CHAPTER

Architecture and design choices for federated learning in modern digital healthcare systems

2

Konstantinos A. Koutsopoulos[1], Christoph Thümmler[2], Angelica Avila Castillo[2], Alice Abend[2], Stefan Covaci[3], Benjamin Ertl[3], Giannis Ledakis[4], Stéphane Lorin[5], Vincent Thouvenot[5], Sahar Haddad[6], Gouenou Coatrieux[6], Reda Bellafqira[6] and Alessandro Bassi[7]

[1]*Qualtek Sprl., Brussels, Belgium*

[2]*6G Health Institute GmbH, Markkleeberg, Germany*

[3]*Agentscape AG, Berlin, Germany*

[4]*UBITECH, Chalandri Athens, Greece*

[5]*Thales SIX GTS France SAS, Campus Polytechnique, Palaiseau cedex, France*

[6]*Inserm, Cyber Health, LaTIM UMR1101, IMT Atlantique, Brest, France*

[7]*Eurescom GmbH, Heidelberg, Germany*

2.1 Introduction

The recent advancements in artificial intelligence (AI) and the integration of machine learning (ML) models into day-to-day tasks and applications (Davenport & Kalakota, 2019) have heightened aspirations for innovative healthcare solutions (Habehh & Gohel, 2021). This potential, however, depends on the availability of medical data that is both of good quality and sufficient quantity (Abraham, 2023; Riskin, 2023). However, apart from these two requirements concerning data sources, a horizontal requirement, arising from a legal standpoint related to privacy protection (Wieringa et al., 2021), poses a challenge regarding how data can be involved and utilized in training and inference workflows.

With the emergence of dataspaces of controlled and fully interoperable infrastructure for data sharing and exploitation and the continuously enriched legal and governance framework concerning data and digital services (Shaping Europe's Digital Future, https), data availability and usability are becoming more straightforward. Consequently, this development is creating better awareness regarding both the value and sensitivity of digital assets (Adekoya & Ekpo, 2022). Although the federation of data in the context of dataspaces addresses to a large extent the vertical business ecosystem, where data ownership and provision are both subject to the governance decisions of a single stakeholder (Usländer &

Federated Learning for Digital Healthcare Systems. DOI: https://doi.org/10.1016/B978-0-443-13897-3.00002-3

Teuscher, 2022), much more complexity is involved when it comes to medical data. The reason is similar to all cases where the interests of individuals are involved. The data subjects who play the primary role in governance decisions do not have a direct connection with the health domains responsible for storing the data on their behalf (Berlage et al., 2022). As the storage, exposure, and usage of data are strictly related to the well-defined consent decisions made by the data subjects, the parties involved need solutions that may require revisiting current practices to allow them to effectively participate in the various dataspace processes. Such solutions should be subject to trust establishment and continuous attestation, expressed in the form of immutable proofs, ensuring adequate and foreseen system and processing integrity (Koutsopoulos, 2022).

The motivation, challenges, and aspirations outlined above have significantly influenced the objectives of the European project PAROMA-MED. The project aims to develop, validate, and evaluate a hybrid-cloud (central and edge) delivery framework that ensures privacy and security for services and applications in federative cross-border environments. This is achieved by providing technologies, tools, and services to support various aspects, including automatic attestation of federation partners; privacy and security by design; continuous risk assessment; privacy-preservation; and trusted data storage and processing in federative environments; AI/ML by design, managed privacy and security operations for automated policy enforcement; and cyberthreat detection and mitigation. The concepts of PAROMA-MED are discussed in this chapter.

2.1.1 Key contributions

The chapter focuses on the following:

i. Current dataspace landscape and evolution of new practices and patterns that relate to the value of data.
ii. Potential of the availability of medical data for the development of ground-breaking AI-based solutions, as well as the privacy concerns and restrictions
iii. Concepts of candidate solution, currently evolving in PAROMA-MED, embracing privacy protection at its core with utilization potential beyond the healthcare domain.

2.1.2 Chapter organization

Section 2.1.2 presents the current landscape of dataspaces and particularly how it may introduce new practices in the healthcare domain due to the potential of medical data for the evolution of AI-based solutions. Section 1.3 presents the details of a privacy-aware and privacy-preserving technical approach that is aligned with the current needs for dataspace establishment in the healthcare domain. This section considers FAIR (findable, accessible, interoperable, reusable) principles as key enabling features for data exploitation and also analyzes how data sovereignty can be assured, and most importantly, the concrete trust establishment mechanisms that will raise user confidence.

2.2 Dataspaces and health domain

2.2.1 State-of-the-art and current practices

The evolution of dataspaces across Europe has been dictated, among others, by the strategic objective of the European Union (EU Strategy, 2022) to become a world leader in digital and data economy. With interoperability being a key aspect (EU Data Act, 2022) that all initiatives in the field are trying to ensure and to avoid fragmentation due to different interpretations and implementations, the Horizon 2020 OpenDEI (OpenDEI, 2021) project brought together experts from several initiatives and organizations to define a set of common design principles and standards, including both technical and governance aspects. One of the key takeaways of this effort has been the definition of a soft infrastructure that identifies the main sector-agnostic building blocks that in turn identify how the participants have to interact either within sector-specific (the ones defined in Common EU Dataspaces and future ones [EU Data Act, 2022]) dataspaces or in intersector scenarios. This soft architecture organizes the building blocks according to the four main categories, three of which relate to technology (interoperability, trust, data value) and one to business and regulation (governance).

Each of the categories identifies a number of important building blocks facilitating the purpose of the category, whereas additional ones can be optionally deployed to aid interoperability and connectivity with additional systems with the data connector architecture by IDSA and the Trust Framework semantics and procedures by Gaia-X identified in this study as the most important (Siska et al., 2023).

Data connectors provide the basic mechanism for enabling a participant to connect and operate within the context of a dataspace, ensuring the support of exchange services and policy enforcement. This, in turn, facilitates technical interoperability. According to the IDSA Reference Architecture Model (IDS RAM, 2023), a connector is deployed, either on the cloud or on local resources, as a set of containers under the command of an application container management functionality. Among the containers, those identified as core containers take control of tasks related to data exchange, metadata management, remote attestation, logging and monitoring, and policy and contract management.

Gaia-X Trust Framework (GAIA-X Trust, 2022) is based on the exchange of verifiable credentials managed and utilized by functional components at any phase of interaction among dataspace participants. The fundamental element of the Gaia-X Trust Framework regards the verification of the validity of the claims stated in the self-descriptions of the participants, including self descriptions and claims of service and resource offerings. To achieve this purpose the utilization of trust anchors and compliance services is predicted. Presently the trust framework deals with participants identified as legal persons and, on an experimental basis, natural persons. It is expected that the roles of provider, consumer, and federator, as defined by the architecture document, will soon be supported by the trust framework.

2.2.2 Impact on machine learning

Dataspaces have a positive impact on ML and AI as they facilitate data access and allow better coordination and interoperability between the participants (EU, 2023). All AI technologies benefit from these new facilitating concepts and solutions, leading to the development of important applications in the healthcare domain, specifically in areas such as diagnosis, treatment planning, and therapy guidance.

In addition to the advantages of ML/AI technologies, the introduction of dataspaces not only transforms the paradigm of data storage and access but also has an impact on ML models. It also brings forth new requirements for the privacy and security of ML, impacting the field (Kerry, 2020).

In the subsequent sections of this chapter we will focus on AI algorithms that aim to protect data and AI processes. These algorithms are designed to ensure protection even in the event of attacks and/or leaks or to minimize their impact. The privacy and security ML techniques outlined here are designed to prevent unnecessary data sharing and exchanges, introduce noise to the data when referencing is necessary, and incorporate traceability measures for all information used and generated by AI models. A more detailed examination of these techniques will be provided in a later paragraph that outlines the proposed approach.

2.2.3 European digital age and development of secure (health) dataspaces

The European Commission is trying to make Europe fit for the digital age. It is determined to promote in Europe the so-called digital decade whose goal is to strengthen digital sovereignty by setting a new set of standards with a focus on data, technology, and infrastructure (EU Digital Decade, 2019).

The Digital Decade policy program contains targets, objectives, and ambitions for 2030 and will guide Europe’s digital transformation. The commission will pursue its digital ambitions through concrete terms such as projected trajectories at the EU and national level, an annual cooperation cycle to monitor and report on progress, and through multicountry projects that combine investments from EU member states and the private sector.

On January 26, 2022, the commission proposed an interinstitutional solemn declaration on digital rights and principles in the digital decade. These new rights include, e.g., prioritizing individuals and their rights in the digital transformation, supporting solidarity and inclusion, ensuring freedom of choice online, promoting participation in the digital public space, enhancing safety and security, and consequently empowering individuals (EU Digital Decade, 2019). These rights and principles will complement the already existing rights reported in the Charter of Fundamental Rights of the European Union, as well as data protection and privacy legislation (GDPR) (Charter of Fundamental Rights of the European Union, 2000; Radley-Gardner et al., 2016).

The European Commission made the proposals in December 2020, and on March 25, 2022, a political agreement was reached on the Digital Markets Act (Digital Markets Act, 2022), and on April 23, 2022, on the Digital Services Act. Together they form a single set of new rules governing digital services in the EU that will be applicable across the whole of the EU. The main goals of the DSA and DMA are to create a safer digital space in which the fundamental rights of all users of digital services will be guaranteed and protected and to establish a level playing field to foster innovation, growth, and competitiveness, both in the European Single Market and globally (Digital Services Act, 2020).

The health domain is composed of four essential contributors: data holders and users, application and service providers, data space governance and operating systems, and cloud service providers (Gaia-X Domain Health Position Paper Version 1.0, 2021). Dataspace is the term that primarily refers to any ecosystem of datasets and data models, including ontologies, data-sharing contracts, and data management services, as well as associated soft competencies such as social interactions, governance, business processes, etc. Such competencies follow a data engineering approach whose goal is to optimize data storage and exchange mechanisms, preserving, generating, and making possible knowledge sharing to others. In contrast, data platforms refer to architectures and repositories of a group of interoperable hardware/software components that follow a software engineering approach (Scerri et al., 2022). The two concepts, data engineering and platforms, are interconnected and need to be considered together, as commercial solutions often do not differentiate between them. Therefore, and due to the special requirements for protecting the privacy of the individual, a distinction was made between technology and infrastructure that stores and processes personal and other data.

The nine European data-sharing spaces outlined by the European Strategy for Data are health, industry, agriculture, finance, mobility, green deals, energy, public administration, and skills. They are essential for the implementation of the European digital market and guide European activities toward the data economy (Scerri et al., 2022).

The BDVA (Big Data Value Association) is a community of experts that has been working on the development of dataspaces for many years. Their vision comprises a data space composed of several individual connected spaces. The dataspace should be able to cut across sectoral, organizational, and geographical boundaries (European Big Data Value Association, 2015).

As previously described, the European strategy for data aims to create a single market for data, which should ensure Europe’s global competitiveness and data sovereignty. The strategy essentially aims to ensure the flow of data within and across EU sectors, making high-quality data available for innovation in the economy and society (while keeping data owners, companies, or individuals in control). The strategy is based on ensuring full compliance with European rules, regulations, and values as well as setting rules for fair access and usage of data in accordance with the existing data governance mechanisms (EU Digital Decade, 2019). Common European Dataspaces will be central in enabling new technologies such as AI and supporting the marketplace for cloud and edge-based services.

2.2.4 Potential

In order to release the full potential of health data, the European Commission is presenting a regulation to set up the European Health Dataspace (EHS). The EHS is a specific ecosystem composed of rules, standards, infrastructure, and a legal governance framework that aims to empower individuals through increased digital access to and control of their electronic personal health data. It supports the use and free movement of health data across the EU for better healthcare delivery, better research, innovation, and policymaking. It enables the EU to use and reuse the full potential of health data offered through a safe and secure exchange. The EHS supports the fostering of a genuine single market for electronic health record systems, medical devices, new technologies, and high-risk AI systems (EU Health Data Space, 2022). Thus, it is a core component of the European Health Union and builds further on the GDPR, the proposed data Governance Act, the draft Data Act, as well as the Network and Information Systems Directives.

Dataspace initiatives aim to access and share highly sensitive personal data in a secure and confidential manner governed and controlled by each EU member state in a consistent way, complying with relevant European regulations. A EHS could facilitate future pandemic management, including fast data transfer and short reaction times. Pattern recognition of disease outbreaks across state borders would be possible. Moreover, national healthcare systems could be relieved of the burden of some bureaucratic processes. For example, doctor appointments necessary to simply transfer data from one medical office to the other by the patient would become obsolete. Doctors would also be able to treat patients at home, which again reduces the pressure on hospital bed occupancy.

In the case of regulators and policy-makers, they will have easier access to health data and be able to make decisions for the better functioning of healthcare systems, leading to a more evidence-based policy-making. This will lead to better access to healthcare, increase its efficiency, reduce costs, and enable new research and innovation.

Also, the industry can benefit from the better availability of electronic health data sourced in an EU-wide market. This will improve people's health by facilitating the production of medical devices and gadgets, leading to improved personalized healthcare coverage.

2.2.5 Challenges

Secure health dataspaces require a complex ecosystem involving many different stakeholders connected by a plethora of regulated processes. The set-up and operation of such dataspaces are associated with many challenges. Various sources of public funding involved in the design of health dataspaces need to be reconciled. EU member states can apply different financing models in accordance with their sovereign legislation. Moreover, highly fragmented and heterogeneous EU markets limit the quick rollout of dataspaces or any other kind of digital framework on a large and transnational scale. Stakeholders may struggle to interpret and map the GDPR rules with the local legislators of the member states. Navigating the complex regulatory landscape and ensuring compliance with data protection laws adds another layer of complexity to the development process.

Data interoperability will be a major challenge for the design of health dataspaces. Medical data is often stored in various formats, collected by different organizations across miscellaneous systems. Data interoperability is a prerequisite for unified dataspaces. Otherwise, practitioners will have difficulty accessing, modifying, and exchanging data. Advances in standardization and the development of data exchange protocols are needed to achieve sufficient data interoperability. A single European or international health data standard such as Fast Healthcare Interoperability Resources (FHIR) should be adopted on a European level, as proposed by a study prepared for the European Parliament's committee on industry, research, and energy (Marcus et al., 2022). FHIR is a healthcare data standard with an application programming interface used to represent and exchange electronic health records (FHIR & Cloud Healthcare API', 2023). The standard enables links between medical data across different systems.

Additionally, special attention needs to be paid to security threats and data breaches. Sensitive health data should be protected from unauthorized access and cyberattacks with security measures such as encryption, intrusion detection systems, and the application of strong access control. Data quality conservation and integrity go hand in hand with data security concerns. Medical data needs to be accurate, reliable, free from errors, and protected from unauthorized tampering. Ensuring long-term data quality is a challenge, especially if data stems from various practitioners across different systems. This, in turn, leads to the establishment of data standards and exchange formats. However, the establishment of data silos is discouraged, even though they might simplify data protection measures. On the other hand, these silos could bring about data stockpiling without any useful function or connectivity. Current solutions are mostly too permissive (e.g., exposing data to the public domain or transferring data usage rights to a single commercial company) or too restrictive (e.g., study-specific point solutions or local-for-local solutions without opportunities for reuse). A balanced middle ground between these two extremes should be found in the development of the modern health dataspaces.

From a global perspective, without a more open European market, innovative companies are forced to focus their strategies on the United States and China.

The European Union and its associated transnational legislation could offer a great opportunity to attract international companies if they were not limited by the national borders and limitations of the member states. It remains a complex task to deploy a health data management system throughout the EU. But it is nevertheless a vital step toward medical research and development in Europe.

2.3 Proposed approach

EHDS will create a common space where natural persons can easily control their electronic health data as far as the fundamental rights of the data subject are concerned. Thus the individual can control to which entities (including humans and services) their data can be made available, as well as the constraints enforced. It will also enable researchers, innovators, and policymakers to use electronic health data on the condition that they fulfill the eligibility criteria for a number of added-value cases and workflows, in a trusted and secure way. The added value cases may range from medical inspection to AI training and evidence traceability.

In accordance with the goals of EHDS, PAROMA-MED aims to resolve the challenges mentioned above and proposes an approach that establishes a data life cycle, empowering individuals in the governance of their data. It allows the sharing of nonidentifiable health data and facilitates the trusted execution of data for researchers and other health professionals.

2.3.1 Data life cycle

The governance of data in PAROMA-MED involves four main phases: (1) secure and trusted addition of health data to the local data lake during data generation; (2) preparation of data for exposure in the federation, including encryption, anonymization, and ensuring data interoperability; (3) allowing the search over nonidentifiable health data available in the data space; and (4) using the data in a secure and trusted way, which may include additional consent from the user. The introduced data life-cycle management complies with any stakeholder's archiving procedures and policies as it retains adequate structuring, indexing, and retrieval. Beyond that, it allows data subject consent (including the right to be forgotten) to be appropriately applied. Furthermore, legacy data is planned to be integrated through the deployment of appropriate adapters with privacy and ownership protection and enforcement mechanisms supported by design. PAROMA-MED plans to provide user support dashboards for managing legacy data inclusion tasks.

2.3.2 Data generation and interoperability

Ideally, data should be generated in direct association with the subject they belong to. Assuming that data is generated following some medical procedures that are performed within the relevant medical infrastructure, the data subject has to be identified in the context of the validation of their prescribed examinations (Fig. 2.1). Upon presentation of the prescription, a trusted medical domain (through components that are continuously attested for integrity and adherence to the foreseen procedures—depicted in the figure in the green box) requires an identity challenge to be sent to the data subject. This step involves verifying both the subject's identity (potentially supported by a digital wallet application) and confirming that the medical domain is verified for its adherence to the proper procedures (Step 1). The resolution of the challenge establishes a time-limited association between the subject and the process (Step 2), concluding with the secure storage of the results under clear governance for future use (Step 3). Adhering to FAIR practices, an FHIR server (as shown in the figure within the secured storage

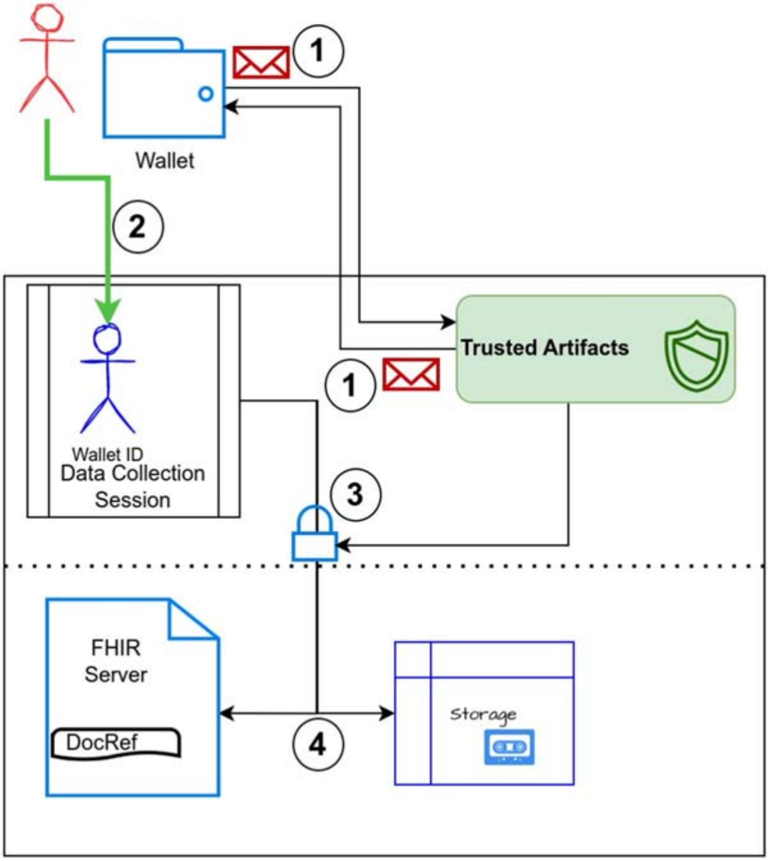

FIGURE 2.1

Data collection steps.

area at the bottom) is used from the moment data is generated. It is protected under the supervision and encryption by trusted domain artifacts of edge nodes for maintaining personal information. Additionally, an object storage solution, also under the same protection measures, is used for storing various medical results. Object references are included within FHIR documents. In the figure below (Fig. 2.1), the data subject is represented through a wallet-based interaction. This aspect has been taken into consideration for further research to ensure compliance with EBSI and eIDAS solutions.

2.3.3 Data exposure

Before the data can be used in the context of specific actions, they have to be discoverable, as far as federation interactions are concerned. For this purpose, a data inventory layer is produced from the data types available inside the protected storage. Inventory updates are performed in batch mode to avoid statistical variation being linked to identities. The process of populating the content of the data inventory layer takes into account constraints from policies specified by individual consents. The outcome is intended to be published for discovery in a dataspace ecosystem through the appropriate connector (illustrated at the border of the trusted domain). The flow is presented in the figure below (Fig. 2.2).

2.3.4 Data discovery

Once the data sources are exposed in the federation they can be utilized in AI model design and training workloads. Exposure does not mean direct population of some external storage system but the availability and participation in dataspace

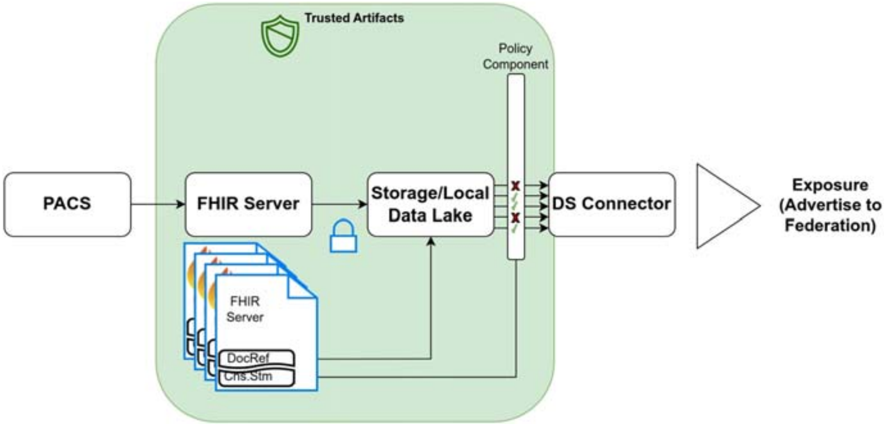

FIGURE 2.2

Dataspace exposure.

interworking sessions for resolutions of queries. Data scientists are able to submit queries to the Dataspace Metadata Brokering subsystem for discovering availability of types and volumes of data according to certain protection levels. The outcome is collected from all the participating domains. Each of the domains contributes to the query resolution by applying internally user policies and resolving the portions of the stored data that can be made available according to the query options.

This process aims at a streamlined and ergonomic approach that relieves the data scientists from the burden of locating data that are highly distributed, but most importantly from the burden of taking all measures to remain with the legal restrictions that private data protection legislations require. This leads to a one stop shop service and enabling mechanism. At this stage data availability is presented under three main categories (assuming local processing in all cases):

- Directly usable data
- Data of application relevance that need additional consent
- Data without known relevance and quantity

If the first category suffices, the ML flow can continue. In the opposite case the consuming side (data scientists on behalf of any organization or by themselves) can suggest rewards for the other two categories in an effort to secure data availability adequate enough to allow proper development of the intended ML model. In such cases the subject is presented with an incoming request containing usage context details to facilitate the creation of a clear decision in the form of an enforceable policy. Subsequently the consent details are updated, and additional usage possibilities are permitted.

2.3.5 Data usage

In the PAROMA-MED approach one of the core concepts is to avoid the transfer of medical data. Instead, with appropriate consent and trust prerequisites, the ability to perform secure computation on health data stored in a node of PAROMA-MED should be possible. As the usage of data is constrained by the intentions and identity of the consumer, based on the options of the producer or the subject whose privacy is to be protected, there is a need to securely enclose the entire flow within the strict borders of an instantiated environment, both in terms of deployed functionality and data with a limited lifespan. The approach is based on the Gaia-X conceptual and composition model, which envisions that resources can be:

- Virtual Resource: It represents static data in any form and necessary information such as dataset, configuration file, license, keypair, an AI model, neural network weights, . . .

- Instantiated Virtual Resource: it represents an instance of a virtual resource. It is equivalent to a service instance and is characterized by endpoints and access rights.

According to the envisaged approach, data at rest (stored in local data lake and FHIR server) serves as virtual resources that can be instantiated within a volatile and isolated software enclosure. This enclosure facilitates the application of the intended processing, forming the depicted data usage layer. This step requires that the data is exposed in a uniform manner irrespective of its actual storage format. Additionally, if this step requires any filtering, encryption, anonymization, or watermarking, it is performed during the provisioning phase for the preparation of the data usage layer (Fig. 2.3).

Once adequate data is available for the model training purposes, data is prepared and remains available for the foreseen processing (constrained in terms of usage and time limits) (Fig. 2.4). The preparation phase, as explained earlier, involves adaptation and encryption/crypto-watermarking according to data owner policies and data user requirements.

The negotiation between the designed application and federation resources is not limited to data discovery and usage. It also encompasses the availability of processing resources, which is also subject to discovery and, in several cases, closely related to the resolution of data availability in cases where data cannot be

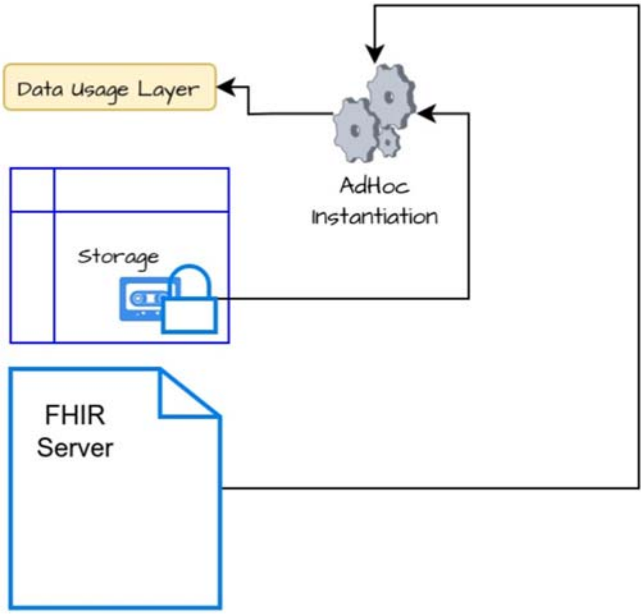

FIGURE 2.3

Data usage layer: provisioning.

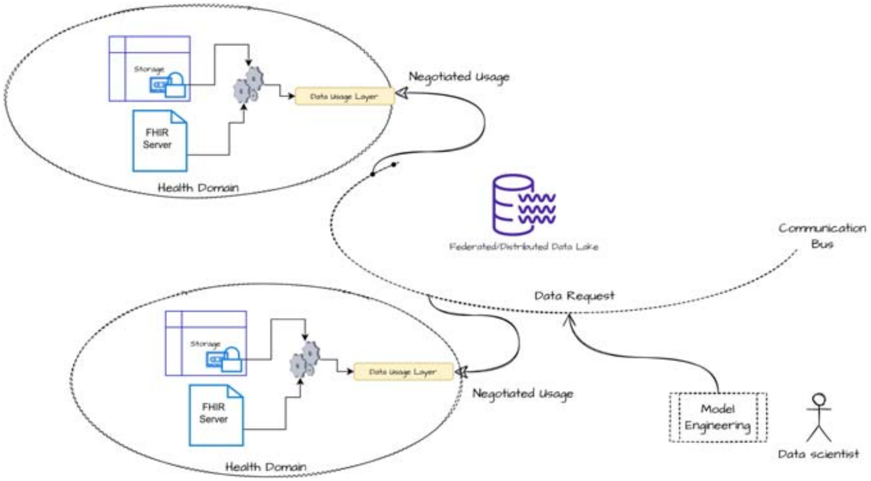

FIGURE 2.4

Model training preparation: data usage negotiation.

transferred outside of the domain borders. Such processes lead to the provisioning of data processing modules that actively participate in the training process. Moreover, processing modules can be deployed on the central cloud for resource-intensive tasks (Fig. 2.5).

The envisaged data usage approach resolves any issues related to the prevention of the use of AI/ML at large that stem from data encryption and protection. This is achieved by bringing training close to the data without necessitating disclosure and transfer among storage systems.

2.4 Dataspaces and participation in ecosystems

2.4.1 Identity governance

The PAROMA-MED approach for identity governance plays a crucial role in ensuring secure and trustworthy data sharing and collaboration. Identity governance within the PAROMA-MED framework focuses on establishing a robust and reliable mechanism for managing identities, access controls, and privacy considerations in federated learning (FL) scenarios.

PAROMA-MED emphasizes the need for a centralized identity management system that governs the identities of all participants involved in the FL ecosystem. This includes healthcare providers, researchers, data custodians, and other stakeholders. The identity management component ensures that each entity is uniquely identified and authenticated prior to their involvement in any data sharing or analysis activities.

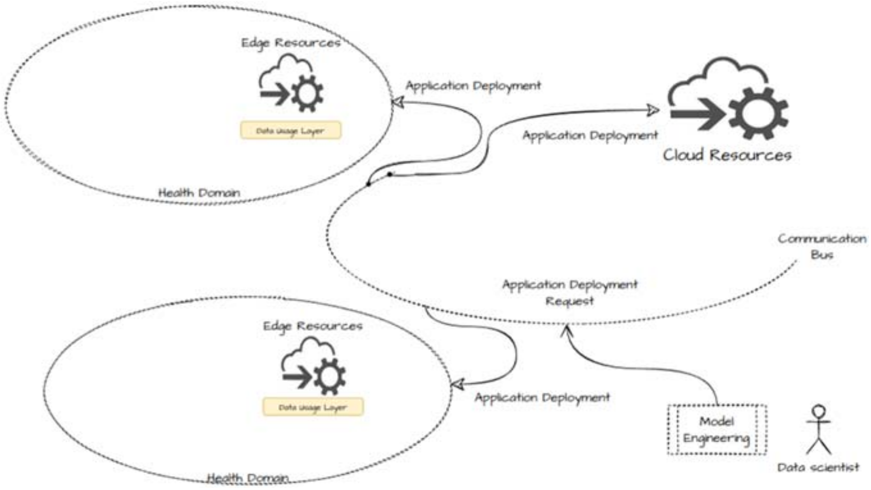

FIGURE 2.5

Model training preparation: deployment of compute modules.

By incorporating RBAC principles to define and enforce access controls within FL systems, different entities can be assigned roles based on their responsibilities and privileges. This approach ensures that only authorized individuals or entities can access specific datasets, participate in collaborative analysis, or contribute to the FL process.

In addition, PAROMA-MED leverages privacy-preserving technologies, such as differential privacy, FL, and secure multiparty computation, to protect sensitive patient data during the collaborative analysis process. These technologies help in minimizing the risk of data breaches and maintaining patient privacy while enabling effective knowledge sharing and model development.

Clear data governance policies that outline the permissible use, access, and sharing of healthcare data and consent management mechanisms ensure that patients have control over their data and can provide informed consent for its use in FL research.

2.4.2 Consent management

The focus on maintaining privacy and control over patient data while enabling its usage for research and learning purposes is one of the key aspects of the consent management approach of PAROMA-MED. Consent management is designed for the seamless integration of data into the architecture and design choices of the system.

The consent management process begins with capturing and recording patient consent for data usage. Various mechanisms, such as consent forms and digital consent processes, are employed to ensure that individuals can make informed decisions about how their data will be used. The aim is to provide transparency and clarity regarding the purpose and scope of data usage.

Consent storage and access control mechanisms are implemented to securely store the consent choices made by individuals and associate them with their data. This ensures that only authorized entities can access the data based on the provided consent, safeguarding patient privacy and ensuring compliance with consent preferences.

To respect the dynamic nature of consent, the PAROMA-MED approach allows individuals to easily revoke or modify their consent preferences. They are provided with interfaces and tools to manage their consent settings, empowering them to have control over the usage of their data and exercise their right to withdraw consent if desired. Moreover, maintaining an audit trail, ensures accountability, transparency, and compliance with privacy regulations. These auditing capabilities include recording the details of consent, such as when it was obtained, the specific terms of consent, any modifications or revocations, and the actions taken based on the provided consent.

In the context of FL, PAROMA-MED's consent management approach is seamlessly integrated into the process. Only data for which explicit consent has been given is included in the FL models. Privacy-preserving techniques, such as data anonymization or encryption, are applied during the learning process to further protect patient privacy.

By incorporating consent management into the architecture and design choices, the principles of responsible and ethical data usage are upheld. The privacy preferences of individuals are respected, and transparency and control over data usage are provided, fostering trust between patients, healthcare providers, and researchers.

2.4.3 Trust establishment

According to the European Data Strategy fact sheet, data processing moves gradually from centralized cloud computing facilities to smart, connected, and edge resources. This pattern, in combination with dataspaces, leads to the formation of data federations, where business domains attach to ecosystems to perform data-related tasks according to specific roles. According to the Gaia-X conceptual model, participants interact within the federation by providing (producers) and consuming (consumers) resources. The basic assumption before engaging in a transaction is that both interacting parties present verifiable credentials to each other, issued by participants and signed by trust anchors. Aiming to enable the automated attestation of federation participants, following the ideas of Gaia-X and the Compliance Service deployment options (licensed, private decentralized, secure private, and public decentralized models) closely, the consolidation of the trust model, grounded in both the secure private and the public decentralized models, can be expressed as follows:

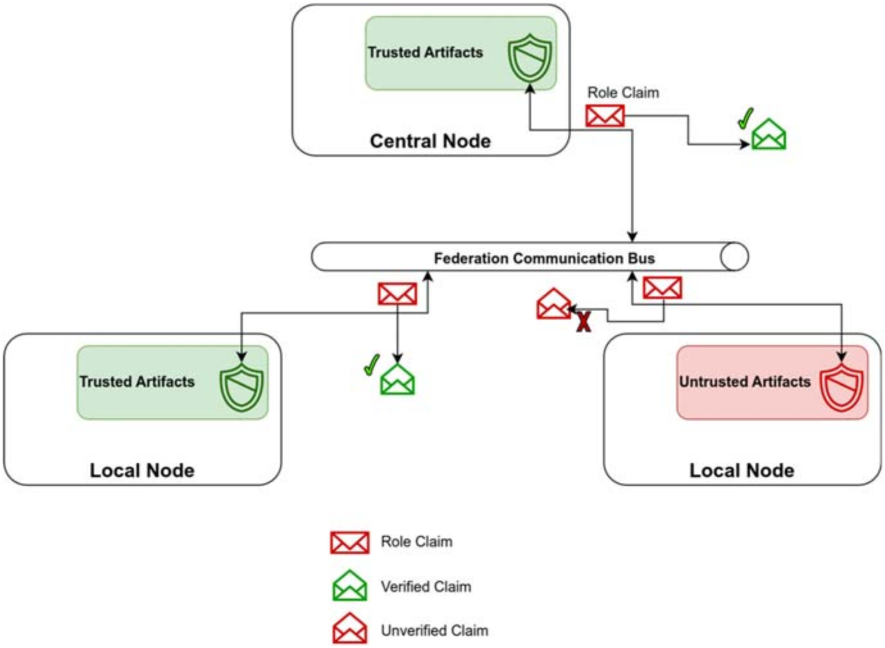

FIGURE 2.6

Trust establishment based on integrity of the participant.

- Elimination of manual process in trust establishment and legal binding (e.g., contracts and/or SLAs)
- Trust is established on proofs related to the integrity of certified components and their exclusive involvement in privacy-sensitive operations, based on trusted execution enclaves and/or trusted platform modules
- Adherence to proper operation is a continuous process, and any verification failure leads to immediate and immutable publication of the status change

This approach is visualized in the following figure (Fig. 2.6). The concept of the Federation Communication Bus serves as a placeholder to be further elaborated and clarified according to the dataspace connector protocols.

2.4.4 Protection and assurance

2.4.4.1 Privacy-preserving technologies

2.4.4.1.1. Agent-based approach

The agent-based approach, specifically using authentication and authorization sidecar proxies along with policy enforcement agents, is an innovative concept in the context of architecture and design choices for FL in modern digital healthcare systems. This approach leverages the benefits of microservice architecture to enhance security, privacy, and consent management.

In this concept, microservices within the FL system are equipped with sidecar proxies responsible for authentication and authorization. These proxies act as intermediaries between the microservices and external systems, handling the authentication process and enforcing access control policies. They authenticate users or entities requesting access to the system and verify their credentials against trusted identity providers. By offloading authentication tasks to dedicated proxies, the microservices can focus on their core functionalities, ensuring a modular and scalable architecture.

Policy enforcement agents are introduced to enforce security, privacy, and consent policies within the system. These agents are responsible for evaluating and enforcing policies related to data access, data sharing, privacy protection, and consent management. They operate in conjunction with the sidecar proxies to enforce fine-grained policies based on user roles, permissions, and other contextual attributes. This enables dynamic and context-aware policy enforcement, ensuring that sensitive data is accessed and shared appropriately and that privacy and consent requirements are adhered to.

By incorporating authentication and authorization sidecar proxies and policy enforcement agents, the agent-based approach enhances the security, privacy, and consent management capabilities of the FL system. It enables centralized policy management and enforcement, ensuring the consistent application of security and privacy controls across microservices. The use of dedicated agents allows for flexibility and agility in adapting to changing policies and regulatory requirements.

Furthermore, this approach promotes interoperability and compatibility with existing authentication and authorization frameworks, enabling seamless integration with external identity providers and policy management systems. It provides a unified and standardized approach to authentication, authorization, and policy enforcement across the FL system, facilitating secure and privacy-preserving data exchange and collaboration. The approach strengthens security, privacy, and consent management capabilities, ensuring the protection of sensitive data and compliance with regulatory requirements while maintaining the flexibility and scalability offered by microservices.

2.4.4.1.2 Machine learning-based approach

FL is the most commonly used ML approach that allows multiple data owners to collaboratively train an ML model without sharing their own training data. Some other approaches are possible, such as model fusion. Here, all participants train a local ML model. When the local trainings have converged, the model weights are smartly aggregated to obtain a model that generalizes well on the data of all participants. Teacher aggregation ensemble is another potential approach. Here, each local data owner trains a local ML model, which is called the teacher model. The teacher models are employed to label a new dataset that is used to train a student ML model that is deployed.

While FL is flexible and resolves data governance and ownership issues, it does not itself guarantee security and privacy unless combined with other methods. Indeed, when using an ML model, information can leak on the learning data, even if the ML objective is to generalize as much information as possible. Many recent works have shown that ML models themselves can be used to derive personal information. In particular, two kinds of attacks have been described:

1. Membership inference attacks: The ability to identify whether a data record was included in the training dataset of the target ML model
2. Attribute inference attacks: The ability to infer missing attributes of a partially known record used in the training dataset by accessing the ML model.

These new needs in terms of security and privacy encourage the use of approaches such as secure multiparty computation (SMPC) or differential privacy (DP) to secure FL processes.

2.4.4.2 Secure multiparty computation

Working on encrypted data is one of the best ways to guarantee security. However, enabling efficient processing of such encrypted data is one of the biggest challenges in the security field. Although fully homomorphic encryption allows one to perform calculations over encrypted data without decrypting it first (Gentry, 2009), it is often judged too slow, complex to use, and impractical. SMPC (Lindell, 2020) is an alternative to homomorphic encryption. It allows owners of private datasets to perform operations on their collective data without disclosing anything except the outcome of the computation. As an example of the SMPC method, private set intersection (PSI) has garnered much attention due to its capability to facilitate efficient comparisons and certain analytics on encrypted data sets. PSI could be used, for example, to determine shared patients between two hospitals without disclosing the specific patient lists held by hospitals.

2.4.4.3 Differential privacy

Intuitively, differential privacy (Dwork, 2014) corresponds to ensuring that the output distribution of a randomized algorithm will not be significantly different considering the presence or absence of one particular individual. An adversary with access to the algorithm will not be able to learn about individuals but will only have access to the global knowledge of the algorithm among them, ensuring the protection of privacy.

2.4.4.4 Data protection and traceability

Data traceability is another major concern for FL, as sensitive data has to be shared between different users. In some cases, data samples may be remotely requested to understand incorrect model behavior, or externalized for annotation when the expertise for annotation is not on site. The risk of information leakage is not negligible. Over 50% of data breaches originate internally. There is a need to be able to hold accountable the entity responsible for the leak and identify it as quickly as possible. Today, this requires a complex and lengthy investigation, generally lasting more than 2 months. Similar issues arise when it comes to models' parameters. Building a model is costly, as it requires expertise in data science and medicine as well as huge computing resources. Herein it is important to protect model ownership.

Data watermarking and model watermarking are technologies that can address such threats in the FL environment. When applied to images, watermarking is defined as the invisible embedding of a message into a host image by imperceptibly modifying its gray values. Watermarking leaves access to the data while keeping it protected by the message (Boenisch, 2021). Depending on its content, the embedded message can fulfill various security services, such as ensuring data authenticity, maintaining data integrity, and enabling data traceability. This may involve embedding proof of ownership or a message tracker to counteract information leaks and identify malicious users. There has been a growing interest in combining watermarking with encryption to achieve both a priori and a posteriori protection simultaneously (Haddad et al., 2021). The integration, known as the crypto-watermarking technique, is designed to provide watermarking-based security services from encrypted data.

2.5 Lessons learned: conclusions and future scope

PAROMA-MED has worked so far on an extended set of functional and nonfunctional requirements that are trying to cover several perspectives from different stakeholders' (data subjects, data scientists, medical experts, medical centers and organizations, application providers, etc.) point of view. The process revealed several aspects regarding interoperability, feasibility, value protection, and adequacy of the technology. More specifically, the role of FHIR has been identified as the most prominent solution with respect to data structuring, management, and storage. Furthermore, involvement in dataspace practices appears to present a clear pathway toward maximizing the utilization of project outcomes, including proposing/contributing to a concrete model that explores the feasibility of close-to-data processing. Furthermore, a clear challenge for utilizing trusted computing practices for the purpose of zero trust and attestation to deliver the trusted components has been made evident. Finally, the interaction with external players, such as medical experts, has significantly clarified the importance of data value protection not only for primary medical data but also for secondary data products that demand domain knowledge.

After fulfilling the identified requirements, the project is currently advancing in the development of key components, including medical imaging adaptation, FHIR server, object storage, FL framework, dataspace connectors, watermarking solutions, and identity and privacy awareness. At the same time technical workshops are being conducted, with the gradual expansion of integration and experimentation scenarios focusing on the overall flow of operations, including data ingestion, data consent management, data advertisement, data discovery, usage negotiation, and ML training.

The current study introduces the main concepts evolving within the context of the European Project PAROMA-MED. With the challenges and potential of dataspaces and federated ML well-identified, the project is soon to enter an experimentation phase. This phase will lead to the realization of the identified concepts to be evaluated in a concrete use case related to the qualitative assessment of cardiac anatomy. Specifically, the project is based on addressing the characterization of myocardial wall thinning using cardiac computed tomography images.

Acknowledgment

This work is funded by the European Union under Grant Agreement 101070222. The views and opinions expressed in this work are solely those of the author(s) and do not necessarily represent the views of the European Union or the European Commission (granting authority). The European Union and the granting authority cannot be held responsible for this work.

References

Abraham, I. Jr., (2023). Importance of data quality in artificial intelligence for healthcare. Accessed June 11, 2023, from https://www.linkedin.com/pulse/importance-data-quality-artificial-intelligence-abraham-ibrahim-jr-/.

Adekoya, O., & Ekpo, E. (2022). Digital assets – An emerging trend in capital markets. Available: https://www.pwc.com/ng/en/assets/pdf/digital-assets.pdf

Berlage, T., Claussen, C., Geisler, S., Velasco, C. A., & Decker, S. (2022). Chapter 18 Medical data spaces in healthcare data ecosystems. *Designing Data Spaces*. Available from https://doi.org/10.1007/978-3-030-93975-5_18.

Boenisch, F. (2021). A systematic review on model watermarking for neural networks. *Frontiers in Big Data*, *4*729663.

Charter of Fundamental Rights of the European Union', no. C 364/3, Dec. 2000, [Online]. Available: https://www.europarl.europa.eu/charter/pdf/text_en.pdf

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. Available from https://doi.org/10.7861/futurehosp.6-2-94.

Dwork, C. (2014). *The algorithmic foundations of differential privacy*.

EU Data Act, EU COMMISSION, On Common European Data Spaces, 23.02.2022.

EU Strategy (2022). *A European strategy for data*. https://digital-strategy.ec.europa.eu/en/policies/strategy-data (June 25, 2023).

EU (May 05, 2023). A European approach to Artificial Intelligence and the role of open data. Accessed May 15, 2023 from https://data.europa.eu/en/news-events/news/european-approach-artificial-intelligence-and-role-open-data.

European Big Data Value Association' (Apr. 08, 2015). BDVA. Accessed May 23, 2023 from https://www.bdva.eu/about.

Europe's Digital Decade: digital targets for 2030'. Accessed May 23, 2023 from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en.

FHIR | Cloud Healthcare API'. Google Cloud. Accessed Jun. 02, 2023 from https://cloud.google.com/healthcare-api/docs/concepts/fhir.

Gaia-X Domain Health Position Paper Version 1.0 2021'. Accessed: May 24, 2023. [Online]. Available: https://www.bmwk.de/Redaktion/DE/Publikationen/Digitale-Welt/211116-pp-health.pdf?__blob = publicationFile&v = 1

Gaia-X Trust Framework - 22.10 Release.

Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. *STOC*, *9*, 169–178.

Habehh, H., & Gohel, S. (2021). Machine learning in healthcare. *Current Genomics*, *22*(4), 291–300. Available from https://doi.org/10.2174/1389202922666210705124359.

Haddad, S., Coatrieux, G., Moreau-Gaudry, A., & Cozic, M. (2021). Joint watermarking-encryption-JPEG-LS for medical image reliability control in encrypted and compressed domains. *IEEE Transactions on Information Forensics and Security*, *15*, 2556–2569.

IDS RAM (2023). Accessed May 31, 2023 from https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/.

Kerry, C.F. (2020). Protecting privacy in an AI-driven world. Accessed June 15, 2023 from https://www.brookings.edu/articles/protecting-privacy-in-an-ai-driven-world/.

Koutsopoulos, K., et al. (2022). Federated machine learning through edge ready architectures with privacy preservation as a service. In: *2022 IEEE Future Networks World Forum (FNWF), Montreal, QC, Canada* (pp. 347–350), doi: 10.1109/FNWF55208.2022.00067.

Lindell, Y. (2020). Secure multi-party computation[Online]. Accessed 25 June 2020 from, Available at: https://eprint.iacr.org/2020/300.pdf.

Marcus, J. S., Martens, B., Carugati, C., Bucher, A., & Godlovitch, I. (2022). The European health data space. *SSRN Journal*. Available from https://doi.org/10.2139/ssrn.4300393.

OpenDEI (April 2021). Design Principles for Data Spaces, Position Paper.

Proposal for a Regulation of the European parliament and of the council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC'. Accessed: May 23, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri = CELEX:52020PC0825

Proposal for a Regulation of the European Parliament and of the council on the European Health Data Space'. Accessed: May 24, 2023. [Online]. Available: https://eur-lex.europa.eu/resource.html?uri = cellar:dbfd8974-cb79-11ec-b6f4-01aa75ed71a1.0001.02/DOC_1&format = PDF

Radley-Gardner, O., Beale, H., & Zimmermann, R. (Eds.), (2016). *REGULATION (EU) 2016/679 OF The European Parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Hart Publishing. Available from 10.5040/9781782258674.

Regulation of the European parliament and of the council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)'. Accessed: May 23, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri = CELEX:32022R1925

Riskin,D. (2023). Why healthcare data quality matters in the age of AI. Accessed October 23, 2023 from https://www.forbes.com/sites/forbestechcouncil/2023/09/05/why-healthcare-data-quality-matters-in-the-age-of-ai/?sh = 2f4d48883bdd.

Scerri, S., Tuikka, T., de Vallejo, I. L., & Curry, E. (2022). 'Common European data spaces: Challenges and opportunities'. In E. Curry, S. Scerri, & T. Tuikka (Eds.), *Data spaces : Design, deployment and future directions* (pp. 337—357). Cham: Springer International Publishing. Available from 10.1007/978-3-030-98636-0_16.

Shaping Europe's digital future, https://digital-strategy.ec.europa.eu/en, (accessed June 25, 2023).

Siska, V., Karagiannis, V., & Drobics, M. (2023). Building a Dataspace: Technical overview, Gaia-X Hub Austria.

Usländer, T., & Teuscher, A. (2022). Industrial data spaces, Chapter 19. *Designing Data Spaces*. Available from https://doi.org/10.1007/978-3-030-93975-5_19.

Wieringa, J., Kannan, P. K., Ma, X., Reutterer, T., Risselada, H., & Skiera, B. (2021). Data analytics in a privacy-concerned world. *Journal of Business Research*, *122*, 915—925. Available from https://doi.org/10.1016/j.jbusres.2019.05.005.