From Hospitals to Researchers: A Data-Trustee Infrastructure to Search and Use FHIR-Data for Retrospective Medical Research

Carolin Poschen
Department of Computer Science
Trier University of Applied Sciences
Trier, Germany
email: c.poschen@hochschule-trier.de

Joscha Grüger
Experience-based Learning Systems
German Research Center for Artificial Intelligence
Trier, Germany
email: joscha.grueger@dfki.de

Britta Berens
Department of Computer Science
Trier University of Applied Sciences
Trier, Germany
email: b.berens@hochschule-trier.de

Helene Christ
BI & Analytics
Dedalus HealthCare GmbH
Trier, Germany
email: helene.christ@dedalus.com

Lukas Meyer
BI & Analytics
Dedalus HealthCare GmbH
Trier, Germany
email: lukas.meyer@dedalus.com

Konstantin Knorr
Department of Computer Science
Trier University of Applied Sciences
Trier, Germany
email: k.knorr@hochschule-trier.de

***Abstract*—The secondary use of clinical data is crucial for advancing medical research, yet it remains challenged by data fragmentation, privacy concerns, and limited availability. This paper presents a data-trustee infrastructure designed to enable secure, privacy-preserving access to retrospective medical data stored in Hospital Information Systems (HIS). The infrastructure leverages Fast Healthcare Interoperability Resources (FHIR) standards to ensure interoperability and employs a modular pipeline. The pipeline extracts, preprocesses, encrypts, and annotates the data in the hospital and then stores data in a trustee repository. A central component—the Study Specification Board—facilitates ethical and formalized study planning, while a privacy-preserving, two-phase search mechanism allows researchers to retrieve relevant data without exposing sensitive information. A demonstrator system has been implemented and successfully integrated with an HIS, confirming the feasibility and practical applicability of the approach. This work represents a significant step toward operationalizing secure clinical data sharing aligned with EU-GDPR and the goals of the European Health Data Space.**

Keywords-medical research; data-trustee infrastructure; data access; privacy; security.

I. Introduction

In recent years, Artificial Intelligence (AI) has brought significant and disruptive changes across various sectors, including healthcare [1]. However, in the medical field, the integration of AI has so far achieved only limited success [2]. This can be partially attributed to the substantial data requirements for training effective AI models [1], which pose a challenge in healthcare due to the sensitive nature of medical data and associated privacy concerns. Additionally, technical barriers persist: medical data are typically generated and stored across numerous hospitals and private practices, resulting in heterogeneous and fragmented data silos with inconsistent formats and limited interoperability [3]. Serving the objectives of the EU Data Strategy, the European Health Data Space strives to unify these fragmented data silos, relying on technical and semantic interoperability [4]. These data spaces would also provide access to large and robust datasets, which are crucial to train AI models [5].

The secondary use of clinical data is increasingly valued as a vital tool for enhancing healthcare and advancing medical research. Using clinical data for medical research offers several key advantages. Since the data are already collected during routine patient care, they are readily available, cost-effective, and eliminate the need for additional patient involvement or physical intervention. This real-world data enables large, diverse sample sizes, making it especially valuable for studying rare diseases [6]. However, secondary analysis of raw health records poses significant challenges, as the data were initially collected for clinical care rather than research. Researchers must navigate fragmented databases, inconsistent representations of clinical concepts, and changes in coding practices over time, all of which complicate data access and preparation [7]. To overcome these challenges and to ensure secure, trustworthy, and legally compliant access to health data, the concept of data trustees has been proposed [8]. Additionally, data trustees can serve as data spaces as proposed by the European Union and outlined above.

This paper introduces a secure data pipeline and a Data-Trustee Infrastructure (DTI) designed to facilitate privacy-compliant secondary use of medical data. Our approach enables the controlled transfer of data from hospitals to researchers through a data trustee, an intermediary that manages and forwards data without having direct access to its contents. Within our pipeline, medical data are collected and preprocessed securely within the hospital's internal infrastructure, then encrypted and stored in a central repository. Descriptive metadata for each data entry are created to keep a general description while storing the original data encrypted. Researchers may access these data only for specific studies that have received approval from an ethics committee. An automated process translates the approved study's data requirements into search parameters and queries the describing data set. Access to the

original, encrypted data and their corresponding keys is granted only if the number of matching records exceeds a predefined threshold, ensuring both data utility and privacy protection.

The remainder of the paper is structured as follows: Related work in the fields of sharing and accessing medical data for research is discussed in Section II. Section III details the architecture of our DTI and the complete pipeline for data in the system, as well as how researchers interact with it to access data. A discussion of our work follows in Section IV, Section V concludes the paper and outlines potential future work.

II. Related Work

When sharing and using medical data for research, it is important to balance the opportunity provided by data against the individual's right to control their own data [9]. With that in mind, [10] presents a review of research on patient perspectives regarding data sharing, covering their motivations, concerns, privacy considerations, and conditions for sharing. Druedahl et al. concluded that hearing patients' voices is crucial for public acceptance, inclusion, and equity in data sharing.

The German Medical Informatics Initiative (MII) [11] established a decentralized, FHIR-based, federated research data infrastructure based on local Data Integration Centers (DIC) at university centers and partner locations, which extract, pseudonymize, and harmonize clinical data using a modular core dataset defined with international standards. Analyses are performed either centrally—based on a harmonized Broad Consent—or via federated learning, where containerized algorithms are distributed to local sites (data-in-place approach). The German Research Data Portal for Health (FDPG) serves as a central entry point for researchers, offering metadata browsing, feasibility queries, and cohort selection. Though different research projects have already requested data through the MII infrastructure, their data application process still requires substantial manual rework and communication between DIC, FDPG staff, and data requesters as described in [12]. Our proposal minimizes the manual rework and communication overhead by storing data centrally, reducing the number of involved parties for data requests and a simplified and intuitive data requirement description.

A data-trustee architecture for medical sleep research data is presented in [13]. Their architecture enables secure, decentralized data sharing based on dynamic patient consent. A key feature of their system is a standardized, FHIR-based feasibility query that allows researchers to search for relevant data before submitting formal access requests. Combined with containerized analysis environments and tamper-proof logging, the platform addresses legal, ethical, and technical challenges in secondary data use. In addition to addressing these challenges, our approach relies on a separate but centralized data storage architecture, aiming to automate as many steps as possible.

The concept of data trusts or data trustees is discussed in different works. While [14] argues for a variety of data trusts, so that data subjects can choose the most suitable one, [15] aims to answer the question "What are design features that assist practitioners in the secure and sovereign selection process of finding a data trustee in a data space?". When designing data trustee models, [16] identifies four ideal-typical archetypes for data trustees in healthcare, namely data brokerage, processing, aggregation, and custody trustees, which differ along their defined meta-dimensions (1) Task & People, (2) Technology, and (3) Structure.

In [17], the authors propose data trusts as a service using blockchain, which they claim may enable transparent data sharing between multiple stakeholders. To share electronic medical records of the same patients between different hospitals, [18] proposes a blockchain-based information system, MedBlock, as an efficient and privacy-preserving scheme to share data between hospitals. However, as [19] points out in their discussion on leveraging blockchain for healthcare data management systems, the integration of blockchain with healthcare systems generates some challenges, such as interoperability, complexity, or integration with existing systems.

Many works discuss the use of Electronic Health Records (EHR) for medical research [7], [20]–[22]. For example, [20] explored challenges and opportunities of sharing and reusing EHR data for clinical research during the COVID-19 pandemic. They highlight limited syntactic and semantic interoperability, regional privacy regulations, and emerging data protectionism as key barriers. To address privacy regulations and prevent uncontrolled data use, they emphasize the role of a data steward who enforces policies to support institutions in overseeing data sharing both legally and comprehensively. To enable retrospective analysis using EHR data, [7] presents a seven-step data preparation workflow, ranging from obtaining an overview of available data, over extracting relevant data, to implementing a data processing pipeline. Although their work discusses different issues regarding the access and preparation of data for secondary use, it is based on the experience of a single hospital, and the workflow would need potential adjustment for different hospitals. Likewise, [21] proposes an automated framework to transform clinical data into Findable, Accessible, Interoperable, Reusable (FAIR) research data. The implementation targets a maximum-care university hospital, yet, as in prior cases, remains institution-specific and may require adaptation for broader applicability. Similar to our approach, [22] proposes a pipeline to convert EHR data into FHIR standard to support AI research. Their workflow comprises five steps: querying hospital databases, mapping data to FHIR, validating the output, transferring it to a database, and exporting it in an AI-friendly format. However, the authors do not address anonymization or pseudonymization, and instead store all data in plain text within a single database.

This work is an extension to [23]. The previous work focused on the architecture design of the main DTI-components. In contrast, this work focuses on the data flow through the pipeline and the researchers' interaction with the system, including the components in the hospital's and researcher's network.

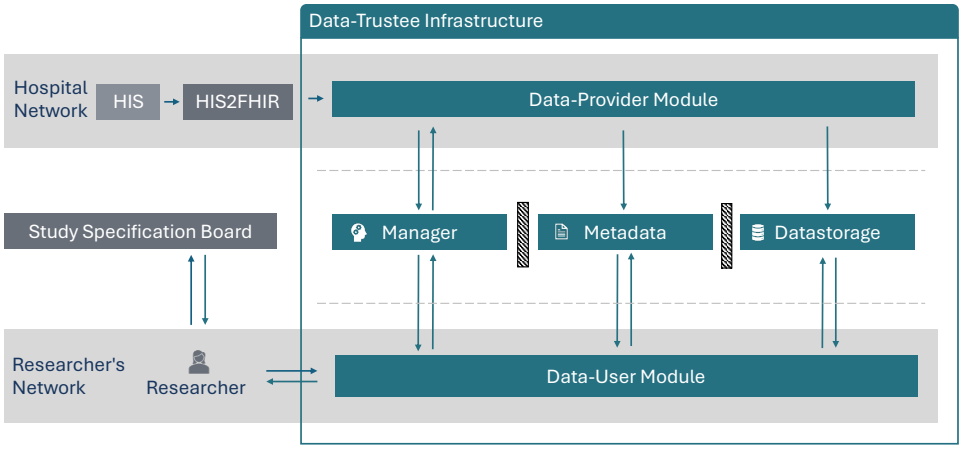

Figure 1. Architecture of the Data-Trustee Infrastructure, as well as the components in the hospital network and the SSB. The arrows indicate data flow between hospital systems, the data-trustee modules, the SSB, and researchers.

III. The Data Pipeline through Our Data-Trustee Infrastructure

The data pipeline through our data-trustee infrastructure, whose architecture was first proposed in [23] and shown in Figure 1, starts in a **Hospital**, where medical data are created during patient care. These data are transformed into the Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR) format by the **HIS2FHIR** module that sends data to the **Data-Provider Module**. This module, still deployed in the hospital's local network, carries out further preprocessing, i.e., splitting, encrypting, and annotating data entries. Split data are sent to and stored in the three independent **core modules** of the DTI. A **Study Specification Board** (SSB) helps researchers to formally define study requirements, especially cohort definitions. The SSB operates independently of the DTI and is deployed externally. Though it supports the usage of the DTI through the provision of *Study Specification Documents* (SSDs), it can be used independently of our DTI. The created SSD is used by the **Data-User Module** to carry out a two-phase, three-party, privacy-preserving search. Upon successful completion of this search, researchers obtain relevant data to conduct their retrospective research.

Any interaction between participating units must be operated on a legal basis, such as contracts and compliance.

A. Data Format and Metadata

At all stages of the DTI, medical data are stored and processed in the HL7 FHIR format, the leading interoperability standard for data exchange in healthcare [24]. The FHIR format consists of different resources, like `Patient` or `Encounter`, which contain the relevant information for the specific resource. The `Encounter` resource stores information on the period, type, class, status, and diagnoses of an encounter, while the `Patient` resource contains demographic and identifying information of a specific patient.

During a preprocessing step in our proposed pipeline, medical data are split into data entries, each containing one resource along with its corresponding information. Each medical entry is annotated with descriptive information, referred to as *Metadata*, which provide generalized and categorized information about the original entry. Such data classification can be the storage of an interpretation of a value, e.g., high blood pressure taking factors such as age and sex into account, instead of the actual numerical value. Generalization may involve taxonomy generalizations such as truncating ICD-10 codes, e.g., storing E10 instead of E10.1. These steps allow us to encrypt the original medical data while maintaining a general description of data in an unencrypted format.

B. Study Specification Board

The interaction of a researcher with the DTI is driven by the need to efficiently access retrospective patient data. To get access to those data, researchers first input information about their planned study including their formalized study parameters into the *Study Specification Board (SSB)*. This also includes a positive vote of an ethics committee on their planned study, in accordance to European Union - General Data Protection Regulation (EU-GDPR) Recital 33, which states that sensitive data must be used in compliance with recognized ethical standards. In addition, some national laws require ethics approval before data access is granted.

The formalization of study parameters, previously suggested in [25] and [26], aims to simplify the definition of cohorts, which can then be used for the search of data in a given database, and the subsequent publication of the study. The SSB is designed to provide a user-friendly interface that does not require knowledge of the FHIR format, but rather provides an intuitive approach of data formalization as proposed in [25]. The formalization is based on metadata.

Once the study proposal and its ethics vote have been evaluated in the SSB, the study is published on the SSB. Additionally, a *Study Specification Document (SSD)* is generated that transforms formalized study parameters into a FHIR-compliant format. The SSD forms the basis for delegating the search of relevant data (cf. Section III-F).

C. Hospital

Digital data recorded in a hospital are stored in the HIS as part of the treatment. For secondary use of data, additional consent is required, therefore Broad Consent [27] is used for consent acquisition. This allows for the storage and secondary use of patient data, without tying the consent to a specific study. Patients are approached during discharge. This allows for the consent process to be integrated into the existing workflow and places the patient in a less stressful situation to consider data donation. The consent document is machine-readable, enabling automatic conversion into the FHIR format.

Once consent is given, a patient's data can be extracted for secondary use by the HIS2FHIR module. To standardize the data, semantic mappings to established medical terminologies are introduced. For example, laboratory values are mapped to LOINC. Data themselves are collected on case level and transformed into the FHIR format. It ensures syntactical correctness and enables the merging of data from different hospitals, although this procedure does not consider the semantic correctness of the data. Regarding the secondary use of data, completed cases ensure data consistency, since they are not likely to change retrospectively. Therefore, only discharged cases are considered. Every case of a patient within a hospital, for which consent has been provided, is extracted. In every subsequent export, only the most recent completed case is extracted, minimizing the extracted data per export. Extraction is scheduled, ensuring it occurs consistently at the same daily time. The extracted data are further processed by the Data-Provider Module.

D. Data-Provider Module

The Data-Provider Module (DPM) is the entry point to the DTI for a hospital. It is uniquely configured for each hospital, with its digital identity directly embedded, and deployed in its local network. It receives medical data as FHIR bundles from the HIS2FHIR module. These FHIR bundles are split into demographic and medical data, and the patient's consent document. Demographic data (DDAT) consist of information stored in the resource type `Patient`, while medical data (MDAT) consist of resource types like `Encounter` or `Observation`.

Data stored in the `Patient` resource are pseudonymized by removing all direct identifiers; information such as the patient's gender, birth year, or their truncated postal code are kept. Each patient is assigned a pseudonym to enable privacy-preserving record linkage if new data of the same patient become available [28]. Additionally, the patient resource is enriched with IDs of each corresponding MDAT entry. In the consent document, the identity of the patient is replaced by their pseudonym.

Each MDAT entry is encrypted (eMDAT) using a newly created, unique Data Encryption Key (DEK). The DEK is a symmetric key of a state-of-the-art cryptographic scheme. To ensure searchability of encrypted medical data while aligning with the formalized study parameters defined in the SSB, a metadata record is created for each MDAT entry as described in Section III-A. Thus, each eMDAT entry and its corresponding metadata entry are assigned the same newly generated unique identifier to ensure consistent linkage. Finally, the DPM sends all documents to their corresponding DTI-core modules.

This preprocessing step is performed within the hospital's local network, yet inside the DTI. This enables a secure split, encryption, and annotation of data prior to their storage in dedicated and physically separated modules. The split is essential for the privacy-preserving design of the DTI and data are only merged again by the researcher.

E. DTI-Core Modules

The DTI at its core consists of three modules, the **Manager** module, the **Metadata** module, and the **Datastorage** module. The DTI operates under a strict separation-of-concerns model: no single module has access to both patient identity and medical content. This architectural principle ensures that sensitive associations can only be reconstructed by the authorized researcher within the Data-User Module. They primarily function as independent storage modules with minimal business logic. However, whenever a request is processed, all steps are authorized, and all returned data are signed with a digital signature of the respective module. This ensures authenticity and integrity of results, particularly when they are forwarded to other modules.

The Manager module consists of three different services, each responsible for a different purpose. The *identity service* stores patient pseudonyms together with their DDAT and MDAT IDs, DEKs are stored in the *key service*, and the consent in the *consent service*. The Manager module enables the search on DDAT based on the given consent and returns all MDAT IDs of patients that fit the search criteria, as further described in Section III-F. After a successful search, the Manager module also provides the corresponding DEKs for all found MDAT IDs and issues a signed receipt of all downloadable eMDAT.

The Metadata module stores metadata provided by multiple DPM and enables querying, allowing searches for relevant metadata. Similarly, the Datastorage module stores eMDAT and returns them for given IDs obtained by the search, provided that the present signed receipt is verified as issued by the Manager module.

F. Data-User Module

The Data-User Module (DUM) is the entry point for the researcher to the DTI. Following a successful evaluation of a study proposal within the SSB, a dedicated instance is uniquely generated for the approved research purpose and made available to the researcher. The SSB exports all formal cohort definitions as a Study Specification Document (SSD), a FHIR-compliant format that is directly embedded in the DUM. The SSD must not be modified; otherwise, the entire module is invalidated. This is enforced by integrity-preserving measures, using digital signatures. Once integrity is ensured, the SSD is used to carry out a privacy-preserving, two-phase, three-party search, first proposed in [29] and formalized in [23]. In the first phase, each SSD cohort definition is split into multiple search queries. Queries related to patient demographics are sent to the Manager

module, which returns the MDAT IDs of patients that match the query criteria. Simultaneously, queries concerning medical data are sent to the Metadata module, which responds with IDs of matching MDAT entries. All returned IDs are grouped by patients, and the system determines which patients match, i.e., meet the entire set of specified criteria. In the second phase of the search algorithm, the eMDAT and their DEKs, along with the DDAT of matching patients are requested from the Datastorage and Manager modules respectively. Within the DUM, eMDAT are decrypted and can be used by the researcher.

The search procedure is designed to be privacy-preserving by enforcing a strict separation of data domains. Only the authorized researcher is able to reconstruct the linkage between demographic and medical data. The Manager module operates exclusively on pseudonymized demographic data and associated identifiers without access to any clinical content. In contrast, the Metadata module processes generalized medical metadata without knowledge of patient identities. At no point can either module independently infer complete patient-level information, thereby preventing re-identification risks, while only providing data specifically for a study upholds data minimization.

IV. Discussion

Our proposed DTI offers a practical and privacy-preserving approach that facilitates the secondary use of clinical data for retrospective medical research. A demonstrator implementing the system design has been developed and evaluated, integrated with an HIS, confirming its feasibility and suitability for practical use in clinical research environments.

A. Strengths and Contributions

A key contribution of this work is the development of a nearly fully-automated pipeline that enables the secure transfer of patient data from hospitals to researchers. By leveraging a modular, standardized architecture based on HL7 FHIR, we enhance interoperability and reduce the technical integration burden across institutions. The process—from in-hospital data preprocessing, encryption, and metadata annotation, to study-specific data retrieval by researchers—is handled in a streamlined, privacy-conscious manner.

Our infrastructure supports researchers in accessing initially distributed datasets via a unified system. The SSB and DUM simplify study setup and automate the formalization and translation of cohort definitions into FHIR-compatible search parameters. This ensures legal and ethical compliance (e.g., with EU-GDPR and ethics committee approval) and reduces researcher workload and administrative overhead.

Furthermore, the two-phase, three-party, privacy-preserving search mechanism ensures that the DTI-core modules cannot infer sensitive links between patient identities and medical content. Only the researcher, within their working environment, can decrypt and reconstruct the data necessary for their approved study.

B. Limitations and Challenges

Despite these strengths, several limitations remain that could affect scalability and adoption. First, participation from hospitals requires technical integration efforts, including the deployment of specific components such as customized HIS2FHIR and DPM. Secondly, each research project requires its own DUM instance. Although this leads to a certain amount of additional work, it is limited in time, as the DUM can be taken out of operation again once the data has been delivered.

Another practical limitation is the manual verification of actors and identities at onboarding. While this step is common across most trusted data-sharing ecosystems, it remains a bottleneck and may benefit from future integration with national digital identity systems.

Moreover, the system currently depends on metadata for search operations. While this approach supports general cohort definitions and preserves privacy, it limits the granularity and specificity of data queries. Highly specialized or narrow study parameters may not be captured by available metadata alone.

Finally, data accessed through the DTI are not fully anonymized. Although encryption, access control, and legal contracts serve as safeguards against re-identification, the lack of guaranteed anonymization represents a residual privacy risk that must be addressed through governance and compliance measures.

V. Conclusion and Future Work

This work provides a concrete and extensible blueprint for operationalizing the principles of the European Health Data Space. It tackles key challenges such as patients' consent and heterogeneous data formats. A two-phase, three-party, privacy-preserving search algorithm guarantees that the patients' data can only be combined by the researcher. This ensures that the other parties cannot access the data, and the researchers are only able to use data they have permission to. While the data are stored centrally, they are split into distinct components. This design allows searches to be performed solely on the metadata, completely isolating the encrypted raw data from the querying process. Furthermore, we automated the entire processes of getting data from hospitals to requesting data for research, thereby eliminating a bottleneck in retrospective medical research.

Future enhancements could include:

- Integration of outpatient care data and general practitioners. This can be achieved by deploying a module similar to the HIS2FHIR component in the practitioner's system.
- Support for the re-import and analysis of research outcomes to promote learning healthcare systems.
- Semantic enrichment of data and improved quality checks to ensure plausibility and consistency.
- Mechanisms for patient-driven consent management and dynamic revocation.
- More expressive query languages for SSDs, potentially combined with privacy-preserving computation techniques like secure multi-party computation or federated analytics.
- Systematic and quantitative evaluation of the DTI.

Overall, while challenges remain, our infrastructure represents a significant step toward bridging the gap between clinical

data silos and the data needs of modern AI-driven healthcare research.

ACKNOWLEDGEMENT

This work is part of the DaTreFo project, funded by the German Federal Ministry of Research, Technology, and Space (16KIS1644).

REFERENCES

[1] A. B. Rashid and A. K. Kausik, "AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications", *Hybrid Advances*, p. 100 277, 2024. DOI: 10.1016/j.hybadv.2024.100277.

[2] A. Zahlan, R. P. Ranjan, and D. Hayes, "Artificial intelligence innovation in healthcare: Literature review, exploratory analysis, and future research", *Technology in Society*, vol. 74, p. 102 321, 2023. DOI: 10.1016/j.techsoc.2023.102321.

[3] T. K. Eisinger-Mathason *et al.*, "Data linkage multiplies research insights across diverse healthcare sectors", *Communications Medicine*, vol. 5, no. 1, p. 58, 2025. DOI: 10.1038/s43856-025-00769-y.

[4] C. Stellmach, M. R. Muzoora, and S. Thun, "Digitalization of Health Data: Interoperability of the Proposed European Health Data Space ", in *Digital Professionalism in Health and Care: Developing the Workforce, Building the Future*, IOS Press, 2022, pp. 132–136. DOI: 10.3233/SHTI220922.

[5] I. Ulnicane, "Artificial Intelligence in the European Union: Policy, ethics and regulation", in *The Routledge Handbook of European Integrations*, Taylor & Francis, 2022. DOI: 10.4324/9780429262081-19.

[6] M. Jungkunz, A. Köngeter, K. Mehlis, E. C. Winkler, and C. Schickhardt, "Secondary Use of Clinical Data in Data-Gathering, Non-Interventional Research or Learning Activities: Definition, Types, and a Framework for Risk Assessment", *Journal of Medical Internet Research*, vol. 23, no. 6, e26631, 2021. DOI: 10.2196/26631.

[7] A. Maletzky *et al.*, "Lifting Hospital Electronic Health Record Data Treasures: Challenges and Opportunities", *JMIR Medical Informatics*, vol. 10, no. 10, e38557, 2022. DOI: 10.2196/38557.

[8] S. Kilz and M. Radic, "Health Data Trustees: A Business Model Perspective", in *The International Conference on Innovations in Computing Research*, Springer, 2024, pp. 618–630.

[9] T. Hulsen, "Sharing Is Caring—Data Sharing Initiatives in Healthcare", *International Journal of Environmental Research and Public Health*, vol. 17, no. 9, p. 3046, 2020. DOI: 10.3390/ijerph17093046.

[10] L. C. Druedahl and S. Kälvemark Sporrong, "Patient Perspectives on Data Sharing", in *The Law and Ethics of Data Sharing in Health Sciences*, Springer, 2023, pp. 51–67. DOI: 10.1007/978-981-99-6540-3_4.

[11] S. C. Semler *et al.*, "The Medical Informatics Initiative at a glance-establishing a health research data infrastructure in Germany", *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, pp. 616–628, 2024. DOI: 10.1007/s00103-024-03887-5.

[12] H.-U. Prokosch *et al.*, "Towards a National Portal for Medical Research Data (FDPG): Vision, Status, and Lessons Learned", in *Caring is Sharing — Exploiting the Value in Data for Health and Innovation*, IOS Press, 2023, pp. 307–311. DOI: 10.3233/SHTI230124.

[13] R. Burmeister, C. Erler, F. Gauger, R. J. Dressle, and B. Feige, "Advancing Sleep Research Through Dynamic Consent and Trustee-Based Medical Data Processing", ICDS, 2024, ISBN: 978-1-68558-169-5.

[14] S. Delacroix and N. D. Lawrence, "Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance", *International Data Privacy Law*, vol. 9, no. 4, pp. 236–252, 2019. DOI: 10.1093/idpl/ipz014.

[15] M. Steinert, D. Tebernum, and M. Hupperz, "Design Features for Data Trustee Selection in Data Spaces", in *International Conference on Data Science, Technology and Applications 2024*, 2024, pp. 559–570. DOI: 10.5220/0012851400003756.

[16] F. Lauf *et al.*, "Exploring Design Characteristics of Data Trustees in Healthcare - Taxonomy and Archetypes", *ECIS 2023 Research Papers*, p. 323, 2023.

[17] R. K. Lomotey, S. Kumi, and R. Deters, "Data Trusts as a Service: Providing a platform for multi-party data sharing", *International Journal of Information Management Data Insights*, vol. 2, no. 1, p. 100 075, 2022. DOI: 10.1016/j.jjimei.2022.100075.

[18] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "Medblock: Efficient and Secure Medical Data Sharing Via Blockchain", *Journal of Medical Systems*, vol. 42, pp. 1–11, 2018. DOI: 10.1007/s10916-018-0993-7.

[19] I. Yaqoob, K. Salah, R. Jayaraman, and Y. Al-Hammadi, "Blockchain for healthcare data management: Opportunities, challenges, and future recommendations", *Neural Computing and Applications*, pp. 1–16, 2022. DOI: 10.1007/s00521-020-05519-w.

[20] A. Dagliati, A. Malovini, V. Tibollo, and R. Bellazzi, "Health informatics and EHR to support clinical research in the COVID-19 pandemic: an overview", *Briefings in Bioinformatics*, vol. 22, no. 2, pp. 812–822, 2021. DOI: 10.1093/bib/bbaa418.

[21] M. Parciak *et al.*, "FAIRness through automation: development of an automated medical data integration infrastructure for FAIR health data in a maximum care university hospital", *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 94, 2023. DOI: 10.1186/s12911-023-02195-3.

[22] E. Williams *et al.*, "A Standardized Clinical Data Harmonization Pipeline for Scalable AI Application Deployment (FHIR-DHP): Validation and Usability Study", *JMIR Medical Informatics*, vol. 11, e43847, 2023. DOI: 10.2196/43847.

[23] C. Poschen, B. Herres, and K. Knorr, "A Threat-Driven Design of a Data-Trustee Infrastructure for Medical Data", in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2024, pp. 6753–6760. DOI: 10.1109/BIBM62325.2024.10822788.

[24] S. N. Duda *et al.*, "HL7 FHIR-based tools and initiatives to support clinical research: a scoping review", *Journal of the American Medical Informatics Association*, vol. 29, no. 9, pp. 1642–1653, 2022. DOI: 10.1093/jamia/ocac105.

[25] C. Poschen, B. Berens, and K. Knorr, "Towards Formalized Study Parameters for Medical Research", in press, to be published at MCCSIS e-health 2025, 2025.

[26] B. Berens, J. Grüger, C. Poschen, and K. Knorr, "A FHIR Specification to Formalize Cohort Definitions", in press, to be published at EFMI Special Topic Conference 2025 Good Evaluation - Better Digital Health, 2025.

[27] D. Hallinan, "Broad consent under the GDPR: An optimistic perspective on a bright future", *Life Sciences, Society and Policy*, vol. 16, no. 1, p. 1, 2020. DOI: 10.1186/s40504-019-0096-3.

[28] A. Gkoulalas-Divanis, D. Vatsalan, D. Karapiperis, and M. Kantarcioglu, "Modern Privacy-Preserving Record Linkage Techniques: An Overview", *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4966–4987, 2021. DOI: 10.1109/TIFS.2021.3114026.

[29] B. Herres, C. Poschen, and K. Knorr, "Privacy-Preserving Search on Medical Data", in *Digital Health and Informatics Innovations for Sustainable Health Care Systems*, IOS Press, 2024, pp. 252–256. DOI: 10.3233/SHTI240392.