Secure Infrastructures and Protection of Personal Data: Approaches for the Development of the NFDIxCS Research Data Management Container

Patrick Brunner [1,*], Stefan Bavendiek [2]

[1] FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Karlsruhe, Germany
[2] University of Hamburg, Hamburg, Germany
[*] Corresponding author: patrick.brunner@fiz-karlsruhe.de

This paper proposes a concept for the sharing of personal data for scientific research purposes and a potential role of the NFDIxCS infrastructure in that process. The concept takes account of the so-called Research Data Management Container (RDMC) that is developed by the NFDIxCS consortium. The RDMC should incorporate research data together with the corresponding metadata, the research software and the required execution environment. With regard to the proposed concept, the paper discusses some legal and technical issues concerning the transmission of personal data for scientific research purposes. The legal part of the paper addresses the purpose limitation principle that is stipulated, inter alia, by the GDPR. It analyses two distinct legal interpretations for the notion of data collection and further processing under the GDPR: a role-based approach and a data-based approach. The role-based approach ties these legal terms to the controller of the personal data, the data-based approach to the data itself. As explained in the paper, both approaches may be based on the wording of the GDPR provisions and recitals. The approaches lead to different legal conditions for the transmission of personal data. Our decision in favour of the data-based approach rests on a teleological interpretation of the purpose limitation principle. This is followed by a technical perspective of the requirements and existing solutions to address some of the discussed issues and legal requirements, with a focus on secure data sharing in the context of research containers. After an overview of the challenges has been provided, several potential solutions based on existing technology are presented.

Keywords: data sharing, personal data, purpose limitation principle, data security

Published in: Vincent Heuveline, Philipp Kling, Florian Heuschkel, Sophie G. Habinger, and Cora F. Krömer (Hrsg.): E-Science-Tage 2025. Research Data Management: Challenges in a Changing World. Heidelberg: heiBOOKS, 2025. DOI: https://doi.org/10.11588/heibooks.1652.c23932 (CC BY-SA 4.0).

1 Introduction

The NFDIxCS consortium has the goal of being a central infrastructure for and together with the computer sciences in Germany to promote the (inter-)national availability and reusability of research data from this field. For this objective, the NFDIxCS consortium is developing a so-called Research Data Management Container (RDMC) to enable researchers in the computer sciences to bundle their research data together with the corresponding research software and the required execution environment in a "time capsule". The RDMC will enable a sustainable, subsequent use of the research data that it holds. The long-term availability of research data is ensured in compliance with the FAIR principles.

The research data in the field of computer science is diverse. For example, it includes data that concerns the interactions between human beings and machines (Saporito et al. 2024). This kind of research data – if it relates to identifiable human beings – is personal data under data protection law.[1] Furthermore, some legal scholars have argued, that computer code in certain instances is personal data as it may relate to natural persons (Purtova and Leenes 2023).

A research project does not end with the acquisition of research findings; the communication of research results and research data is an essential part of scientific research. In general, scientific research builds upon the knowledge, produced by previous research findings. In that regard, it is necessary to document research results and keep the underlying research data available; on the one hand to ensure the verifiability of the research results and on the other hand to promote future scientific research. In contrast, the availability of personaldata must be limited to what is necessary in order to protect data subjects. Furthermore, data subjects must be provided with information and the means to exert their rights in relation to the processing of their personal data. Under EU Data protection law, research data that is personal data needs to be anonymized as soon as the research purposes allow it (Art. 89 para. 1 General Data Protection Regulation, GDPR). Long-term availability, transfer and subsequent use of personal data without a data protection concept is not compatible with EU data protection laws. The following concept is a proposition for the NFDIxCS infrastructure to facilitate access to and transfer of personal data for research purposes within the computer sciences.

[1] Keeping in mind that the highest competent authority for the interpretation of EU law, the Court of Justice for the European Union (CJEU), follows a broad understanding for the notion of personal data, cf. CJEU, judgment of 07.03.2024, OC, C-479/22, ECLI:EU:C:2024:215, para. 44 et seq.

2 Scope and limitation of this paper

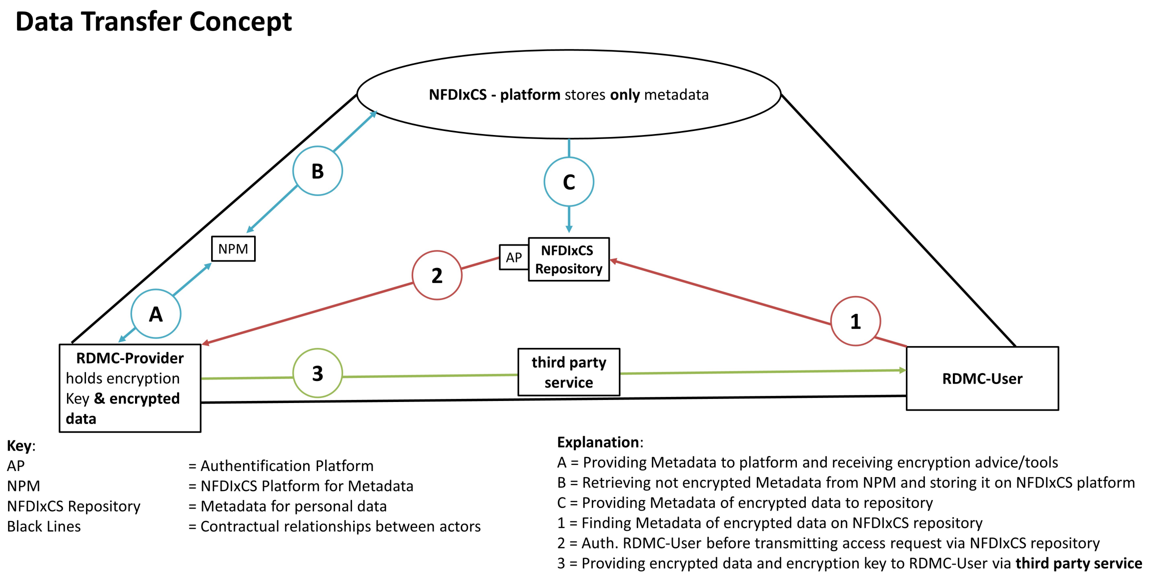
This paper proposes a data transfer concept for the NFDIxCS infrastructure, with the aim of facilitating the lawful sharing and access to personal data in research (Section 3). There are also other viable concepts for the sharing of personal data for scientific research purposes.

The process explanation should give the reader an overview of our proposed concept that is based on a NFDIxCS platform and a NFDIxCS repository that allows for the findability of research data from the computer sciences. With a view to the proposed concept, we then address some principles for the processing of personal data under the GDPR (Art. 5 GDPR) that need specific attention. From a legal perspective (Section 4), we will discuss the conceptual transfer of personal data in light of the purpose limitation principle (Art. 5 (1)(b) GDPR). From a technical perspective (Section 5), we present different solutions for ensuring the confidentiality and integrity of personal data (Art. 5 (1)(f) GDPR) during transmission.

3 Process explanation for the proposed data transfer concept

At the first level, the RDMC provider (the data holder) provides metadata for their research data that is personal data (in the following also referred to as research data) to the NFDIxCS platform via the 'NFDIxCS Platform for Metadata' (NPM) (A, see Figure 1). At the same time, the NPM suggests encryption methods or specific tools for encrypting the research data. The research data remains with the RDMC provider (e.g. on the servers of the research institute) and is not uploaded to the NPM. Next, the metadata is retrieved from the NPM, stored on the NFDIxCS platform (B), and sent to the NFDIxCS repository, where it becomes publicly viewable (C).

At the second level, the RDMC user (a user registered with the NFDIxCS platform) searches the NFDIxCS repository for relevant research datasets and finds relevant metadata (1). After verifying their identity via the 'Authentication Platform' (AP), the RDMC-User can send a standardised access request to the RDMC-Provider (2). The RDMC-Provider then checks the access request to ensure it is legally admissible, before deciding whether to send the encrypted research data and the encryption key to the RDMC-User via a secure third-party service (3).



Data Transfer Concept
NFDIxCS - platform stores only metadata
B
C
NPM
AP
NFDIxCS Repository
2
A
1
RDMC-Provider holds encryption Key & encrypted data
3
third party service
RDMC-User

Key:

AP	= Authentification Platform
NPM	= NFDIxCS Platform for Metadata
NFDIxCS Repository	= Metadata for personal data
Black Lines	= Contractual relationships between actors

Explanation:
A = Providing Metadata to platform and receiving encryption advice/tools
B = Retrieving not encrypted Metadata from NPM and storing it on NFDIxCS platform
C = Providing Metadata of encrypted data to repository
1 = Finding Metadata of encrypted data on NFDIxCS repository
2 = Auth. RDMC-User before transmitting access request via NFDIxCS repository
3 = Providing encrypted data and encryption key to RDMC-User via **third party service**

Figure 1: Graphic of the Data Transfer Concept.

4 Legal issues pertaining to the concept

This part of the paper addresses the principle of purpose limitation (Art. 5(1)(b) GDPR), which is a key legal issue in the conceptual transfer of personal data.

In the proposed concept, if the RDMC-Provider wishes to transmit personal data to an RDMC-User, the question arises as to whether this is permitted under data protection law. The following discussion is based on the premise that the RDMC-Provider initially collects personal data from data subjects for a specific research project, and later transmits it to an RDMC-User for another research project. This allows the research data to be reused in a scientific context.

4.1 The purpose limitation principle with regard to scientific research

Personal data is any information that relates to an identified or identifiably person, the data subject (Art. 4(1) GDPR). i.e., the usage, of personal data in the EU is regulated by multiple data protection laws, most importantly the General Data Protection Regulation (GDPR).

The subsequent use of personal data is particularly important under data protection law. The GDPR's purpose limitation principle stipulates that personal data shall be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes" (Art. 5(1)(b) GDPR). Data collection is an active activity by which the collecting entity obtains personal data for the first time (Roßnagel 2025a, DS-GVO Art. 4 Nr. 2, para. 15).

The purpose of the data collection establishes the basis for assessing the admissibility of data processing under data protection law and the associated obligations in a particular processing situation. Thus, the purpose specification serves, among other things, to assess the lawfulness of the processing within the meaning of Art. 6(1) GDPR.[2] Furthermore, the purpose specification creates transparency regarding the reason for the processing. It prevents data collection without cause.

The processing of personal data for scientific research purposes is broadly interpreted under the GDPR. In a non-exhaustive manner, the legislator explains that scientific research purposes should include "technological development and demonstration, fundamental research, applied research and privately funded research" (Recital 159). However, processing personal data for research purposes usually requires specification of a particular research project (European Data Protection Board 2018, para. 10).

4.2 Further processing personal data for secondary research purposes

Any processing activity after the initial data collection constitutes a further processing of personal data under the GDPR. If personal data is collected for one or more purpose(s) (hereinafter also referred to as: primary purposes) and is later processed for a different purpose, the latter constitutes a secondary use of the personal data. Under data protection law, this is referred to as a further processing of personal data for purposes, that were not specified at the time of the original data collection (hereinafter also referred to as: further processing for secondary purposes).

With regard to the purpose limitation principle, further processing of personal data for secondary purposes is permissible under the same conditions as the original processing of that data, if the data subjects to whom the data refer have consented to the processing of their personal data for the secondary purposes (Art. 6(4) GDPR). The technical implementation of dynamic consent models facilitates the ongoing inclusion of data subjects in processing activities concerning their personal data (Lay et al. 2024). These models can support the permissibility of further processing personal data for secondary purposes by enabling the data subjects to give and change their consent for a particular processing pur-

[2] Cf. CJEU, 'SS' SIA, C-175/20, ECLI:EU:C:2022:124, para. 66.

pose via a dedicated platform. However, there is a risk that data subjects, after initially providing their data, will not respond to consent requests for secondary purposes.

Further processing of personal data for secondary purposes is also permissible if the processing is based on a legal basis in Union or Member State law that constitutes a necessary and proportionate measure in a democratic society to protect the objectives set out in Article 23(1) GDPR (Art. 6(4) GDPR). These objectives are primarily those of general public interest, such as public security and public health.

Finally, further processing is permitted if the purposes of the original data collection are not incompatible with the secondary purposes. This generally requires a compatibility test, taking into account, inter alia, any links between the primary and secondary purposes, the context of the initial data collection, the sensitivity of the personal data concerned, the potential consequences of further processing for the data subjects, and the existence of appropriate safeguards (Article 6 (4) GDPR).

With regard to scientific research, the further processing of personal data for secondary purposes is privileged. Further processing of personal data for secondary scientific or historical research purposes is not considered incompatible with the purposes of the initial data collection (Art. 5(1)(b) GDPR), if appropriate safeguards are in place to protect the data subjects (Art. 89(1) GDPR). If this is the case, the compatibility between the primary processing purpose and the secondary scientific or historical research purposes is assumed without the need for a compatibility test.[3]

4.3 Safeguards for the processing of personal data for scientific or historical research purposes

Safeguards mentioned by Art. 89(1) GDPR to protect the data subjects are primarily the anonymisation and the pseudonymisation of personal data. Personal data must always be kept in a pseudonymised form during the research process, i.e., "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" (Art. 4(5) GDPR). Anonymisation is the process of changing personal data so that the data subjects can no longer be identified with "means reasonably likely to be used, (. . .) either by the controller or by another person" (Recital 26). Personal data must be anonymised as soon as the research purposes, for which the data is being processed, allow it (Art. 89(1) GDPR).

Additional safeguarding measures (cf. §§ 27(1), 22(2) of the German Bundesdatenschutzgesetz) include but are not limited to:

[3] Critical on the legal implications of Art. 5(1)(b) GDPR: Roßnagel 2025b, DS-GVO Art. 5, para. 109.

- encrypting personal data;
- controlling access to personal data (need-to-know principle);
- logging of access and processing activities;
- training for people involved in processing personal data;
- involving a data protection officer;
- securing the capabilities and the availability, integrity, confidentiality and resilience of systems used for data processing;
- establishing procedures to ensure compliance with data protection laws when personal data is further processed for other purposes (including by transmission of personal data);
- establishing a procedure for regular audits of the aforementioned measures.

4.4 A question of interpretation: The difference between data collection and further processing

Different interpretational approaches exist regarding the understanding of the interdependent concepts of data collection and further processing in data protection law. A distinction is made between a role-based and a data-based interpretation of the purpose limitation principle.

4.5 Role-based approach

The role-based approach of the purpose limitation principle (Becker et al. 2022), links the distinction between data collection and further processing to the controller. This approach is founded on the understanding that the GDPR's rules always relate to the processing of personal data and the purposes for which the data is being processed (Becker et al. 2022, p. 138). The controller decides the purposes and means by which personal data is being processed in a particular situation (Art. 4(7) GDPR). It is the controller's responsibility to comply with the GDPR's rules when processing personal data (Art. 5(2) and Art. 24(1) GDPR). Therefore, the notion of further processing personal data should be tied to the respective controller, who has previously collected the data for primary purposes (Becker et al. 2022, p. 138). Accordingly, if a controller obtains personal data for the first time, this is considered a data collection and not a further processing of personal data, regardless of whether the data has previously been processed by another controller.

For the proposed concept, this means that the transmission of personal data by the RDMC-Provider constitutes further processing of that data. This is because the reason for the transmission is to enable the subsequent research by the RDMC-User, which is a distinct purpose from the research purposes for which the data was originally collected by the RDMC-Provider. However, the data procurement by the RDMC-User represents a new initial data collection for the purposes specified in the data access request to the RDMC-Provider.

Different provisions and a recital of the GDPR support this approach:

Art. 5(1)(b) GDPR: Becker et al. base their approach on the wording of Art. 5(1)(b) GDPR, which sets out the purpose limitation principle by requiring personal data to be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes", but does not specify from which source personal data shall be collected. This allows for the interpretation that a "primary" collection of personal data may also occur, where data has previously been processed by another controller (Becker et al. 2022, p. 38).[4]

Art. 6(4)(b) GDPR: If personal data is to be further processed for secondary purposes, a compatibility test is usually required to ensure that the collection purposes and the secondary purposes are not incompatible, in accordance with the purpose limitation principle. As previously mentioned, this compatibility test must take into account various aspects of the original data collection and the further processing, including „the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller" (Art. 6(4)(b) GDPR). This indicates a connection between the original collection of personal data and the controller who wishes to further process personal data for secondary purposes. This suggests that further processing of personal data for secondary purposes is preceded by the collection of the data by the same controller.

Recital 50 GDPR: The legislator, having regard to the compatibility test, explains that 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 (. . .)" (emphasis by the author) the various aspects mentioned by Art. 6 (4) GDPR. Consequently, the legislator holds that a controller who wishes to further process personal data has previously processed the data for primary purposes, and that the lawfulness of the original data processing is a prerequisite for the permissibility of further processing that data.

[4] Cf. also Becker et al. 2022, The authors mention the "generic manner" of data collection, which is unspecific in relation to the data source, also with reference to the wording of Articles 17(1)(a) and 25(2) GDPR.

4.6 Data-based approach

The data-based approach of the purpose limitation principle, implicitly argued by the literature and data protection authorities (European Data Protection Board 2021, p. 6; Specht-Riemenschneider 2023, p. 651; Article 29 Working Party 2013, p. 26[5]), considers the data-lifecycle for the notion of further processing. According to this approach, any processing of personal data that has previously been collected by a controller constitutes further processing of that data, irrespective of the controller for the further processing operation. Thus, the procurement of personal data that has previously been processed by another controller qualifies as further processing and is subject to a compatibility assessment.

For the proposed concept, this would mean that the transmission of personal data by the RDMC-Provider to the RDMC-User constitutes a further processing operation. Likewise, the procurement and subsequent processing of personal data by the RDMC-User for their own research purposes and under their legal responsibility also constitutes further processing of personal data. It is therefore limited by the purposes for which the data was initially collected by the RDMC-Provider.

Under the GDPR, this approach may be based on the following provisions and the following recital:

Art. 5(1) GDPR: The principles for processing personal data refer explicitly to the personal data itself, allowing the interpretation that these principles pertain to the personal data itself. In this regard, Art. 5(1)(b) GDPR refers to the collection of personal data “for specified, legitimate and explicit purposes” as well as further processing without explicitly mentioning the respective controller as a decisive factor. The controller is, however, responsible for complying with these principles (Art. 5(2) GDPR).

Art. 17(2) GDPR: Under certain circumstances, data subjects have the right to demand that the data controller erases personal data related to them. This is the case, for example, if the personal data has been processed unlawfully. If the controller is obliged to erase personal data that they have published, they must take “reasonable steps” to “inform controllers which are processing the personal data that a data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data”

[5] The WP29 Opinion 03/2013 has not been officially endorsed by the EDPB, see: https://www.edpb.europa.eu/our-work-tools/general-guidance/endorsed-wp29-guidelines_en; *Visited on March 26, 2025.* With a view to the “determination of ‘purposes’ ”, the EDPB has referenced this WP29 opinion, see EDPB, 2020, Guidelines 05/2020 on consent under Regulation 2016/679, Version 1.1, page 13, footnote 28, available at: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en; *Visited on March 26, 2025.*

(Art. 17(2) GDPR). This indicates that a change of the data controller does not constitute new and independent data processing.

Art. 19 GDPR: Referring to Art. 19 GDPR, Frenzel argues that the controller who collected the personal data is obligated to convey the collection purposes to data recipients (Frenzel 2021, para. 29). Art. 19 GDPR obliges the controller to "communicate any rectification or erasure of personal data or restriction of processing" to each recipient to whom the personal data have been disclosed.

Recital 50 GDPR: With regard to the data collection and the purpose limitation principle, the first sentence of recital 50 GDPR refers to purposes "for which the personal data were initially collected", allowing the interpretation that once personal data has been collected for the first time, any subsequent processing is considered further processing of that personal data, irrespective of the controller.

4.7 Conceptual relevance of the different approaches

Both approaches are supported by the GDPR's provisions and (non-binding) recitals. A distinction between the two approaches is also relevant for the concept.

Under the role-based approach, the procurement of personal data by the RDMC-User from the RDMC-Provider is considered an initial data collection. Any further processing by the RDMC-User is limited by the purposes specified by the RDMC-User at the time of the procurement.

Under the data-based approach, the procurement of personal data by the RDMC-User from the RDMC-Provider is considered a further processing of that data. Any further processing by the RDMC-User is limited by the purposes for which the data was initially collected by the RDMC-Provider. In order to assess the legal permissibility of any further processing, the RDMC-User must be aware of the purposes and context of the initial data collection by the RDMC-Provider.

The data-based approach requires the RDMC-Provider to communicate the original data collection purposes to the RDMC-User to ensure that the data is further processed in compliance with the purpose limitation principle.

4.8 Evaluating the approaches based on a teleological interpretation of the purpose limitation principle

As demonstrated, the wording of the GDPR's provisions and recitals can support both approaches. In addition, the teleological interpretation takes into account the aim of the purpose limitation principle stipulated by Art. 5(1)(b) GDPR.

By specifying the processing purpose, the controller sets out the reasons for why they wish to process the personal data. This purpose specification ensures transparent and predictable data processing (Article 29 Working Party 2013, p. 13). It also establishes the basis for assessing other processing principles and related obligations (Article 5 GDPR) in a specific processing context.

Based on the specified and communicated purposes at the time of the data collection, the data subjects can decide whether to provide their data to the controller and whether they wish to influence the processing of their data by asserting their data subject rights. The permissibility of further processing personal data for secondary purposes must take into account, among other things, the reasonable expectations of the data subject based on the initial data collection (Article 6 (4)(b) and Recital 50 GDPR). The purpose limitation principle is therefore fundamental to the data subject's control of their personal data (cf. Recital 7 GDPR).

In this respect, the role-based approach is problematic. It leads to a break in the collection purposes when data is transferred from one controller to another who intends to process the data for secondary purposes. Under this approach, the transmission of personal data for secondary purposes leads to a shift in the basis that limits further processing in relation to the purposes of the initial data collection. This poses the risk that data subjects lose sight of, and control over, the purpose(s) for which their data is being processed. This is because data subjects may not be actively involved in the data transmission and the specification of the new collection purpose(s). This could happen, for example, if data subjects do not respond to consent requests sent to them. This loss of control continues and intensifies with subsequent data transmissions that lead to different collection purposes. Consequently, there is a risk that personal data will be processed for reasons unrelated to the initial purpose(s) for which it was provided by the data subjects and that the data subjects could not have anticipated such a subsequent processing of their personal data. From the data subject's perspective, the role-based approach allows for an unpredictable processing of their personal data.

The data-based approach limits further processing of personal data by any controller with regard to the purposes specified at the time of the initial data collection. If the data is initially collected directly from data subjects, they can assess the reasons and extent to which their data is being processed based on the specified purposes and decide accordingly whether to provide their data. However, even if the data is not collected directly from the data subjects, communicating the collection purposes enables them to assess any potential further processing of their data. Based on the specified purposes and related expectations by the data subjects, they can assert their rights.

The data-based approach does not prevent subsequent use of personal data from being efficient. Although the initial collection purposes are not inherent to the personal data itself, this can be addressed through the use of high-quality metadata sets that explain these purposes.

The data-based approach leads to a more effective protection of data subjects in relation to the processing of their personal data. It is consistent with the legislative aim of enabling data subjects' control of their personal data (Recital 7 GDPR).

For our concept we follow the data-based approach.

5 Technical issues pertaining to the concept

The exchange of private information between two parties over untrusted networks is a long-standing challenge in Computer Science. Regarding privacy, the primary security objective to achieve is confidentiality of data (Art. 5 (1)(f) GDPR), while integrity and authenticity are also part of the problem. Solutions involving applied cryptography have been developed, they also add additional challenges like key management and usability as the most prominent issues. In order to address this challenge within the NFDIxCS project, we are going to describe currently existing solutions that solve this objective. The key challenge that we face in NFDIxCS is that we as a platform provider do not want to have access to the plaintext data at any given time. This will include a significant risk by taking our infrastructure out of the chain of trust. If RDMC containers include sensitive data, we will provide the user with an option to request this data from the Data Owner. The protected data will however not be processed on the NFDIxCS platform. This leaves us with options to suggest for the Data Owner or third-party service providers.

The proposed options would allow the Data Owner to exchange the protected dataset with a privileged user securely, without that third part service gaining access to the plaintext data.

5.1 Potential solution: End-to-End Encrypted cloud storage

Several service providers offer solutions for end-to-end encrypted storage space. These solutions vary and come with different drawbacks that affect both the security and usability of the solutions. One possible solution, both for third party providers and for self-hosting, is the end-to-end encrypted app for the Nextcloud platform (Nextcloud GmbH 2021). This approach requires the users that intend to exchange encrypted data, to both have an established account registered with the same Nextcloud server. Additionally, the feature only works through the Nextcloud client applications on users end devices, while web access is out of scope. Consequently, this approach also does not allow exchanging data with unregistered users. The data is encrypted by generated 2048-bit RSA keys within the user's native client applications. The public key pairs can then be found by other registered users on the Nextcloud server and be used to encrypt data intended for the owner of the private key. In case the access to the key material or the end device is lost, a mnemonic passphrase is generated as a recovery key and offered for download to the user during the setup process. This however requires the user to manage this mnemonic passphrase securely and possibly for a long period of time, which presents one of the most difficult challenges on their own. Consequently, many users will either forget this recovery phrase or not store them securely. The use of 2048-bit RSA keys is also problematic, considering that current recommendations for RSA key length is already at 3072-bit (Bundesamt für Sicherheit in der Informationstechnik 2025).

Alternative solutions offered by service providers can be found with "Proton Drive" (Proton AG 2022) or the open-source solution provided by "Filen" (Filen Cloud Dienste UG 2025). Both providers implement encrypted cloud storage independent of the used end device and work in web browser sessions as well. While their internal implementations differ, they both make use of the user's password for both authentication and symmetric encryption. This allows any user to log in to their encrypted cloud storage through any end device or web browser and decrypt their data on their own device, either through a client application or within a browser session. Another advantage provided by these solutions is the option to share encrypted data by generating a URL with an additional generated password that is used to decrypt the shared file. While the URL is generated by the server application, the creation of the password and re-encryption of the shared file takes place on the users end device. Consequently, the end-to-end encryption is maintained while this approach also allows sharing the encrypted files with users outside the provider's registered user base. Unlike the previously described solution this approach is much more flexible since it allows secure file sharing without requiring specific client applications and even with yet unknown recipients.

5.2 Potential solution: Browser side encryption

Part of the above-described solutions was the option to generate URLs and keys to enable the download of files by users outside the hosting service. This option can also be implemented directly into an upload service without user registration and management. One such implementation can be found with "Firefox Send" which is now developed as a community drive fork (Visée 2025) of the original Mozilla project. This service enables a file upload interface that can be used to upload files which are then encrypted on the client within the browser session before the encrypted file is uploaded. A download URL is then generated with the decryption key as a parameter generated by the client. The hosting server only sees the encrypted data and is thereby removed from the chain of trust regarding the confidentiality of the contents. The advantage with this solution is the simplicity and lack of user management. A potential issue with this approach is the potential for abuse, which was also named as a reason for the shutdown of the original "Firefox Send" project (Mozilla Foundation 2023), which was faced with malicious users mounting malware and phishing attacks through the service. In the context of the NFDIxCS project, this issue could be addressed however by providing this service only to users that are authenticated by a research institution through the single-sign-on service like the academic cloud project (Gesellschaft für wissenschaftliche Datenverarbeitung 2025). For repeated download and persistent data management the previously discussed cloud storage solutions may provide a more suitable solution that integrated this client-side URL generation for e2ee download.

5.3 Unsolved problems

All proposed solutions include different drawbacks that need to be considered. The most pressing issue for solutions that require authentication is the use of user defined passwords for authentication and encryption. While the introduced solutions handle both parts appropriately in their implementation, the use of passwords is generally a significant security concern. In terms of authentication, it has long been known that passwords are a very weak method (Zviran and Haga 1999) that are commonly the target of attacks (Bundesamt für Sicherheit in der Informationstechnik 2019). The introduction of two-factor-authentication can help to mitigate some of these risks, but most 2fa methods will not address all these threats. A primary attack vector against authentication methods to this day is still the use of phishing (Hong 2012), which includes tricking the user into providing their password to a fake web site. Most second factor authentication methods are also vulnerable to phishing. One commonly available alternative that solves these issues entirely can be found with the Fido2 specification (Fido Alliance 2020), which allows for entirely phishing resistant authentication process. This method is already available in all commonly used browsers and operating systems, including mobile devices. The most widespread implementation of Fido2 can be found with "Passkeys", which has become popular on mobile platforms like Android and IOS.

5.4 Further Research

Future solutions need to address the problem of phishing resistant authentication while also providing an option for client-side encryption. Such a solution needs to be both convenient to use and secure to be implemented across common operating systems and devices. The Fido2 specification has already shown great progress in terms of authentication and is actively being pushed by large vendors to have received widespread adoption. Future work will address potential solutions based on this method that involves client-side encryption in common web applications, which will enhance the existing solutions and address remaining issues.

6 Conclusion

The transmission of personal data by the RDMC-Provider to the RDMC-User is a sensitive issue that needs consideration of various legal, ethical and technical aspects.

With regard to the purpose limitation principle, the transmission of personal data for secondary purposes constitutes a further processing of the personal data. Such a transmission is only legally permissible, if the purposes of the intended further processing are not incompatible with the purposes of the initial data collection. Processing personal data for historical or scientific research purposes is not incompatible with the initial data collection purposes, provided that necessary safeguards exist to protect the data subjects. Following the data-based approach of the purpose limitation principle, the RDMC-Provider must communicate the purposes of the initial data collection to the RDMC-User before transmitting the personal data. The RDMC-User must communicate the further processing purposes as well as the existing safeguards to the RDMC-Provider, who needs to assess the legal admissibility of the further processing before transmitting the personal data. The RDMC-Provider and the RDMC-User should lay down the binding conditions for the further processing of the personal data via a written contract. Both, the RDMC-Provider and the RDMC-User must inform the data subjects of the intended further processing of their personal data with enough time in advance for the data subjects to potentially assert their rights prior to the further processing.

For the transmission of protected data between the RDMC-Provider and RDMC-User, several technical solutions are available to design this process in a secure and reliable manner without requiring technical knowledge or training of the users. The use of end-to-end encryption will ensure that the potential for compromised data is limited to a minimum and avoid the common issues with cloud-based services and data protection.

Acknowledgements

The authors would like to thank the anonymous reviewers and the editors for their helpful feedback and comments on the manuscript. Furthermore, the authors would like to thank Dr. Dara Hallinan for his suggestions on linguistic improvements to the manuscript.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – (NFDI-52/1) – 501930651.

Authorship Contributions

- Part 1-4 of the paper is authored by Patrick Brunner.
- Part 5 of the paper is authored by Stefan Bavendiek.
- The abstract and Part 6 are co-authored.

Conflict of Interest

The authors work on the NFDIxCS project.

Bibliography

Article 29 Working Party. 2013. *Opinion 03/2013 on purpose limitation. WP 203.* Visited on March 26, 2025. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

Becker, Regina, Davit Chokoshvili, Giovanni Comandé, Edward S. Dove, Alison Hall, Colin Mitchell, Fruzsina Molnár-Gábor, Pilar Nicolàs, Sini Tervo, and Adrian Thorogood. 2022. "Secondary Use of Personal Health Data: When Is It "Further Processing" Under the GDPR, and What Are the Implications for Data Controllers?" *European Journal of Health Law* 30 (2): 129–157. https://doi.org/10.1163/15718093-bja10094.

Bundesamt für Sicherheit in der Informationstechnik. 2019. *The state of IT security in Germany in 2019.* Technical report. Visited on June 27, 2025. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/Securitysituation/IT-Security-Situation-in-Germany-2019.pdf?__blob=publicationFile&v=3.

———. 2025. *Cryptographic Mechanisms: Recommendations and Key Lengths.* Technical report. Visited on June 27, 2025. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/TechGuidelines/TG02102/BSI-TR-02102-1.pdf.

European Data Protection Board. 2018. *Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak.* Visited on March 26, 2025. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-032020-processing-data-concerning-health-purpose_en.

European Data Protection Board. 2021. *Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research.* Visited on March 25, 2025. https://www.edpb.europa.eu/our-work-tools/our-documents/other-guidance/edpb-document-response-request-european-commission_en.

Fido Alliance. 2020. *FIDO UAF Architectural Overview.* Visited on March 25, 2025. https://fidoalliance.org/specs/fido-uaf-v1.2-ps-20201020/fido-uaf-overview-v1.2-ps-20201020.pdf.

Filen Cloud Dienste UG. 2025. *Documentation: Cryptography.* Visited on March 27, 2025. https://docs.filen.io/docs/api/guides/cryptography/.

Frenzel, Eike Michael. 2021. "DS-GVO Art. 5 Grundsätze für die Verarbeitung personenbezogener Daten". In *Datenschutz-Grundverordnung Bundesdatenschutzgesetz: DS-GVO BDSG,* 3rd edition, edited by Boris P. Paal and Daniel A. Pauly. München: C. H. Beck. ISBN: 978-3-406-75374-9.

Gesellschaft für wissenschaftliche Datenverarbeitung. 2025. *Academic Cloud.* Visited on March 27, 2025. https://academiccloud.de.

Hong, Jason. 2012. "The state of phishing attacks". *Communications of the ACM* 55 (1): 74–81. https://doi.org/10.1145/2063176.2063197.

Lay, Winnie, Loretta Gasparini, William Siero, and Elizabeth K Hughes. 2024. "A rapid review of the benefits and challenges of dynamic consent". *Research Ethics* 21 (1): 180–202. https://doi.org/10.1177/17470161241278064.

Mozilla Foundation. 2023. *What happened to Firefox Send?* Visited on March 27, 2025. https://support.mozilla.org/en-US/kb/what-happened-firefox-send.

Nextcloud GmbH. 2021. *Nextcloud end-to-end encryption RFC.* Visited on March 27, 2025. https://github.com/nextcloud/end_to_end_encryption_rfc/blob/master/RFC.md.

Proton AG. 2022. *The Proton Drive security model.* Visited on March 27, 2025. https://proton.me/blog/protondrive-security.

Purtova, Nadezhda, and Ronald Leenes. 2023. "Code as personal data: implications for data protection law and regulation of algorithms". *International Data Privacy Law* 13 (4): 245–266. https://doi.org/10.1093/idpl/ipad019.

Roßnagel, Alexander. 2025a. "DS-GVO Art. 4 Nr. 2 Begriffsbestimmung "Verarbeitung"". In *Datenschutzrecht,* 2nd edition, edited by Spiro Simitis, Gerrit Hornung, and Indra Spiecker. Nomos Verlag. ISBN: 978-3-8487-8958-0, visited on June 27, 2025. https://beck-online.beck.de/Bcid/Y-400-W-SimHorSpiKoDatenSchR-G-EWG_DSGVO-A-4-N-2.

———. 2025b. “DS-GVO Art. 5 Grundsätze für die Verarbeitung personenbezogener Daten”. In *Datenschutzrecht,* 2nd edition, edited by Spiro Simitis, Gerrit Hornung, and Indra Spiecker. Nomos Verlag. ISBN: 978-3-8487-8958-0.

Saporito, Antonio, Parinaz Tabari, Mattia De Rosa, Vittorio Fuccella, and Gennaro Costagliola. 2024. “Exploring the Privacy Horizons: A Survey on HCI and HRI”. In *Computational Science and Its Applications – ICCSA 2024 Workshops,* 113–125. Springer Nature Switzerland. ISBN: 9783031653186. https://doi.org/10.1007/978-3-031-65318-6_8.

Specht-Riemenschneider, Louisa. 2023. “Datennutz und Datenschutz: Zum Verhältnis zwischen Datenwirtschaftsrecht und DSGVO”. *Zeitschrift für Europäisches Privatrecht,* number 3, 638–672.

Visée, Tim. 2025. *Send.* Visited on March 27, 2025. https://github.com/timvisee/send.

Zviran, Moshe, and William J. Haga. 1999. “Password security: an empirical study”. *Journal of Management Information Systems* 15 (4): 161–185. Visited on June 27, 2025. https://hdl.handle.net/10945/40319.