERSONAL DATA SHARING, Emerging Use Cases and Technologies. European Union Agency for Cybersecurity (ENISA). [January 2023]'.
[234] Section 4.3.1 of the document 'ENGINEERING PERSONAL DATA SHARING, Emerging Use Cases and Technologies. European Union Agency for Cybersecurity (ENISA). [January 2023]'
[235] Recital 15 of the DGA
[236] Paragraph 53 of the document 'Guidelines 1/2020 on processing of personal data in the context of connected vehicles and mobility related applications (EDPB) [9 March 2021]'.

Therefore, the need to implement an agile mechanism to manage a lifecycle of consent may be considered. A Data Subject may decide, at any time, to arbitrarily modify its data processing demands, or to revoke consent to processing or restrict processing on certain data users. Therefore, a Data Mediator also has to permanently monitor the ongoing processing instances of each Data Subject's data, in order to respond to a Data Subject's change of mind within a reasonable period of time[237].

Finally, it should be noted that lawful data processing, even when measures such as anonymisation are applied, might not solve all ethical problems, such as, for example, those related to personal objections to certain private sector stakeholders (e.g. pharmaceutical, insurance, etc.)[238].

E. RETENTION OF PERSONAL DATA AND LIMITATION OF PROCESSING

In accordance with the principle of time limitation[239] in the processing of personal data, personal data to which access is given by Data Space interveners shall be kept for no longer than is necessary for the purposes of the processing. The governance mechanisms should allow for the establishment by the Access Request Supervisor of retention periods as well as management thereof.

The application of this principle, and that of limitation of the purpose of processing[240], makes it essential to implement the aforementioned traceability mechanisms for communicating and executing the restrictions on data processing. Management includes the possibility of an update of the conditions of consent or limitation of processing by the Data Subject, their execution by Mediators and Users, together with an active notification addressed to the Data Subject.

Data Space governance mechanisms should establish guidelines and tools for compliance with these principles, as well as reviews or audits by competent authorities to ensure that the Data Space complies with the rules of effective enforcement of rights, automatic tools to detect or ensure that a local copy of the data in the Data User for further use is not possible, or others.

F. ANONYMISATION AND REIDENTIFICATION

The concept of anonymization has been developed by the European Data Protection Committee, the Agencia Española de Protección de Datos (Spanish Data Protection Agency) and other entities in various guides, technical notes and tools[241]. In turn, anonymization has been addressed in the use cases in the previous chapter.

Recital 15 of the DGA states that in the case of re-use of data from public sector bodies and where access to personal data is implemented by transmission of personal data, the personal data must be anonymised. However, when the provision of anonymised or

[237] Section 4.1 del of the document 'ENGINEERING PERSONAL DATA SHARING, Emerging Use Cases and Technologies. European Union Agency for Cybersecurity (ENISA). [January 2023]'.
[238] Paragraph 20 of the document *'Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]'*
[239] As set out in article 5(1)(e) of the GDPR.
[240] Article 5(1)(b) of the GDPR
[241] Consult the section on Anonymisation and Pseudonymisation of the Innovation and Technology Area of the AEPD website.

modified data does not meet the needs of the Reuser of such data, it leaves open the possibility of using Secure Processing Environments, which have been described in the previous chapter.

The Reuser has an obligation of confidentiality prohibiting the disclosure of any information that jeopardises the rights and interests of third parties. In addition, Re-users shall be prohibited from re-identifying any data subject to whom the data relates and shall be obliged to take technical and operational measures to prevent re-identification[242]. However, the possibility to conduct research on anonymisation techniques[243] where these involve evidence of re-identification shall not be prohibited.

The analysis of a possible re-identification of Data Subjects will always have to be present and will have to be carried out in a way between the parties involved in a processing within the framework of the Data Spaces. Anonymisation processing is not a trivial process and when data comes from several sources, the risk of re-identification increases. The Data Mediator must employ suitable professionals, knowledgeable in state-of-the-art anonymisation techniques, and also experienced in non-personal data re-identification attacks. It has to be determined by analysis and practical evidence that re-identification of the dataset is not possible, considering worst case conditions, such as re-identification attempts by insiders or outsiders, with access to ancillary data, including data available by illegal means, by court orders or by intelligence agencies, and considering that adequate resources are available and extrapolating the possible evolution of known techniques. If all or part of the dataset can be re-identified under these conditions, there is no risk of re-identification, the dataset is simply not anonymised.

However, a residual probability of re-identification must always be assumed. This residual probability means accepting that there is no such thing as total and absolute infallibility. In any case, the controller can be required to do as stated in the previous paragraph: apply proactive accountability with appropriate measures to ensure compliance taking into account the nature, context, scope, purposes and risks to rights and freedoms, as well as review and update, such as incorporating re-anonymisation measures.

Finally, note the limitations to international transfers of non-personal data that could be established by law in case of risk of re-identification, which are briefly discussed in the section on International Transfers[244].

G. ENRICHMENT OF DATA SETS

In relation to the interaction between Mediators and the risks of reidentification[245] one question to be raised would be the possible impact of the enrichment of data sets, through different data sources and the way they are managed, especially when they may go beyond the purpose of a Data Space.

[242] Article 5(5) of the DGA
[243] Recital 8 of the DGA
[244] Article 5(13) of the DGA
[245] See case in the AEPD blog article entitled 'Anonymization III: The risk of re-identification | AEPD [February 2023]

Diversity of data sources

In the framework of the Data Space, it is necessary to consider to what extent it is possible to manage the aggregation of data from different sources, either by Data Holders or by Data Mediators, including in the framework of processing in Secure Processing Environments.

Likewise, it must be ensured how to guarantee that the data conform to what is foreseen for a given Data Space and do not involve data from different sources and from the same Data Subjects, or the possible limitation on Data Users not to cross-reference information from different Data Spaces when it is not considered lawful or appropriate.

In this respect, the use of data protection architectures from the design and PET technologies PET described in previous sections are of great importance.

Unrestricted access sources

The fact that personal data are accessible without restriction, via the Internet or other means, is not a basis for the legitimisation of the processing of personal data.

The application of scraping techniques[246] that do not process personal data does not fall under data protection law. However, as noted above, it will have to be analysed that the joint processing of different non-personal data may lead to the identification of natural persons.

H. INTERNATIONAL TRANSFERS OF DATA

International transfers of personal data are subject to the provisions of Chapter V of the GDPR. Beyond strict compliance, such transfers may present risks to the rights and freedoms of Data Subjects, for example in terms of non-legitimate access to personal data and ineffective monitoring of the same data[247]. Such risks[248] will be higher for some types of processing and data, and need to be assessed and managed in the framework of the governance and design of the Data Space. The EDPB and the EDPS interpret[249], in the light of the CJEU rulings[250] [251], that European law requires, in certain specific circumstances, to

[246] Web scraping is a technique used by software programmes to extract information from websites. Usually, these programs simulate the browsing of a human on the Internet.
[247] Paragraph 105 and 108 of the document 'EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space [12 July 2022]'
[248] Footnote 15 of the document "Guidelines on Data Protection Impact Assessment (DPIA) and for determining whether the processing is 'likely to involve a high risk' for the purposes of Regulation (EU) 2016/679 (EDPS) [11 April]"
[249] Paragraph106 of the document "EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space [12 July 2022]"
[250] Judgment of the Court (Grand Chamber), 8 April 2014, *Digital Rights Ireland Ltd*, joined Cases C-293/12 and C-594/12; para 68. See also the Opinion of Advocate General Cruz Villalón delivered on 12 December 2013 in the same case at para 78 and 79, noting that the absence provision that lays down the requirement to 'store the data to be retained in the territory of a Member State, under the jurisdiction of a Member State', 'increases the risk of use which is incompatible with the requirements resulting from the right to privacy' and 'considerably increases the risk that such data may be accessible or disclosed in infringement of that legislation'.
[251] Judgment of the Court (Grand Chamber) of 21 December 2016, Tele2 Sverige AB v Post- och telestyrelsen and Secretary of State for the Home Department v Tom Watson and Others, joined Cases C-203/15 and C-698/15, para 122. See also the opinion of Advocate General Saugmandsgaard Øe delivered on 19 July 2016, Tele2 Sverige AB v Post- och telestyrelsen and Secretary of State for the Home Department v Tom Watson and Others, joined Cases C-203/15 and C-698/15, para 239 to 241

impose the obligation of storage in the EU in order to mitigate such a risk, for example in relation to the European Health Data Space[252].

This aspect must be taken into account when implementing cloud services. Diligence must take into account not only the location of servers in the EU, but also the collateral processing of such data that may be carried out by these services[253] for multiple reasons and that may end up materialising in international transfers. There are also implementations that may involve international data transfers that are less obvious. For example, when certain SDKs[254] are integrated in the development of mobile and web applications to add a wide range of functionalities, such as high-availability databases, web and app analytics, payment gateways or user interface features[255] [256].

An obligation to store personal data in the EU does not exclude transfers to third countries or international organisations. Indeed, it is possible to reconcile a general requirement to store personal data in the EU with specific transfers that allow compliance with Chapter V of the GDPR (e.g., in the context of scientific research, disbursement of care or international cooperation)[257].

In any event, it is essential to avoid an inconsistent and fragmented approach across the EU with regard to the criteria for international data transfers[258]. Different degrees of protection of Data Subjects in different countries would impede the free flow of data within the EU, as interveners would be reluctant to allow access, when it involves communication of data, from EU countries that do not guarantee the protection of personal data. For this reason, in the event that there is no accepted global criterion for international transfers, it is recommended that the most protective criterion be implemented at the national level so that the entities in our country have the best consideration for compliance.

In the case of re-use of data from public sector bodies, the DGA establishes in its Chapter VII 'International access and transfer' and in Recitals 21 to 24 conditions for transfers of non-personal data. Although these conditions would in principle be outside the competence of the GDPR, there is a reference to the 'protection of privacy and personal data' when restricting the transfer of 'protected data'[259] or 'highly sensitive non-personal data'[260], leaving it to the national or European legislator to restrict the transfer of certain

[252] Paragraph 111 of the document 'EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space [12 July 2022]'
[253] These can range from maintenance operations, remote monitoring from outside the EU, need for compliance with third country regulations, etc.
[254] *Software Development Kit*. It is a set of tools typically provided by the manufacturer of a hardware platform, or an operating system or programming language, which facilitate the development of new applications for the specific product and its application environment.
[255] EDPS sanctions the European Parliament for illegal EU-US data transfers - among other violations (Stripe payment gateway and Google Analytics, link to decision at NOYB)
[256] Google Fonts case, a German court fined a website because when connecting to Google Fonts to download a text font it makes a connection to a US server and that implies that the IP (personal data) is transferred outside the EEA. German Court Fines Website Owner for Violating the GDPR by Using Google-Hosted Fonts – WP Tavern
[257] Paragraph 108 of the document 'EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space [12 July 2022]'
[258] Paragraph 110 of the document 'EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space [12 July 2022]'
[259] Chapter II of the DGA
[260] Recital 24 de la DGA

categories of non-personal data held by public sector bodies where it may, inter alia, entail a risk of re-identification of anonymised non-personal data[261].

In addition[262], the public sector body, natural or legal person who has been granted the right to re-use data of certain categories of protected data (in this case personal data[263]) held by public sector bodies, the data intermediation service provider or the recognised non-profit data management organisation shall inform the Data Holder concerned that an administrative authority of a third country has requested access to his or her data, before acting on such request, except in cases where the request serves law enforcement purposes and as long as necessary to preserve the effectiveness of relevant law enforcement activities.

I. GOVERNANCE, DATA PROTECTION POLICIES, PROCEDURES AND CODES OF CONDUCT

One of the objectives of Data Spaces is to establish a data governance framework. In fact, it has been mentioned throughout the document and is considered of vital importance in order to implement data protection by design. This section is included at the end of the text because this governance framework must include many of the obligations and recommendations developed in the text to ensure compliance with the GDPR.

Particularly, Article 24(2) of the GDPR with regard to the responsibility of the controller provides that *'Where proportionate in relation to processing activities, the measures referred to in paragraph 1 shall include the implementation of appropriate data protection policies by the controller*". The Data Space is proportionally suitable definition and implementation of data protection policies in the data governance framework.

The EDPB and EDPS have already stated, in the framework of the EHDS, that 'success will also depend on the establishment of strong data governance and effective safeguards for the rights and interests of individuals that are fully compliant with the GDPR'[264].

This data protection policy, to be applied to all those involved in the Data Space, must state how the principles and rights set out in the data protection regulation and the guidelines in this document are to be implemented in a concrete, practical and effective manner. The data protection policy does not set out the 'what', which is already developed in the standard, but should describe the 'how' to ensure and be able to demonstrate compliance with the GDPR. Therefore, a data protection policy for processing in the framework of a Data Space should contain references to, inter alia:

1. The involvement of DPOs and data protection advisors in the design of Data Spaces and the processing operations with them.
2. The procedures for authorising the processing of personal data in the Data Space.
3. The precise definition of the purposes of the processing.

[261] Article 5(13) of the DGA
[262] Article 31(5) of the DGA
[263] Article 3(1)(d) of the DGA
[264] EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space [12 July 2022]

4. The establishment of the legal bases for processing.
5. The determination of controller/processor/sub-processor roles for each of the interveners.
6. Risk management for the risks and freedoms of data subjects, including, where appropriate, a DPIA, coordinated between the interveners in the processing.
7. The management of the involvement of several Data Space Mediators in a data processing operation to ensure compliance with the GDPR. In the previous case, when the Data Mediators belong to different Data Spaces, the management and coordination between them should be considered.
8. The definition of data use licenses that include legal safeguards for the management of risks to the rights and freedoms of Data Subjects.
9. The management and limitations of the enrichment of datasets.
10. Processes to determine, and where appropriate remedy, possible reidentification of non-personal or anonymised data sets from different sources.
11. In the case of consent-based processing, the establishment of criteria for the granularity of consent, and resources for the management of the consent lifecycle, in particular, for the modification of consent or its withdrawal.
12. Procedures for resolving doubts about the application of granular consent to specific processing.
13. Guarantees that personal data available in the Data Space are not collected unlawfully or used for purposes that were not originally intended, are disproportionate or lack an adequate legal basis[265].
14. Publicise the conditions for allowing further processing of personal data[266].
15. Publicise the results of the DPIAs.
16. Publicise the anonymisation mechanisms used.
17. The application of the principles of data minimisation in the access to personal data in the Data Space, of data protection by design and by default when deploying the Data Space architecture, as well as in the implementation of the different use cases, introducing limitations on the data retention periods for each of the interveners, if applicable.
18. The technical and organisational procedures guaranteeing security for the protection of the rights and freedoms of individuals.
19. Defining the requirements of Secure Processing Environments and Trusted Execution Environments.
20. The procedure to ensure that the processing in a Secure Processing Environment meets the possible requirements of conducting a DPIA, of consulting the supervisory authority under Articles 35 and 36 of the GDPR, and that the risks to the right and interests of data subjects have been found to be minimal, among others.

[265] Paragraph 19 of the document 'Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]'
[266] Paragraph 19 of the document 'Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]'

21. The conditions of interoperability ensuring the implementation of measures for compliance with data protection rules.
22. The integrated mechanism for the management of personal data breaches, which shortens reaction times, allows for real-time knowledge of incidents for all interveners, and provides effective protection of the impact for data subjects and society.
23. The application of the principle of transparency, so as to allow the Data Subject to have control over the processing of his or her personal data and traceability of to whom his or her data are being communicated and how this communication was made (request, transfer, etc.).
24. The establishment of mechanisms to ensure respect for the principle of transparency in the case of automated individual decisions, including profiling, and the right of the data subject to express his or her point of view, to challenge the decision and to obtain human intervention by the controller.
25. The establishment of mechanisms to ensure respect for the data protection rights of the Data Subject (data subject's right of access, right of rectification, right of erasure, right to limitation of processing, right to data portability, right to object).
26. The establishment of mechanisms for the Data Subject to object to a processing operation that involves the reformatting of his or her personal data.
27. The establishment of conservation periods, as well as management of these periods.
28. The establishment of mechanisms to ensure respect for the digital rights of the data subject (right to digital security, protection of minors on the Internet) included in Spanish Organic Law 3/2018.
29. Determine criteria and procedures for human supervision in the process of authorising the communication of personal data between controllers where appropriate and guarantees in automated data access processes.
30. Ensure compliance with existing codes of conduct and certification programmes.
31. The establishment of mechanisms to ensure that international data transfers are based on adequacy decision or, if not, that adequate safeguards are in place.
32. Implementing activity logs and conducting audits on them to ensure accountability[267].
33. The creation of an iterative process to ensure compliance with the GDPR based on the principle of proactive accountability.
34. Having an entity in charge of the supervision of the Data Space, at least from the perspective of compliance with data protection regulations.

[267] Paragraph 26 of the document *Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]*

The course of action defined in the Data Protection Policy must be effective, efficient and enforceable, and to this end it must be reflected in the internal rules and procedures to be implemented[268].

In this respect, the importance that codes of conduct could have in the Data Space should be highlighted, as well as the role that the supervisor of the code of conduct could play in relation to the supervision of access requests.

[268] Section 'IV GOVERNANCE OF RISKS TO RIGHTS AND FREEDOMS' from the Guideline 'Risk Management and Impact Assessment in the Processing of Personal Data' from AEPD.

VII. REFERENCES

Regulatory framework:

- 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, and repealing Directive 95/46/CE (General Data Protection Regulation)
- 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).
- Proposal for a Regulation of the European Parliament and of the Council on harmonised rules for fair access to and use of data (Data Act). [23/02/2022]
- Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain legislative acts of the Union. [21/04/2021]
- Proposal for a Regulation of the European Parliament and of the Council on the European Data Space for Health. [03/05/2022]
- Directive (UE) 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 version).
- Spanish Law 37/2007 of 16 November 2007 on the re-use of public sector information (consolidated version).
- Spanish Law 34/2002 of 11 July 2002 on information society services and electronic commerce (consolidated version).
- Spanish Law 40/2015, of 1 October, on the Legal Regime of the Public Sector (consolidated version).
- Spanish Royal Decree 311/2022 of 3 May, which regulates the National Security Scheme (consolidated version).
- Spanish Royal Decree 4/2010 of 8 January 2010 regulating the National Interoperability Scheme in the field of e-Government (consolidated version).
- Spanish Technical Standard for Interoperability (TSI) with regard to cataloguing and metadata.

Guidelines and technical notes published by the Agencia Española de Protección de Datos (Spanish Data Protection Authority):

- Risk Management and Impact Assessment in the Processing of Personal Data
- Guidelines for conducting a data protection impact assessment in regulatory development
- List of the types of data processing that require a DPIA (art 35.4)
- Guidelines on Personal Data Breach Notification
- Infographic: Personal Data Breach Communication

- Guidelines to manage data breach risk in public sector bodies massive data communications
- 10 Misunderstandings related to anonymisation
- K-anonymity as a privacy measure
- Guidance and safeguards in personal data anonymisation procedures [only available in Spanish]
- Introduction to the Hash Function as a Personal Data Pseudonymisation Technique
- Guidance for customers contracting Cloud Computing services [only available in Spanish]
- Guidance for Cloud Computing Service Providers [only available in Spanish]
- 10 Misunderstandings about Machine Learning (in collaboration with the EDPS)
- Audit Requirements for Personal Data Processing Activities involving AI
- GDPR compliance of processing that embed Artificial Intelligence. An introduction

Tools of the Agencia Española de Protección de Datos (Spanish Data Protection Authority):

- Tool for the analysis of risk factors: EVALUA-RISK v2 [only available in Spanish]
- Tool to assess the personal data breach notification to the Data Protection Authority: ASESORA BRECHA [only available in Spanish]
- Tool to assess the obligation to communicate a personal data breach to the data subjects: COMUNICA-BRECHA RGPD [only available in Spanish]

Articles published by the Agencia Española de Protección de Datos (Spanish Data Protection Authority):

- Encryption and Privacy: Encryption in the GDPR [November 2019]
- Encryption and Privacy II: Lifespan of personal data [January2020]
- Data breach: communication to the data subject [February 2020]
- Data protection and security [April 2020]
- Encryption and Privacy III: Homomorphic encryption [June 2020]
- Anonymisation and pseudonymisation [October 2021]
- Anonymisation and pseudonymisation (II): Differential privacy [October 2021]
- Without privacy there is no cybersecurity [February 2022]
- Anonymization III: The risk of re-identification [February2023]
- When to review data protection measures [February 2023]
- AI: System vs Processing, Means vs Purposes [April 2023]
- Federated Learning: Artificial Intelligence without compromising privacy [April 2023]

Other national publications:

- Evaluation of the mobility study in Spain with Big Data technology during the state of alarm for the COVID-19 crisis management by the DPO of the Ministry of Transport, Mobility and Urban Agenda [only available in Spanish]
- Presentation of the Public Administration's Data Platform. Measure 6: Transparent management and exchange of data. General Secretariat for Digital Administration. Ministry of Economic Affairs and Digital Transformation [only available in Spanish]
- Article published by the Spanish Data Office "Tool to elaborate se cases in data spaces [28/12/2020]"
- Article published by the Spanish Data Office "The IDS-RAM reference architecture model and its role in data spaces [26/04/2022]"
- Article published by the Spanish Data Office "Gaia-X and European data spaces [28/04/2022]"
- Article published by the Spanish Data Office "Features for the creation of data spaces [05/07/2022]"
- Article published by the Spanish Data Office "What are the main elements of a data space? [20/10/2022]"
- Article published by the Spanish Data Office "X-ray of the national Tourism dataspace: Challenges and opportunities for the tourism sector [07/02/2023]"
- Article published by the Spanish Data Office "UNE specifications – Government, management, and data quality [31/03/2023]"
- Spanish Data Office's tool for developing use cases in data spaces [only available in Spanish]
- Data Sharing scenarios. Francisco Javier Esteve Pradera, june 2022 – Bulletin nº 91 [only available in Spanish]

Publications of the European Data Protection Committee and the Article 29 Data Protection Working Party:

- WP 169 Opinion 1/2010 on the concepts of 'controller' and 'processor' (Article 29 Data Protection Working Party) [16 February 2010]
- WP 211 Opinion 01/2014 on the application of necessity and proportionality concepts and data protection within the law enforcement sector (Article 29 Data Protection Working Party) [27 February 2014]
- WP 216 Opinion 05/2014 on anonymisation techniques (Article 29 Data Protection Working Party) [10 April 2014]
- WP 248 Guidelines on data protection impact assessment (DPIA) and for determining whether processing is 'likely to involve a high risk' for the purposes of Regulation (EU) 2016/679 (Article 29 Data Protection Working Party) [4 October 2017]
- Guidelines 5/2020 on consent within the meaning of Regulation (EU) 2016/679 (EDPB) [4 May 2020]

- Guidelines 4/2019 on Article 25 Data protection by design and by default (EDPB) [20 October 2020]
- Guidelines 1/2020 on processing of personal data in the context of connected vehicles and mobility related applications (EDPB) [9 March 2021]

Joint publications of the European Data Protection Committee and the European Data Protection Supervisor:

- Joint Opinion 3/2021 of the EDPB and the EDPS on the Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) [10 March 2021]
- EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space [12 July 2022]

Publications of the European Data Protection Supervisor:

- Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit (EDPS) [11 April 2017]
- EDPS Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the protection of personal data [19 December 2019]
- Opinion 3/2020 on the European strategy for data (EDPS) [16 June 2020]
- Preliminary Opinion 8/2020 on the European Health Data Space (EDPS) [17 November 2020]

Publications of the European Commission:

- Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union (COM(2019) 250 final) [29 May 2019]
- Communication on the Precautionary Principle (COM(2000)1 final) [2 February 2000]
- Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European Data Strategy (COM (2020) 66 final) [19 February 2020]
- COMMISSION STAFF WORKING DOCUMENT on Common European Data Spaces. European Commission (SWD (2022) 45 final) [23 February 2022]

Publications of the European Cybersecurity Agency:

- Privacy by design in big data. An overview of privacy enhancing technologies in the era of big data analytics European Union Agency for Cybersecurity (ENISA). [17 December 2015]
- RECOMMENDATIONS ON SHAPING TECHNOLOGY ACCORDING TO GDPR PROVISIONS, An overview on data pseudonymisation. European Union Agency for Cybersecurity (ENISA). [November 2018]
- DATA PSEUDONYMISATION: ADVANCED TECHNIQUES & USE CASES, Technical analysis of cybersecurity measures in data protection and privacy. European Union Agency for Cybersecurity (ENISA). [January 2021]

- DATA PROTECTION ENGINEERING, From Theory to Practice European. Union Agency for Cybersecurity (ENISA) [January 2022]
- ENGINEERING PERSONAL DATA SHARING, Emerging Use Cases and Technologies. European Union Agency for Cybersecurity (ENISA) [January 2023]

Other international publications:

- PDPC SINGAPURE: Guide to Basic Anonymisation [March 2022]
- PDPC SINGAPURE: Basic Data Anonymisation Tool [March 2022]
- What is a Data Space? Definition of the concept Data Space. White Paper 1/2022. (Gaia-x – Hub Germany) [September 2022]