The European Health Data Space

Examining A New Era in Data Protection

Edited by Santa Slokenberga, Katharina Ó Cathaoir and Mahsa Shabani

First published 2025

ISBN: 978-1-032-82288-4 (hbk)
ISBN: 978-1-032-89684-7 (pbk)
ISBN: 978-1-003-54411-1 (ebk)

Chapter 11

Federated networks and secondary uses of health data

Challenges in ensuring appropriate safeguards for sharing health data under the GDPR and EHDS

Magdalena Kogut-Czarkowska and Mahsa Shabani

DOI: 10.4324/9781003544111-14

The funder of the Open Access version of this chapter is Justitierådet Edvard Cassels stiftelse, Stockholm

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

11 Federated networks and secondary uses of health data

Challenges in ensuring appropriate safeguards for sharing health data under the GDPR and EHDS

Magdalena Kogut-Czarkowska and Mahsa Shabani

11.1 Introduction

Sharing and accessing diverse types of health data across borders play a crucial role in advancing biomedical research,[1] developing targeted therapies[2] and enhancing diagnostic tools.[3] Traditionally, the research community has relied on large, centralised repositories such as disease-specific databases to publish and exchange health data for scientific purposes. However, this centralised model faced challenges such as compliance with stringent and fragmented legal frameworks and lack of adequate incentives to share data,[4] among other obstacles. These barriers to data sharing prompted the exploration of alternative approaches.[5]

One way to address the challenges associated with traditional health data sharing is through technical solutions, such as the use of decentralised approaches that allow data holders to store data in a local hub rather than a centralised repository. Federated networks are a prominent example of such de-centralised approaches to data sharing, which would offer secure environment for data use.[6] Unlike centralised data pools, federated networks enable data holders (such as hospitals or biobanks acting as controllers of health data) to retain control over data re-use,[7] as the data remains within the secure environment of the data holder and is not duplicated. The features of federated networks seem to introduce a transformative approach to data sharing, addressing both practical and regulatory restrictions associated with the centralised approaches.

On the policy and regulatory level, the concept of data federation rapidly evolved into a new paradigm and made its way into national data sharing infrastructures[8] and multiple EU initiatives. Some examples include GAIA-X federated data infrastructure for Europe, European Genome-phenome Archive (EGA) work on decentralising genetic and phenotypic databases, which originally were a centralised resource, European Cancer Imaging Initiative and various projects supported by EU funding, among others. The imperative to facilitate and expand cross-border exchange of health data through secure, federated repositories was embedded into the 2020 European strategy for data[9] paving the way for the European flagship initiative, the European Health Data Space (EHDS).[10]

DOI: 10.4324/9781003544111-14

Despite the growing focus on de-centralised approaches, particularly federated networks, there remains a lack of comprehensive legal analysis regarding federated infrastructures concerning regulatory requirements for sharing and reusing health data for research purposes.[11] This includes relevant requirements outlined in the General Data Protection Regulation (GDPR)[12] and the EHDS.

It is crucial to first investigate, whether federated networks have been integrated and explained in the recent EU policies and regulatory frameworks for health data sharing and secondary uses, including those related to the EHDS as part of transformation in secondary uses of health data. Second, clarifications are required on the expected function of federated networks within the framework of applicable data protection regulations. In response, this chapter discusses whether and how federated networks can be considered as a form of appropriate safeguards, as recognised by the GDPR, and further supported by the EHDS. Indeed, under the GDPR, the adoption of appropriate security measures, including technical measures, is recognised by Article 89(1), which should be read in conjunction with Article 9(2)(j) when special categories of data, including health data, are to be processed for, inter alia, research purposes. For its part, the EHDS continues to support the need for safeguards and sees its role in establishing them for responsible secondary uses of health data. For this analysis, this chapter specifically examines three aspects, namely whether federated networks can be considered as a tool for data minimisation, their compliance with relevant data protection requirements, and their impact on the protection of fundamental rights and interests of data subjects.

The results of this investigation should provide invaluable insights into whether federated networks should be considered for use under the EHDS. This consideration aligns with the aim of the EHDS to introduce sufficient safeguards for the utilisation of health data for secondary purposes, including for scientific purposes.

11.2 Federated networks in the context of health data sharing

The term federated network was coined already in 1979,[13] and followed by the first conceptual model of federated architecture for office information management described in 1985.[14] With technological advancements, the concept evolved, yet at its core remaining a model consisting of multiple databases or repositories where the data is stored. This model is often contrasted with a centralised one, where the data is copied from the original location and stored in a central storage.[15] Although federated networks can be used for various purposes, in the context of health research, this concept is associated with research infrastructure which allows data users, mostly being researchers in the health field (such as developers of AI tools for healthcare), to seamlessly conduct research using data from multiple repositories. Repositories are connected by a centralised infrastructure or – at least – an agreement on data sharing or governance principles.[16] The individual repositories are sometimes called the nodes[17] or hubs.[18] The entities or persons that are responsible for the nodes are

referred to as data custodians,[19] controllers,[20] owners[21] or stewards.[22] To facilitate data use, the federated network may include an interface which allows the users to search for or access data across various nodes as if they were using a single database.[23]

Currently, numerous federated networks are designed to conduct AI model training in a federated manner, a process commonly referred to as federated learning.[24] In federated learning (also referred to as model-to-data[25]), the data stays in their place of generation or storage and are never transferred to the central location (cloud or server). The training of the AI model takes place locally in individual nodes, and the parameters of each local model are aggregated into a global model.[26] This process is sometimes depicted as 'bringing questions to data rather than moving data'.[27]

In scientific literature, the term federated network is often interchangeably used with federated repository,[28] federated data sharing,[29] federated infrastructure,[30] federated data systems,[31] federation of data[32] or simply federation[33] and others. While this concept is frequently explored from architectural, health[34] or even sociotechnical[35] perspectives, the legal sources on the topic scarce. As shown in Table 11.1, although prominent policy documents and regulatory frameworks reference federated networks or similarconcepts, a comprehensive legal analysis of the concept is absent from these documents.

11.3 Federated networks as appropriate safeguards

11.3.1 Appropriate safeguards under Article 89(1) GDPR

Compliance with Article 89(1) GDPR, read in conjunction with Article 9(2)(j), is pivotal for sharing and using personal data for health scientific research. The provisions mandate that the processing of personal data for scientific purposes adheres to appropriate safeguards, in accordance with the GDPR, to protect the rights and freedoms of the data subjects. The safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. The lack of a legal definition for the term appropriate safeguards in the GDPR, coupled with the opaque wording of Article 89(1), has drawn much criticism[51,52] and fuelled various interpretations of the requirements for establishing these safeguards.

Safeguards must comply with the GDPR, and Article 89(1) highlights two examples: pseudonymisation and anonymisation. However, no criteria for assessing other measures as potential safeguards are provided. Moreover, the provision has not been interpreted by the EU Court of Justice, and the EU regulatory guidance remains incomplete. Opinions of the EDPS and the EDPB[53] list various examples of safeguards, such as conducting a data protection impact assessment (DPIA), appointment of data protection officer (DPO), notifying a data breach without undue delay, guaranteeing data security and data minimisation through anonymisation or pseudonymisation, access limitations and obtaining the informed consent of research participants as an ethical

Table 11.1 References to federated networks in selected legal, policy and regulatory documents relating to use of personal data for scientific purposes

Type of document	*Name of the document*	*References to federated networks*
Policy document	A European strategy for data[36]	States that the data sharing infrastructures 'should support the creation of European data pools enabling Big Data analytics and machine learning, in a manner compliant with data protection legislation and competition law, allowing the emergence of data-driven ecosystems. These pools may be organised in a centralised or a distributed way'. A footnote explains: 'In the latter case the data are not moved to a central place in order to analyse them together with other data assets. The analytical tools come to the data, not the other way around. This makes it easier to keep the data secure and to ensure control over who accesses, what data, for what purposes.'
EU law or EU law proposal	GDPRData Governance Act (DGA)[37]	Terms federated or federated networks not found.
	EHDS[38]	Does not define or use of the term federated network, however the infrastructure foreseen for secondary data sharing, HealthData@EU, will by design not be a central one, but rather will a network of existing databases. The impact assessment report[39] accompanying EHDS refers to the proposed model as 'federated governance structure'. Recital 80 EHDS states that due to the sensitivity of health data, concepts such as 'bring questions to data instead of moving data' should be respected whenever possible. Article 76(2) allows registries or databases from a number of Member States which 'organise themselves into a single network of registries or databases at Union level', to designate a coordinator to ensure the provision of data from the registries' network for secondary use. Article 96(1) also mentions federated EU dataset catalogue connecting the national dataset catalogues.

Guidelines issued by EU advisory and supervisory bodies in the area of data protection and security – European Data Protection Supervisor (EDPS), the European Data Protection Board (EDBP) or the European Union Agency for Cybersecurity (ENISA)	Joint Opinion 03/2021 on the Proposal for a Regulation of the European Parliament and of the Council on European data governance Data Governance Act,[40]	The term federated networks not found.
	EDPS and EDPB opinions on EHDS[41,42]	
	Opinion on Data Strategy for Europe[43]	
	EDPS Preliminary Opinion on data protection and scientific research[44]	The term federating appears only in a footnote and pertains to a citation of EC press release, without analysing it further.
Other guidance and materials on websites of EU agencies and bodies	EDPS website[45]	Website includes a short note on federated learning[46] and on federated social media platforms,[47] however does not include an analysis on federated databases for data sharing.
	Documents published by ENISA	The term federated appears in the context of federated identity[48] and federated cloud environments.[49] Document on Engineering personal data protection in EU Data Spaces[50] identifies building blocks for ensuring personal data protection in data spaces, however does not mention federated networks.

requirement. However, the authorities do not provide a rationale for their selection, nor do they establish criteria for evaluating other potential safeguards. In 2021, the EDPB acknowledged the complexity and importance of explaining the meaning of the term and referred to future guidance for a detailed analysis.[54] This guidance has yet to be issued.

Challenges also arise when differentiating between the appropriate safeguards and the technical and organisational measures which are also mentioned in Article 89(1). Some approaches emphasize understanding safeguards through their objectives, citing the second sentence of the provision ('those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation') as an indirect definition of the safeguards. In this context, the Study on the appropriate safeguards[55] examined technical and organisational measures outlined in regulatory documents across selected EU countries. This led to a narrow understanding of the safeguards, with focus on principles of data minimisation and data security. Other scholars adapt a broader approach, built around the contextual purpose of the safeguards, which is to uphold the data subjects' rights and freedoms defined under European human rights law,[56] i.e., the European Convention of Human Rights (ECHR)[57] and the Charter Fundamental Rights of the European Union (EU Charter).[58] For example, Staunton, Slokenberga and Mascalzoni[59] argue that data protection is much more than a technical issue requiring technical solutions and point out that 'safeguards must respond to the multitude of legal, ethical and social risks that are associated with the sharing of data'. In turn, they extract informed consent (as an ethical requirement in research), independent review and oversight, accountable processes to govern access and use of samples and data, clear and transparent policies and processes, security, and training and education as possible appropriate safeguards for biobank, databank and genetic research.

These varied views lead to large diversity of cited examples of safeguards, ranging from strictly technical measures, such as encryption or data access logs,[60] to organisational or legal practices, such as implementation of procedures to support the exercise of data subjects' rights, use of anonymous or pseudonymous data, appointment of a data protection officer, feedback mechanisms on the results of scientific research, professional accreditations of persons involved in research, creating a specific supervisory body for research endeavour, or the creation of a code of conduct.[61]

It is also not clear which (if any) authority is responsible for designating the safeguards and in which manner they need to be introduced. One interpretation hinges on Recital 156 GDPR, which states, 'Member States should provide for appropriate safeguards for the processing of personal data for … scientific or historical research purposes', suggesting that it is the member states[62] that should define the appropriate safeguards in their national laws. Examples of such safeguards are provided for in the Swedish national legislation on ethical approvals[63] and Section 64 of UK's Digital Economy Act 2017

(pre-Brexit).[64] One may however argue that if researchers can only rely on EU countries to specify the safeguards, absence of national legislation would prevent making use of research exemption rules. Some thus suggest that also the parties that conduct research may design appropriate safeguards after taking into consideration the risks relevant to the processing activity.[65,66] Researchers (as controllers) can thus implement the safeguards based on the principles stemming from the GDPR (such as data minimisation, proportionality and security)[67] or on principles beyond data protection, such as ethical requirements ensuring protections for the research participant.[68] The EDPS in the preliminary opinion on data protection and scientific research also emphasises that the safeguards to be implemented by data controllers and processors should depend on the risk to individuals, however does not take a clear stance on the provenance of the safeguards.

In summary, it can be acknowledged that there is currently no consensus over the interpretation of Article 89(1). It can be argued, however, that GDPR clearly states that safeguards must remain in accordance with GDPR, and that while appropriate safeguards need to lead to implementation of technical and organisational measures, their objective is the protection of rights and freedoms of the data subjects. In turn, the technical and organisational measures imposed by the safeguards must ensure *at least* the principle of data minimisation (including necessity and proportionality), and possibly other Article 5 GDPR principles such as data security and data protection by design and by default. Hence, safeguards deemed appropriate must: (i) ensure the implementation of technical and organisational measures that support these principles, (ii) comply with the GDPR, and (iii) safeguard the rights and freedoms of data subjects.

11.3.2 Safeguards for secondary use of health data under the EHDS proposal

Notably, the EHDS does refer to appropriate safeguards as recognised under Article 89(1) GDPR. In fact, the act uses the term sufficient safeguards in Recital 4 when stating:

> Given the sensitivity of personal health data, this Regulation seeks to provide sufficient safeguards at both Union and national level to ensure a high degree of data protection, security, confidentiality and ethical use. Such safeguards are necessary to promote trust in safe handling of electronic health data of natural persons for primary use and secondary use as defined in this Regulation.

Moreover, Recital 52 states that the EHDS provides a legal basis in accordance with GDPR for the secondary use of personal electronic health data, *including the safeguards* to permit the processing of special categories of data, required under Article 9(2), points (g) to (j) GDPR, in terms of lawful purposes, trusted governance for providing access to health data (through the involvement of

health data access bodies) and processing in a secure environment, as well as arrangements for data processing, set out in the data permit. A breakdown of the EHDS safeguards is provided by Jane Reichel in Chapter 10.

As mentioned in Table 11.1, the EHDS does not refer to data federation. Still, when comparing the proposed framework for secondary uses of health data with the core elements of federated networks, a number of similarities emerge. First and foremost, the EHDS does not anticipate the creation of a large centralised electronic health data pool. On the contrary, the data will be stored by the respective health data holders and only released following the issuance of a data permit. This is similar to the concept of data nodes which are the constituting elements of federated networks. Furthermore, the EHDS restricts access to data through a secure processing environment and prohibits copying personal health data from this environment. The concept of accessing data without transferring it, while adhering to the GDPR principle of data minimisation, is integral to many federated infrastructures, as discussed further below. Given these similarities, it is worth examining the safeguards provided by existing federated networks, not only in the context of the GDPR but also in view of the framework of the EHDS. Conversely, looking at federated networks and their approach to aligning with the requirements of appropriate safeguards outlined in Article 89(1) GDPR may also enhance understanding of the EHDS framework.

11.4 Can federated networks be considered as appropriate safeguards?

11.4.1 Federated networks and data minimisation

Safeguards must ensure that technical and organisational measures are implemented to uphold, in particular, the principle of data minimisation. This principle stipulates that personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (Article 5(1)(c) GDPR). Recital 156 also mentions necessity and proportionality, which support the principle of data minimisation. Data pseudonymisation and anonymisation are mentioned in Article 89(1) as examples of practices that translate the principle of data minimisation into concrete measures. The use of the term in particular in Article 89(1) emphasises that the measures should serve to support the principle of data minimisation but does not exclude that they may also ensure other GDPR principles, such as the principle of integrity and confidentiality (also known as data security, Article 5(1)(f) GDPR) and data protection by design and by default (Article 25 GDPR). Moreover, the role of those principles in addressing the requirements of Article 89(1) was emphasised in the EDPB guidelines.[69] These principles require technical and organisational measures to secure the data to be considered in the design of processing operations. This ensures that privacy and data protection principles are safeguarded from the outset, with the highest possible level of privacy protection by default (for example, limiting the number of persons who have access to the data).

In federated networks, data remains stored on a local node with its original owner (data custodian) and is not copied to a central database. In this way, the data storage model avoids duplication of personal data to the central repository,[70] at the point of deposit for further reuse. The data is also typically pseudonymised or anonymised by the node owner. These measures could be viewed as supporting the principle of data minimisation by design and by default as recognised by the GDPR and further supported by the EHDS. It has also been claimed that data in the node is to be held 'as securely as the local infrastructure will allow'.[71] However, in practice, the individual node holders may not have sufficient resources nor skills to ensure data security of their node. Furthermore, data breaches may occur during data communication.[72] Hence, whether the measures in the federated network will respect the principle of integrity and confidentiality will depend on the participants of the network, in particular the node holders, and the features of connecting infrastructure.

Data minimisation, integrity and confidentiality principles also need to be considered from the perspective of the data users' access to data. Depending on the model of federated networks, access to the data may be provided in various ways. Direct access to data means that the user can visit the data and analyse each database separately but are only allowed to take summary statistics or aggregated information from the database. It has been claimed that direct access limits copying and transmission of data, reduces security risks and allows continuous monitoring of user activity.[73] Yet, if, during the direct access, the users can view the data in the node (for example, through an Internet browser), at least temporary copies of personal data will be generated.[74] Therefore, the impact of possible temporary copying of database content during direct access must be evaluated against data minimisation and data security principles. Federated networks may need to find additional measures to address the risks associated with such access, such as data access monitoring, data use agreements and logging of user activity.[75]

Arguably, indirect access to data may be a remedy that better addresses the requirement of data confidentiality and limits the scope of data processing. Indirect access mandates that while data remain behind secure firewalls at all times, users submit algorithms or queries and receive summary or performance statistics, which are vetted and executed by the data provider.[76] Thus, this is associated with a higher level of data security. However, this feature may reduce the usability of the data. If the users do not have access to data, they cannot verify its quality and check the harmonisation of data in different federated nodes, which may translate to errors and research challenges.[77] Therefore, although restrictive indirect data access may uphold the principles of data minimisation, integrity and confidentiality, a balance still needs to be achieved, to ensure that the data remains suitable for research purposes. Moreover, the use of health data for AI training, including in federated learning, may pose its own security challenges, such as the potential for a partially trained model to still reveal information about the training data.[78]

In summary, the storage of data in federated nodes, as opposed to copying the data to a central repository, can be perceived as ensuring data minimisation. However, questions arise whether accessing data in the federated node inherently has advantages in terms of minimising the data. This will depend on the levels of permissions that users have within the specific federated database. Moreover, deployment of multiple nodes in the network may bring about additional privacy threats.

Under the EHDS, the principle of data minimisation is ensured at various levels, such as by the requirement of anonymisation or pseudonymisation of the electronic health data made available for secondary use (Articles 66(2) and 3), provision of access only to electronic health data which is relevant for the purpose of processing indicated in the data access application by the data user (Article 66(1)), and by limiting access to electronic health data to access through secure processing environments (Article 73(1)). These environments resemble those federated networks, in which the data cannot be copied by the user, and at all times remains in the node. The technical, information security and interoperability requirements for the secure processing environments will be further elaborated by the Commission by way of implementing acts. The questions on minimising the risk of unauthorised processing of data during direct and indirect access which are outlined for federated networks may inform the development of such requirements for secure environments under EHDS.

11.4.2 Federated networks and compliance with GDPR and the relevant points from the EHDS

The second condition of the appropriate safeguards according to Article 89(1) GDPR is that they need to comply with the GDPR. This, in turn, raises the question of whether the use of federated networks for sharing of health data can be considered a GDPR compliant measure. Some of the GDPR requirements relevant to the sharing and use of personal data for scientific research include defining and complying with the responsibilities of the relevant role of a controller or processor of the data (Articles 5(2), 24 and 28 GDPR, among others), demonstrating a legal basis for data processing (Articles 6 and 9 GDPR), complying with restrictions on data transfers outside the European Economic Area (EEA) to countries that do not provide an adequate level of data protection (Chapter V of GDPR), and ensuring data accuracy (Article 5(1)(c) GDPR). Below, we discuss how sharing personal data across federated networks may impact compliance with these requirements.

11.4.2.1 Federated networks and roles and responsibilities of various parties regarding controllership over data

The data custodians and data users, as well as any other parties involved in the federated networks (such as providers of elements of its infrastructure), need to understand their role in data processing under the GDPR. The possible

roles are controller (Article 4(7)), which determines the purposes and means of processing (Article 4(7) GDPR) including joint controller (Article 26), and processor (Article 4(8)), which processes personal data on behalf of the controller (Article 4(8) GDPR). These roles cannot be contractually modified, as they depend on the law or factual circumstances.[79] The role determines the responsibilities of the participant of the network. For example, a controller must ensure a legal basis for the processing. A processor does not need to demonstrate such legal basis but must observe the restrictions of the data processing agreements and instructions of the controller (Article 28 GDPR). The identification of the controllers and processors for each processing operation is a prerequisite for the preparation of privacy notices (pursuant to Articles 13 or 14), records of processing operations (Article 30), agreements between entities involved in processing operations (for example, data processing agreements pursuant to Article 28(3) or joint controller arrangement under Article 26) and other documents necessary to demonstrate compliance with the GDPR.

The development of a federated network requires complex multi-party discussions and various converging decisions to be made, taking into account the entire data lifecycle. These decisions relate to various issues such as categories of data to be shared in the federated network, tools to be used for storage, data management and access, collaboration of various components of the infrastructure to ensure data security, data access levels and policies. Such decisions are often beyond the control of any single data custodian and must be made jointly by all data custodians in close coordination with the infrastructure component providers. The scope of these decisions and the manner in which they are made may affect the assessment of the GPDR roles of federated network stakeholders.

Under CJEU jurisprudence[80] and the EDPB Guidelines,[81] the controller does not necessarily have to possess a copy of the data it controls or even have access to it, as long as it gives instructions in relation to the processing. Thus, the argument that since the data is not moved from the node to a new user, the node owner remains the controller and the existence of the federation does not change that role[82] may not always be true. Furthermore, harmonisation of data access rules (discussed below) may tip the assessment of the GDPR role of the custodians from individual controllers to joint controllership. To complicate matters further, if the nodes delegate the decisions to approve users of the federated networks to a central authority (such as data access committee), depending on the terms of reference and access procedures,[83] the authority may become a controller of the data in the federated networks or act as a processor on behalf of the node custodians. Users who use data for their own purposes will typically be independent controllers.[84] However, this qualification may change if they have common scientific goals with the data custodians or other users.

Proper qualification as a controller has various implications under the GDPR. For example, actors qualified as controllers are responsible for ensuring the legal basis for processing (Articles 6 and 9), informing data subjects (Articles 13 and 14), implementing the principles of data protection by design

and by default (Article 25 GDPR) and reporting data breaches (Article 33). Establishing common governance over the data in the federated network complicates the issue of legal responsibility. For example, how should the data access governance structure be explained to data subjects in a privacy notice? If a single gatekeeper allows users to access all data in the repository, who is liable if a malicious user leaks the data? Which participant in the federated networks should report the breach to the authorities and take corrective action? As put by Bak and others, 'because this responsibility is spread out over multiple partners instead of being allocated to a central controller, the risk that no one takes proper responsibility for data governance is increased'.[85]

In light of the above, planning a federated network to determine steps for GDPR compliance requires early discussion and assessment of stakeholder roles, followed by a written agreement that addresses established roles and responsibilities.[86] The agreement should cover both responsibilities to data subjects and obligations of potential users of the data. In practice, however, it may be difficult to identify and allocate responsibilities, and nodes may have conflicting perceptions of their qualifications. The matrix can become even more complex when additional parties are involved in building the shared infrastructure or providing certain components of it (such as user access and authorisation services).

Difficulties in proper assignment of GDPR roles and diverging views on this matter are clearly shown in the legislative process of EHDS. In the EHDS proposal, both the data users and health data access bodies (HDABs) were considered joint controllers of the electronic health data processed in accordance with the data permit. During the legislative process, this position was amended to state now that the health data holder and HDABs are simply controllers for a specific part of the process of secondary use of health data and according to their respective roles, without reference to joint controllership (Article 74).

11.4.2.2 Federated networks and nature of data processing

Counterintuitively, although the term "sharing of data" suggests a single action, multiple operations are required to enable health data sharing with users in federated networks. These operations involve both the data custodians and the data users. This is an important consideration, because under the GDPR, the processing of personal data by the controller requires a legal basis under Article 6 and, for special categories of data (such as health data), identification of an exception under Article 9(1) GDPR. The various steps taken by data controllers and data users when sharing and using personal data in the federated networks need to be examined to determine whether they could be considered as processing, which requires a legal basis.

Processing of personal data is defined broadly in the GDPR as 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 combination, restriction, erasure or destruction (Article 4(2) GDPR).

Custodians are typically responsible for collecting and preparing the data (for example, pseudonymising it and adapting it to the agreed standards) and keeping it stored in the node. The collection, adaptation, and storage of data are clear examples of processing under the Article 4(2) GDPR definition and, therefore, require the custodian to establish a legal basis for these activities. In turn, the user typically performs a search for data stored in the nodes before accessing the data. Searching the central data catalogue or reviewing the metadata of each node without using the actual data is not an operation on the data itself, and therefore does not constitute data processing under the GDPR. Once the user finds a suitable dataset and is given access to it, they can start using that data for their research purposes. The use of the data is also mentioned as an example of processing in Article 4(2) GDPR. Some federated networks allow the users to query or process data in the local node,[87] but without providing direct access to the data. For example, a user's AI model can visit[88] the data during federated learning. The user initiates this activity and as a result the data is used to fulfil the user's scientific goals. Given the broad definition of processing under GDPR, which includes any use of personal data, we agree with the view[89] that the data analysis performed at the local node is a data processing operation. Similarly, conducting federated AI training on personal data is subject to GDPR.[90] Moreover, while the federated network may not facilitate direct user interaction with the data, it is ultimately the user who determines the purpose for which the data is queried for scientific research and the means (for example, specific query) by which it is done. In turn, the user acting as controller for the use of health data in their research activity must demonstrate a legal basis for this processing.

Consequently, the processing of data for storage in the federated node by the data holder would be considered a different processing purpose than the access and use of the same data by one or more data users (researchers). In order to comply with the GDPR, all controllers involved in the processing of data in the federated network must ensure an appropriate legal basis for their processing operations. In practice, a single consent from the data subject may not sufficiently cover multiple operations that are often performed by different actors in the federated networks, so alternative approaches (such as dynamic consent[91]) may need to be implemented to ensure that all controllers can demonstrate a valid legal basis for their processing.

The EHDS provides an interesting insight into this discussion. According to Recital 52, the EHDS will provide the legal basis (and obligation) for the data holder to disclose the requested electronic health data to the HDAB and secure the legal basis for the processing performed by these administrative bodies. However, to obtain a data permit, the data applicant will have to demonstrate a legal basis pursuant to Article 6 GDPR. This recital highlights the complexity of ensuring a legal basis for all participants in the data sharing framework, even when governance rules are imposed by a legislative act such as the EHDS.

11.4.2.3 Other challenges in deployment of federated networks

Federated networks may also not comply with other GDPR principles, such as the need to respect the data accuracy principle, as well as broader issues such as infrastructure sustainability. Under Article 5(1)(d) GDPR, the data must be accurate and, where necessary, kept up to date. This raises the question of whether federated networks can ensure the accuracy and quality of research data. This is not a straightforward issue. Some doubt whether error-free data sets are even possible, especially in the medical field.[92] Dealing with distributed structures adds another layer of complexity. Federated networks consist of many distributed data sources (nodes). If the data within a node is restricted to the custodian's access, only the custodian can vouch for its accuracy. As a result, attempting to verify all data nodes by an outside entity becomes infeasible. As a practical solution, the participating nodes may establish a process of validating datasets, define measures to evaluate their quality and even *ex ante* accountability measures.[93] Still, such an overarching agreement on how the data will be collected and prepared may trigger the qualification of nodes as joint controllers, as further discussed above.

Compliance with Chapter V of GDPR when transferring personal data to countries outside the European Economic Area which do not provide for adequate level of data protection (non-adequate third countries) is also an area for consideration. These strict legal conditions drive the question of whether accessing data through the federated network constitutes a data transfer. While comprehensive analysis of this topic remains beyond the scope of this chapter, given the broad interpretation of data transfers by the recent EDPB guidelines,[94] such qualification is not excluded.

Sustainability and continued reliability of data sources in federated networks also require further consideration. Users should be able to document the data sources for their results so that these results can be verified and reproduced. Often the structure of the network is fluid and the data in the nodes may be altered or removed by its custodians. This has implications for reproductivity of the research results. To become an attractive resource for the potential users, the node custodians must agree on the rules of withdrawing or modifying the data in the network, and those rules should be transparent and allow the users to plan and document their research.

As shown in this part of the chapter, the answer to the question of whether federated networks are GDPR compliant is multifaceted and depends on different decisions made during the implementation of the network. While there does not appear to be any feature of the federated network model that inherently violates the requirements of the GDPR, given the different possible implementations of the federated approach, GDPR compliance must be carefully considered and documented during the planning and implementation of the network.

The EHDS addresses some of the key points indicated above. It provides for data quality and utility labels (Article 78) which can be used by the data users to find a dataset that fits their needs. Moreover, it provides additional restrictions regarding access to non-personal health data from third countries (Article

61). Finally, the inclusion of obligations to make health data available for secondary use (Article 41) may shift perspectives on the long-term sustainability of existing data repositories, provided they align with the requirements of the EHDS.

11.4.3 Federated networks and protection of rights and freedoms of the data subjects

According to the third overarching requirement of Article 89(1) GDPR, the appropriate safeguards should protect the rights and freedoms of the data subject. These rights and freedoms can be viewed, in a broad sense, as deriving directly from human rights law. For example, in discussing the safeguards mandated by the legislative measures implemented by the DEA, J. Bell[95] has referred to European human rights law, in particular Article 8 of ECHR and Articles 7 and 8 of the EU Charter and related case law to assess the enforcement mechanisms behind the legislative measures.

However, the requirement to respect the rights and freedoms of natural persons can also be understood in a narrower sense, as referring to the right of protection of personal data. In particular, Recital 75 of GDPR lists examples of the risks to the rights and freedoms which may result from personal data processing, including among others, physical, material or non-material damage, 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. Many of those risks are associated with data subjects' loss of control over their personal data. In the context of use of data for research purposes, mitigating such risks is often associated with not only compliance with GDPR but also with ethical standards, such as Oviedo Convention[96] or OECD Guidelines on Human Biobanks and Genetic Research Databases.[97]

As mentioned above, participation in a federated network is tied to an agreement which the data custodian must accept when joining a federated network. We argue that the content of this agreement can be a key factor in ensuring that the rights and freedoms of the data subjects are respected. Specifically, the measures to address risks to the rights and freedoms of data subjects should be evaluated when designing the data access governance rules to be adopted by the network of custodians. Data access governance regulates who are the users who have access to data, for what purposes and under what conditions.[98] Through an agreement on data governance, each node custodian may cede some of its autonomy to a central authority or to rules of common governance. This may include connecting the federated node to a central infrastructure[99] or, in a loosely defined federated network, adhering to agreements on common rules or standards[100] (for example, on data or technology). Common rules should determine how the access to the data in the node is granted. These rules should be designed not only in compliance with GDPR but also to respect the ethical standards mentioned above.

When establishing conditions for data access, it is important to consider not only the legitimate interests of potential data users in pursuing their scientific goals but also to protect the rights and freedoms of data subjects. Delegating the authority to make data access decisions to a common governance body[101] of the federated network (such as data access committee, DAC) may serve to streamline the process. It is argued that collective bodies, such as DACs are better suited to assess the data use requests objectively and transparently.[102] However, to ensure that the DAC decision making process is sufficiently independent from professional interests of the data custodians and potential users, the DAC should consist of a diverse group of stakeholders, with different expertise. It is particularly important to include representatives from patient organisations to ensure that the perspectives and needs of patients are effectively represented.[103] Moreover, the transparency of the data access process could be supported by technical solutions where federated networks collect and make public certain information about data usage.[104] Furthermore, the legal terms on the data access accepted by the data user can also be an instrument to ensure protections for the rights and freedoms of data subjects by providing legally binding rules on confidentiality, security measures to be adopted by the data user and restrictions on attempting to re-identify the data subjects.[105] Data use conditions for data users should also take into consideration any restrictions stemming from data subjects' informed consent or ethics approvals.

When in effect, the EHDS will change the rules of the game for seeking health data for scientific research by introducing harmonised rules for accessing data for secondary use. Yet, before its adoption, existing federated networks are challenged with the design of their own data governance framework, in a way that minimises privacy risks and fosters research goals. Decisions on such framework can be informed by some of the safeguarding elements which are provided in the EHDS, in particular standard common processes for data access applications (Article 67), conditions imposed on the data user in the data permit (Article 68) and reporting requirements for HDABs to ensure the transparency of the process (Article 59). One notable difference, however, is that while federated networks may deploy various governance bodies, including data access committees, under the EHDS the data access decision hinges, in principle, on the health data access bodies. The mechanisms aimed to shield the rights and freedoms of the patients under the EHDS framework are evaluated by others in this book.

11.5 Conclusions

Federated networks have emerged as a response to challenges in sharing health data. They offer advantages such as cost reduction through shared infrastructure and precise control over data access permissions at various levels.[106] From a legal perspective, properly implemented federated networks comply with GDPR requirements and support data minimisation and security.

However, it is premature to declare them a universal solution for complex legal demands for data sharing. Federated networks must be evaluated on a

case-by-case basis, and further research is needed to define GDPR-compliant characteristics. Depending on the implementation, roles and responsibilities of stakeholders within the federated infrastructure can vary, affecting the ability to assess whether their technical and organisational measures can support GDPR principles of data minimisation (Article 5(1)(c)), data accuracy (Article 5(1)(d)) or integrity and confidentiality (Article 5(1)(f)). The sustainability and reliability of data sources in federated networks also warrant further examination, as they may affect data security and accountability. There is also a need for further exploration into how the means to protect the rights and freedoms of data subjects should be reflected in the rules of data governance and usage within the network. The absence of standardised guidelines for constructing these networks within health infrastructures for data sharing complicates the assessment of federated networks as appropriate safeguards under Article 89(1) GDPR.

Understanding the benefits and risks associated with sharing data in federated networks may inform legislative decisions and create a more robust EHDS framework. At the same time, some safeguards foreseen by the EHDS should be carefully evaluated by existing and future federated networks. These safeguards may serve as guiding examples to improve GDPR compliance and balance the interests and obligations of data subjects, data holders, and data users.

Another point of consideration is to what extent the framework of health data sharing provided for in the EHDS will be mandatory and exclusive. EHDS provides that the regulation 'shall not affect access to electronic health data for secondary use agreed in the framework of contractual or administrative arrangements between public or private entities' (Article 1(8)). This suggests that existing health data repositories (either centralised or federated) may be allowed to function based on contractual agreements and provide their own governance rules for reuse of deposited data. Nevertheless, federated network developers should pay close attention to the EHDS proposal to understand the potential impact on their sharing model, as well as for (indirect) guidance on how to address the legal challenges inherent in multi-stakeholder data sharing.

Acknowledgements/Funding

The work of MKC was funded by the European Union's research and innovation programme under grant agreement number 952179 – INCISIVE and grant agreement number 101095382 – FLUTE.

Notes

1 Sarah J. MacEachern and Nils D. Forkert, 'Machine learning for precision medicine' (2021) 64(4) *Genome* 416.
2 Anna D. Barker and Jerry SH Lee, 'Translating 'Big Data' in oncology for clinical benefit: progress or paralysis' (2022) 82(11) *Cancer Res*. 2072.
3 Eric J Topol, 'High-performance medicine: the convergence of human and artificial intelligence' (2019) 25(1) *Nat. Med*. 44.

4 David Peloquin and others, 'Disruptive and avoidable: GDPR challenges to secondary research uses of data' (2020) 28(6) *Eur. J. Hum. Genet*. 697; Olga Tzortzatou and others, 'Biobanking Across Europe Post-GDPR: A Deliberately Fragmented Landscape' in Santa Slokenberga, Olga Tzortzatou and Jane Reichel (eds), *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe* (Springer Nature, 2021).; Birgit Wouters and others, 'Putting the GDPR into practice: Difficulties and uncertainties experienced in the conduct of big data health research' (2021) 7 *Eur. Data Prot. L. Rev*. 206.
5 World Economic Forum, 'Sharing sensitive health data in a federated data consortium model: An eight-step guide' (2020) https://www.weforum.org/reports/sharing-sensitive-health-data-in-a-federated-data-consortium-model-an-eight-step-guide/ accessed on 28 September 2023.
6 Alexander Bernier and others, 'Reconciling the biomedical data commons and the GDPR: three lessons from the EUCAN ELSI collaboratory' (2023) *Eur. J. Hum. Genet*. 1.
7 James Casaletto and others 'Federated analysis for privacy-preserving data sharing: a technical and legal primer' (2023) 24 *Annu. Rev. Genomics Hum. Genet*. 347.
8 Raivo Ruusalepp, 'A comparative study of international approaches to enabling the sharing of research data' (2008) https://era.ed.ac.uk/bitstream/handle/1842/3361/Ruusalepp%20Data_Sharing_Report.pdf?sequence=1 accessed on 28 September 2023.
9 European Commission, 'A European strategy for data' (Communication) (2020) 66 final.
10 European Commission, 'Proposal for a regulation – The European Health Data Space' COM(2022) 197 final. On 24 April 2024 the European Parliament passed a legislative resolution with a view to the adoption of Regulation (EU) 2024/… of the European Parliament and of the Council on the European Health Data Space P9_TC1-COD(2022)0140 (EHDS compromise text).
11 For example, Peloquin (n 4); Casaletto (n 7); Rita T. Lawlor, 'The impact of GDPR on data sharing for European cancer research' (2023) 24(1) *Lancet Oncol*. 6; Regina Becker and others, 'Applying GDPR roles and responsibilities to scientific data sharing' (2022) 12(3) *Int. Data Priv. Law* 207.
12 European Parliament and the Council Regulation 2016/679 of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data (General Data Protection Regulation) [2016] OJ L 119/1.
13 Amit P. Sheth and James A. Larson, 'Federated database systems for managing distributed, heterogeneous, and autonomous databases' (1990) 22(3) *ACM Comput. Surv*. 183.
14 Dennis Heimbigner and Dennis McLeod, 'A federated architecture for information management' (1985) 3(3) *ACM Trans. Inf. Syst*. 253.
15 Adrian Thorogood and others, 'International federation of genomic medicine databases using GA4GH standards' (2021) 1(2) *Cell Genomics* 100032.
16 Christine Suver and others, 'Bringing code to data: Do not forget governance' (2020) 22(7) *JMIR* e18087.
17 World Economic Forum, 'Federated data systems: balancing innovation and trust in the use of sensitive data' (2019) https://www3.weforum.org/docs/WEF_Federated_Data_Systems_2019.pdf accessed on 9 November 2023.
18 Stéphane Goldstein, 'The evolving landscape of Federated Research Data Infrastructures Knowledge Exchange', (2017) *Zenodo*.
19 Yannick Marcon and others, 'Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD' (2021) 17(3) *PLoS Comput. Biol*. e1008880.

20 Richard Milne and others, 'A concentric circles view of health data relations facilitates understanding of sociotechnical challenges for learning health systems and the role of federated data networks' (2022) 5 *Front. Big Data* 945739.
21 Rolf Gedeborg and others, 'Federated analyses of multiple data sources in drug safety studies' (2023) 32(3) *Pharmacoepidemiol Drug Saf*. 279.
22 Suver and others (n 16).
23 Global Alliance for Genomics and Health, 'A federated ecosystem for sharing genomic, clinical data' (2016) 352(6291) *Science* 1278.
24 Brendan McMahan and others, 'Communication-efficient learning of deep networks from decentralized data' (2017) 54 Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR.
25 Suver and others (n 16).
26 Alissa Brauneck and others, 'Federated Machine Learning, Privacy-Enhancing Technologies, and Data Protection Laws in Medical Research: Scoping Review' (2023) 25 *JMIR* 41588.
27 Gary Saunders and others, 'Leveraging European infrastructures to access 1 million human genomes by 2022' (2019) 20(11) *Nat. Rev. Genet*. 693.
28 J. Damián Segrelles Quilis and others, 'A federated cloud architecture for processing of cancer images on a distributed storage' (2023) 139 *Future Gener. Comput. Syst*. 38.
29 Kari A. Stephens and others, 'Implementing partnership-driven clinical federated electronic health record data sharing networks' (2016) 93 *Int. J. Med. Inform*. 26.
30 Goldstein (n 18).
31 Suver and others (n 16).
32 Marc Fiume and others, 'Federated discovery and sharing of genomic data using Beacons' (2019) 37 *Nat. Biotechnol*. 220.
33 Luís António Bastião Silva, 'Federated architecture for biomedical data integration' (Dissertation, Universidade de Aveiro 2015) http://hdl.handle.net/10773/15759 accessed on 14 May 2024.
34 Griffin M. Weber, 'Federated queries of clinical data repositories: Scaling to a national network' (2015) 55 *J. Biomed. Informat*. 231.
35 Marcelline R. Harris, Lisa A. Ferguson, and Airong Luo, 'Infrastructuring an organizational node for a federated research and data network: A case study from a sociotechnical perspective' (2021) 5(1) *J. Clin. Transl. Sci*. e186.
36 European Strategy for Data (n 9).
37 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) [2022] OJ L 152/1.
38 EHDS.
39 European Commission, 'Commission Staff Working Document, Impact Assessment Report accompanying the document Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space' (3 May 2022) Swd(2022) 131 final.
40 EDPB-EDPS Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act), adopted on 10 March 2021.
41 EDPS, Preliminary Opinion 8/2020 on the European Health Data Space, adopted on 17 November 2020.
42 EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space, adopted on 12 July 2022.
43 EDPS, Opinion 3/2020 on the European strategy for data, adopted on 16 June 2020.
44 EDPS, A Preliminary Opinion on data protection and scientific research, adopted on 6 January 2020.

45 EDPS website https://edps.europa.eu/_en accessed on 25 October 2023.
46 Federated Learning https://edps.europa.eu/press-publications/publications/techsonar/federated-learning_en accessed on 5 January 2024.
47 TechDispatch #1/2022 – Federated Social Media Platforms https://edps.europa.eu/data-protection/our-work/publications/techdispatch/2022-07-26-techdispatch-12022-federated-social-media-platforms_en accessed on 5 January 2024.
48 ENISA 'Engineering Personal Data Sharing Emerging Use Cases and Technologies' January 2023 https://www.enisa.europa.eu/publications/engineering-personal-data-sharing accessed on 5 January 2024.
49 ENISA 'Fog And Edge Computing in 5G Security Opportunities And Challenges' March 2023. https://www.enisa.europa.eu/publications/fog-and-edge-computing-in-5g accessed on 5 January 2024.
50 ENISA, 'Engineering personal data protection in EU Data Spaces' January 2023 https://www.enisa.europa.eu/publications/engineering-personal-data-protection-in-eu-data-spaces accessed on 5 January 2024.
51 Kart Pormeister, 'Genetic data and the research exemption: is the GDPR going too far?' (2017) 7 *Int. Data Priv. Law* 137.
52 Jessica Bell and others, 'Balancing Data Subjects' Rights and Public Interest Research' (2019) 5 *European Data Protection Law Review* 43.
53 EDPS, Opinion on safeguards and derogations under Article 89 GDPR in the context of a proposal for a Regulation on integrated farm statistics, adopted 20 November 2017; EDPS, A Preliminary Opinion on data protection and scientific research (n 44); EDPB, Response to the Request from the European Commission for Clarifications on the Consistent Application of the GDPR, Focusing on Health Research, adopted on 2 February 2021.
54 EDPS, opinion on safeguards and derogations under Article 89 GDPR.
55 Els Kindt and others, 'Study on the Appropriate Safeguards under Article 89(1) GDPR for the Processing of Personal Data for Scientific Research. Final Report' (2021, EDPS/2019/02–08).
56 Bell and others (n 52).
57 Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 1950 Council of Europe European Treaty Series 5.
58 Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.
59 Ciara Staunton, Santa Slokenberga and Deborah Mascalzoni, 'The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks' (2019) 27(8) *European J. Hum. Gen*. 1159.
60 Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (Oxford Academic, online edn, 2020).
61 Anne-Marie Duguet and Jean Herveg, 'Safeguards and Derogations Relating to Processing for Scientific Purposes: Art. 89 Analysis for Biobank Research' in Santa Slokenberga, Olga Tzortzatou and Jane Reichel (eds) *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe* (Springer Nature 2021); Christopher F. Mondschein and Cosimo Monda 'The EU's General Data Protection Regulation (GDPR) in a research context' in Pieter Kubben, Michel Dumontier and Andre Dekker (eds), *Fundamentals of Clinical Data Science* (Springer Nature 2019).
62 Annagrazia Altavilla and others, 'The secondary use of paediatric data under GDPR: looking for new safeguards for research' (2019) 3 *EPLR* 156.
63 Jane Reichel, 'The GDPR and processing of personal data for research purposes: what about case law?' (2021) 27(1) *European Pub. Law* 167.
64 Bell and others (n 52).
65 Kuner (n 60).

66 Mondschein and Monda (n 61).
67 Giulia Schneider, 'Health data pools under European policy and data protection law: research as a new efficiency defence?' (2020) 11 *JIPITEC* 49 https://www.jipitec.eu/archive/issues/jipitec-11-1-2020/5082 accessed on 14 May 2024.
68 Ciara Staunton and others, 'Appropriate safeguards and Art. 89 of the GDPR: considerations for biobank, databank and genetic research' (2022) 13 *Front. Gen*.
69 EDPB Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak, adopted 21 April 2020.
70 Stephanie Rossello, Luis Muñoz-González, and Roberto Díaz Morales, 'Data protection by design in AI? The case of federated learning' (2021) 166 *Computerrecht* 273. https://ssrn.com/abstract=3879613 accessed on 14 May 2024.
71 Milne and others (n 20).
72 Brauneck and others (n 26).
73 Thorogood and others (n 15).
74 C-360/13 *Public Relations Consultants Association Ltd v Newspaper Licensing Agency Ltd and Others* ECLI:EU:C:2014:1195.
75 Maria Alvarellos and others, 'Democratizing clinical-genomic data: How federated platforms can promote benefits sharing in genomics' (2023) 13 *Front. Genet*. 3725.
76 Thorogood and others (n 15).
77 Suver and others (n 16).
78 Bjørn Aslak Juliussen and others, 'The third country problem under the GDPR: enhancing protection of data transfers with technology' (2023) 13(3) *Int. Data Priv. Law* 225.
79 EDPB Guidelines 07/2020 on the concepts of controller and processor in the GDPR, adopted on 07 July 2021.
80 C-25/17 *Jehovan todistajat – uskonnollinen yhdyskunta* ECLI:EU:C:2018:551; C-40/17 *Fashion ID GmbH & Co.KG v Verbraucherzentrale NRW eV* ECLI:EU:C:2019:629.
81 EDPB Guidelines 07/2020 (n 79).
82 Milne and others (n 20).
83 Foundation 29 *Health Data: the Playbook* https://www.healthdata29.org/playbook accessed on 10 November 2023.
84 Becker and others (n 11).
85 Marieke Bak and others 'Federated learning is not a cure-all for data ethics' (2024) *Nat Mach Intell **6**, pages 370–372 (2024)*.
86 Gedeborg and others (n 21).
87 Harry Hallock and others, 'Federated networks for distributed analysis of health data' (2021) 9 *Front Public Health* 712569.
88 Thorogood and others (n 15).
89 Gedeborg and others (n 21).
90 Rossello, Muñoz-González, and Morales (n 70).
91 Jane Kaye and others, 'Dynamic consent: a patient interface for twenty-first century research networks' (2015) 23(2) *Eur. J. Hum. Gen*. 141.
92 Brauneck and others (n 26).
93 Brauneck and others (n 26).
94 EDPB, Guidelines 05/2021 on the Interplay between the application of Article 3 and the provisions on international transfers as per Chapter V of the GDPR, adopted 14 February 2023.
95 Bell and others (n 52).
96 Council of Europe, Convention for the Protection of Huan Rights and of the Huan Being with regard to the Application of Biology and Medicine (Convention on Human Rights and Biomedicine or the Oviedo Convention).

97 Organization for Economic Cooperation and Development, 'OECD guidelines on human biobanks and genetic research databases' (2009).
98 Mahsa Shabani, Adrian Thorogood and Madeleine Murtagh, 'Data access governance' in Graeme Laurie and others (eds), *The Cambridge Handbook of Health Research Regulation* (Cambridge University Press, 2021).
99 World Economic Forum (n 17).
100 Thorogood and others (n 15).
101 Alvarellos and others (n 75).
102 Thijs Devriendt and others, 'Data sharing platforms: instruments to inform and shape science policy on data sharing? (2022) 127(6) *Scientometrics* 3007.
103 Milne and others (n 20).
104 Devriendt and others (n 102).
105 Shabani, Thorogood and Murtagh (n 98).
106 World Economic Forum (n 17).

List of sources

EU treaties, legislative acts and international conventions

Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.
Council of Europe, Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 1950 European Treaty Series 5.
Council of Europe, Convention for the Protection of Human Rights and of the Human Being with regard to the Application of Biology and Medicine (Convention on Human Rights and Biomedicine or the Oviedo Convention) European Treaty Series 164, [1997].
European Parliament and the Council Regulation 2016/679 of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data (General Data Protection Regulation) [2016] OJ L 119/1.
European Parliament and the Council 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) [2022] OJ L 152/1.

Official documents

EDPB, Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak, adopted on 21 April 2020.
EDPB, Response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research, adopted on 2 February 2021.
EDPB, Guidelines 07/2020 on the concepts of controller and processor in the GDPR, adopted on 07 July 2021.
EDPB, Guidelines 05/2021 on the interplay between the application of Art. 3 and the provisions on international transfers as per Chapter V of the GDPR, adopted on 14 February 2023.
EDPB-EDPS, Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act), adopted on 10 March 2021.
EDPB-EDPS, Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space, adopted on 12 July 2022.

EDPS, Opinion on safeguards and derogations under Article 89 GDPR in the context of a proposal for a Regulation on integrated farm statistics, adopted on 20 November 2017.
EDPS, A Preliminary Opinion on data protection and scientific research, adopted on 6 January 2020.
EDPS, Opinion 3/2020 on the European strategy for data, adopted on 16 June 2020.
EDPS, Preliminary Opinion 8/2020 on the European Health Data Space, adopted on 17 November 2020.
EDPS https://edps.europa.eu/_en accessed on 25 October 2023.
ENISA 'Engineering Personal Data Sharing Emerging Use Cases and Technologies' January 2023 https://www.enisa.europa.eu/publications/engineering-personal-data-sharing accessed on 5 January 2024.
ENISA 'Fog And Edge Computing in 5G Security Opportunities And Challenges' March 2023 https://www.enisa.europa.eu/publications/fog-and-edge-computing-in-5g accessed on 5 January 2024.
ENISA, 'Engineering Personal Data Protection in EU Data Spaces' January 2023 https://www.enisa.europa.eu/publications/engineering-personal-data-protection-in-eu-data-spaces accessed on 5 January 2024.
European Commission, 'A European strategy for data' (Communication) (2020) 66 final.
European Commission, Proposal for a regulation – The European Health Data Space COM(2022) 197 final.
European Commission, 'Commission Staff Working Document, Impact Assessment Report accompanying the document Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space' (3 May 2022) Swd(2022) 131 final.
European Parliament position adopted at first reading on 24 April 2024 with a view to the adoption of Regulation (EU) 2024/… of the European Parliament and of the Council on the European Health Data Space P9_TC1-COD(2022)0140.

Literature

Altavilla A and others, 'The Secondary Use of Paediatric Data under GDPR: Looking for New Safeguards for Research' (2019) 3 *EPLR* 156. DOI:10.21552/eplr/2019/4/6
Alvarellos M and others, 'Democratizing Clinical-Genomic Data: How Federated Platforms Can Promote Benefits Sharing in Genomics' (2023) 13 *Frontiers in Genetics* 3725. DOI: 10.3389/fgene.2022.104545
Bak M and others, 'Federated Learning Is Not a Cure-All for Data Ethics' (2024) *Nat Mach Intell* 1. DOI:10.1038/s42256-024-00813-x
Barker A and Lee J, 'Translating 'Big Data' in Oncology for Clinical Benefit: Progress or Paralysis' (2022) 82(11) *Cancer Research* 2072. DOI: 10.1139/gen-2020-0131
Becker R and others, 'Applying GDPR Roles and Responsibilities to Scientific Data Sharing' (2022) 12(3) *Int. Data Priv. Law* 207. DOI: 10.2139/SSRN.3851128
Bell J and others, 'Balancing Data Subjects' Rights and Public Interest Research' (2019) 5 *Eur. Data Prot. Law Rev.* 43. DOI:10.21552/EDPL/2019/1/8
Bernier A and others, 'Reconciling the Biomedical Data Commons and the GDPR: Three Lessons From the EUCAN ELSI Collaboratory' (2023) 1 *European Journal of Human Genetics* DOI:10.1038/s41431-023-01403-y.
Brauneck A and others, 'Federated Machine Learning, Privacy-Enhancing Technologies, and Data Protection Laws in Medical Research: Scoping Review' (2023) 25 *JMIR* 41588. DOI:10.2196/41588
Casaletto J and others, 'Federated Analysis for Privacy-Preserving Data Sharing: A Technical and Legal Primer' (2023) 24 *Annual Review of Genomics and Human Genetics* 347. DOI:10.1146/annurev-genom-110122-084756

Suver Ch and others, 'Bringing code to data: Do not forget governance' (2020) 22(7) *JMIR* e18087. DOI:10.2196/preprints.18087

Devriendt T and others, 'Data Sharing Platforms: Instruments to Inform and Shape Science Policy on Data Sharing? (2022) 127(6) *Scientometrics* 3007. DOI:10.1007/s11192-022-04361-2

Duguet AM and Herveg J, 'Safeguards and Derogations Relating to Processing for Scientific Purposes: Art. 89 Analysis for Biobank Research' in Slokenberga, S, Tzortzatou, O and Reichel, J (eds) *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe* (Springer Nature, 2021). DOI:10.1007/978-3-030-49388-2_7

Federated Learning https://edps.europa.eu/press-publications/publications/techsonar/federated-learning_en accessed on 5 January 2024.

Fiume M and others, 'Federated Discovery and Sharing of Genomic Data Using Beacons' (2019) *Nature Biotechnology* 37¸ 220. DOI:10.1038/s41587-019-0046-x

Foundation 29 Health Data: The Playbook https://www.healthdata29.org/playbook accessed on 10 November 2023.

Gedeborg R and others, 'Federated Analyses of Multiple Data Sources in Drug Safety Studies' (2023) 32(3) *Pharmacoepidemiology and Drug Safety* 279. DOI:10.1002/pds.5587

Global Alliance for Genomics and Health, 'A Federated Ecosystem for Sharing Genomic, Clinical Data' (2016) 352(6291) *Science* 1278. DOI: 10.1126/science.aaf6162

Goldstein S, 'The Evolving Landscape of Federated Research Data Infrastructures Knowledge Exchange', (2017). DOI:10.5281/ZENODO.1064730

Hallock H and others, 'Federated Networks for Distributed Analysis of Health Data' (2021) *Frontiers in Public Health* 9¸ 712569. DOI:10.3389/fpubh.2021.712569

Harris M, Ferguson L and Luo A, 'Infrastructuring an Organizational Node for a Federated Research and Data Network: A Case Study From a Sociotechnical Perspective' (2021) *J. Clin. Transl. Sci*. 5(1), e186. DOI:10.1017/cts.2021.846

Heimbigner D and McLeod D, 'A Federated Architecture for Information Management' (1985) 3(3) *ACM Transactions on Information Systems* 253. DOI:10.1145/4229.4233

Juliussen BA and others, 'The Third Country Problem Under the GDPR: Enhancing Protection of Data Transfers With Technology' (2023) 13(3) *Int. Data Priv. Law* 225. DOI:10.1093/idpl/ipad013

Kaye J and others, 'Dynamic Consent: A Patient interface for Twenty-First Century Research Networks' (2015) 23(2) *European Journal of Human Genetics* 141. DOI:10.1038/ejhg.2014.71

Kindt E and others, 'Study on the Appropriate Safeguards under Article 89(1) GDPR for the Processing of Personal Data for Scientific Research. Final Report' (2021, EDPS/2019/02–08).

Kuner Ch and others (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (Oxford Academic, online edn, 2020).

Lawlor R, 'The Impact of GDPR on Data Sharing for European Cancer Research' (2023) 24(1) *The Lancet Oncology* 6. DOI: 10.1016/S1470-2045(22)00653-2

MacEachern S and Forkert N, 'Machine Learning for Precision Medicine' (2021) 64(4) *Genome* 416. DOI: 10.1139/gen-2020-0131

Marcon Y and others, 'Orchestrating Privacy-Protected Big Data Analyses of Data From Different Resources with R and DataSHIELD' (2021) 17(3) *PLoS Computational Biology* e1008880. DOI:10.1371/journal.pcbi.1008880

McMahan B and others, 'Communication-Efficient Learning of Deep Networks From Decentralized Data.' (2017) Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR. 54. DOI: 10.48550/arXiv.1602.05629

Milne R and others, 'A Concentric Circles View of Health Data Relations Facilitates Understanding of Sociotechnical Challenges for Learning Health Systems and the Role of Federated Data Networks' (2022) *Front. Big Data* 5, 945739. DOI:10.3389/fdata.2022.945739

Mondschein Ch and Monda C, 'The EU's General Data Protection Regulation (GDPR) in a Research Context' in Kubben, P, Dumontier, P and Dekker, A (eds), *Fundamentals of clinical data science*'. (Springer Nature, 2019). DOI:10.1007/978-3-319-99713-1_5

Organization for Economic Cooperation and Development, 'OECD Guidelines on Human Biobanks and Genetic Research Databases' (2009).

Peloquin D and others, 'Disruptive and Avoidable: GDPR Challenges to Secondary Research Uses of Data' (2020) 28(6) *European Journal of Human Genetics* 697. DOI:10.1038/s41431-020-0596-x

Pormeister K, 'Genetic Data and the Research Exemption: Is the GDPR Going too Far?' (2017) 7 *Int. Data Priv. Law* 137. DOI: 10.1093/IDPL/IPX006

Reichel J, 'The GDPR and Processing of Personal Data for Research Purposes: What About Case Law?' (2021) 27(1) *Eur. Public Law* 167. DOI: 10.54648/euro2021007

Rossello S, Muñoz-González L and Díaz Morales R, 'Data protection by design in AI? The case of federated learning' (2021) 166 *Computerrecht* 273. https://ssrn.com/abstract=3879613 accessed on 14 May 2024.

Ruusalepp R, 'A Comparative Study of International Approaches to Enabling the Sharing of Research Data' (2008) https://era.ed.ac.uk/bitstream/handle/1842/3361/Ruusalepp%20Data_Sharing_Report.pdf?sequence=1 accessed on 28 September 2023

Saunders G and others, 'Leveraging European Infrastructures to Access 1 Million Human Genomes by 2022' (2019) 20(11) *Nature Reviews. Genetics* 693. DOI: 10.1038/s41576-019-0156-9.

Schneider G, 'Health Data Pools Under European Policy and Data Protection Law: Research as a New Efficiency Defence?' (2020) 11 JIPITEC 49. https://www.jipitec.eu/archive/issues/jipitec-11-1-2020/5082 accessed on 14 May 2024.

Segrelles Quilis D and others, 'A Federated Cloud Architecture for Processing of Cancer Images on a Distributed Storage' (2023) 139 *Future Gener. Comput. Syst*. 38. DOI: 10.1016/j.future.2022.09.019

Shabani M, Thorogood A and Murtagh M, 'Data access governance' in Graeme Laurie and others (eds), *The Cambridge Handbook of Health Research Regulation* (Cambridge University Press, 2021). DOI: 10.1017/9781108620024.023

Sheth A and Larson J, 'Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases' (1990) 22(3) *ACM Computing Surveys* 183. DOI:10.1145/96602.96604

Silva L, 'Federated Architecture for Biomedical Data Integration' (Dissertation, Universidade de Aveiro 2015) http://hdl.handle.net/10773/15759 accessed on 14 May 2024.

Staunton C and others, 'Appropriate Safeguards and Art. 89 of the GDPR: Considerations for Biobank, Databank and Genetic Research' (2022) 13 *Frontiers in Genetics*. DOI: 10.3389/fgene.2022.719317

Staunton C, Slokenberga S and Mascalzoni D, 'The GDPR and the Research Exemption: Considerations on the Necessary Safeguards for Research Biobanks' (2019) 27(8) *European Journal of Human Genetics* 1159. DOI: 10.1038/s41431-019-0386-5

Stephens K and others, 'Implementing Partnership-Driven Clinical Federated Electronic Health Record Data Sharing Networks' (2016) 93 *International Journal of Medical Informatics* 26. DOI:10.1016/j.ijmedinf.2016.05.008

TechDispatch #1/2022 – Federated Social Media Platforms, https://edps.europa.eu/data-protection/our-work/publications/techdispatch/2022-07-26-techdispatch-12022-federated-social-media-platforms_en accessed on 5 January 2024.

Thorogood A and others, 'International Federation of Genomic Medicine Databases Using GA4GH Standards' (2021) 1(2) *Cell Genom*. 100032. DOI:10.1016/j.xgen.2021.100032

Topol E, 'High-Performance Medicine: The Convergence of Human and Artificial Intelligence' (2019) 25(1) *Nature Medicine* 44. DOI:10.1038/s41591-018-0300-7

Tzortzatou O and others, 'Biobanking Across Europe Post-GDPR: A Deliberately Fragmented Landscape' in Slokenberga, S, Tzortzatou, O and Reichel, J (eds), *GDPR*

and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe (Springer Nature, 2021). DOI:10.1007/978-3-030-49388-2_22

Weber G, 'Federated Queries of Clinical Data Repositories: Scaling to a National Network' (2015) 55 *Journal of Biomedical Informatics* 231. DOI:10.1016/j.jbi.2015.04.012

World Economic Forum, 'Federated Data Systems: Balancing Innovation and Trust in the Use of Sensitive Data' (2019), https://www3.weforum.org/docs/WEF_Federated_Data_Systems_2019.pdf accessed on 9 November 2023.

World Economic Forum, 'Sharing Sensitive Health Data in a Federated Data Consortium Model: An Eight-Step Guide' (2020). https://www.weforum.org/reports/sharing-sensitive-health-data-in-a-federated-data-consortium-model-an-eight-step-guide/ accessed on 28 September 2023.

Wouters B and others, 'Putting the GDPR Into Practice: Difficulties and Uncertainties Experienced in the Conduct of Big Data Health Research' (2021) 7 *Eur. Data Prot. L. Rev*. 206. DOI: 10.21552/edpl/2021/2/9

Case law

C-360/13 *Public Relations Consultants Association Ltd v Newspaper Licensing Agency Ltd and Others* ECLI:EU:C:2014:1195.

C-25/17 *Jehovan todistajat – uskonnollinen yhdyskunta* ECLI:EU:C:2018:551.

C-40/17 *Fashion ID GmbH & Co.KG v Verbraucherzentrale NRW eV Fashion ID* ECLI:EU:C:2019:629.