Article

Not peer-reviewed version

Gaia-X-Med: Experiences with Building Dataspaces for Medical Applications

Bennet Gerlach , Hannes Hesse , Stefan Fischer * , Martin Leucker

Posted Date: 17 October 2024

doi: 10.20944/preprints202410.1403.v1

Keywords: Gaia-X; data sovereignty; federated ecosystems; sovereign data infrastructure; secure authentication; digital contract negotiation; medical and healthcare applications

 doi:10.20944/preprints202410.1403.v1

Article

Gaia-X-Med: Experiences with Building Dataspaces for Medical Applications

Bennet Gerlach [1,†], Hannes Hesse [2,†], Stefan Fischer [1,*] and Martin Leucker [1,2]

[1] University of Lübeck, Germany
[2] Unitransferklinik Lübeck, Germany
* Correspondence: stefan.fischer@uni-luebeck.de; Tel.: +49 451-3101-6400
† These authors contributed equally to this work.

Abstract: Gaia-X, a European initiative, aims to create a digital sovereignty framework for service ecosystems in the future internet. Its applicability to the health domain has been explored in the Gaia-X-Med project, which aimed to establish a common dataspace for various medical use cases based on Gaia-X principles. This paper presents the approach to secure authentication and digital contract negotiation central to this endeavour and discusses the challenges that arose during adoption of the Gaia-X framework, in particular relating to the strict requirements of the European healthcare domain with regards to privacy and consent regulations. By exploring the practical implications of Gaia-X in the healthcare context, this paper aims to contribute to the ongoing discussions surrounding digital sovereignty of both citizens and corporations, as well as its realization via future internet technologies.

Keywords: Gaia-X; data sovereignty; federated ecosystems; sovereign data infrastructure; secure authentication; digital contract negotiation; medical and healthcare applications

1. Introduction

Gaia-X is a European initiative aimed at creating a sovereign data infrastructure. Launched in 2019, it seeks to establish a federated ecosystem where data can be shared and processed securely and transparently within Europe. The project aims to ensure data sovereignty by keeping data within European borders, protecting it from foreign influence, and promoting European technological leadership. Gaia-X also emphasizes interoperability and standardization to facilitate data exchange between different platforms and providers. By fostering a thriving data economy, Gaia-X aims to drive innovation, create new business opportunities, and strengthen Europe's competitiveness in the global market. Gaia-X is a key initiative shaping the future of the Internet, buy building a *trusted*, decentralized, and interoperable digital ecosystem that emphasizes *data sovereignty*, *security*, and *transparency*—crucial elements for the Future Internet. By enabling *federated cloud services* and promoting European standards, Gaia-X contributes to a more open, secure, and user-centric Internet, which aligns with the goals of the Future Internet vision, where data flows freely but securely across borders, industries, and applications.

The official Gaia-X website[1] provides comprehensive information about the initiative's mission, goals, governance, and its role in creating a federated, secure, and transparent European data infrastructure. The European Commission's Gaia-X webpage[2] outlines how the initiative supports the EU's broader digital strategy, promoting data sovereignty and innovation while aligning with European values and regulations.

Gaia-X is a highly complex initiative, with numerous white papers detailing its architecture, principles, and vision for a federated European data infrastructure. While some early implementations

[1] https://www.gaia-x.eu/
[2] https://ec.europa.eu/digital-strategy

 doi:10.20944/preprints202410.1403.v1

have emerged, demonstrating the potential of Gaia-X in sectors like cloud services and data sharing, building fully Gaia-X-conformant solutions remains a significant challenge. The intricacies of aligning with its strict standards, such as data sovereignty, interoperability, and transparency, present hurdles for businesses and developers. Moreover, the integration of diverse cloud providers, adherence to European regulations, and ensuring seamless data portability across borders further complicate the path to widespread Gaia-X adoption.

Gaia-X-Med is a project that explores the application of Gaia-X concepts within the healthcare sector, focusing on six eHealth use cases. The project was performed from July 1st, 2022 to June 30th, 2024 and was financed by the State of Schleswig-Holstein, Germany. By leveraging Gaia-X's principles of data sovereignty, security, and interoperability, Gaia-X-Med aimed to facilitate secure data sharing and collaboration between healthcare providers, researchers, and medical institutions. The project addressed critical challenges such as ensuring patient data privacy, enabling cross-border data exchange, and enhancing access to medical information for research and treatment purposes. Through these use cases, Gaia-X-Med demonstrated how the Gaia-X infrastructure can drive innovation in healthcare, improve patient outcomes, and support the development of personalized medicine, while maintaining compliance with strict regulatory frameworks like GDPR (General Data Protection Regulation[3]).

1.1. Problem Statement/Motivation

Data exchange in digital service ecosystems in the health domain has huge potential for both research and commercial use, yet supporting digital sovereignty remains challenging. It is thus our intention to test the applicability of the Gaia-X approach to establishing such systems via the six eHealth use cases of the Gaia-X-Med project, which cover a broad spectrum of applications, giving insight to its applicability to the domain as a whole.

1.2. Contributions

This paper evaluates the Gaia-X approach for its ability to enable data exchange in service ecosystems as applied to the six use cases from the health domain of the Gaia-X-Med project. In order to do so, we adapted the Gaia-X concepts to fit the requirements of those use cases, including defining and developing four main processes. While doing so, we explain how to instantiate them in order to realize a Gaia-X based application. We also present detailed explanations of the Gaia-X related parts of the implementations of the six use cases common to all of them. Furthermore, we discuss our lessons learned while implementing the use cases overall and whether the Gaia-X approach could satisfy their requirements placed upon it. Lastly, we discuss remaining challenges as encountered in our use cases, but also relevant to Gaia-X based approaches in the health domain at large, leading to future research directions.

1.3. Related Work/State of the Art

Data sovereignty, for a long time a somewhat fuzzy, mainly political term, has grown to inherit a more precise technical meaning: To remain in control of one's personal data. In the context of data exchanges in a digital service economy, this principle has become central to many data space concepts with decentralized federations of participants with full control over their data up to the point of the actual exchange. Many requirements and challenges have been identified, among them interoperability within the data spaces, but also trust among its participants [1–5] and even relating to data security after the exchange. [6] The long-standing effort to realize such data spaces culminated in the International Data Spaces (IDS) initiative [4], whose concepts and implementations [7,8] provided

[3] https://gdpr.eu
[4] https://internationaldataspaces.org/

 doi:10.20944/preprints202410.1403.v1

the basis for the Gaia-X specification documents. Gaia-X-based approaches have been demonstrated to be successful in some application domains, especially mobility [9], [5] with promising approaches in others, like smart cities [10] or agriculture [11]. To that end, implementations of these concepts like Cross Federation Service Components (XFSC) [12], formerly Gaia-X Federation Services (GXFS), and Eclipse Dataspace Components (EDC) [13] have emerged. While attempted [14], the same kind of successful general application of Gaia-X approaches can so far not be seen within the health domain, however.

To give context, much effort has been invested on this besides Gaia-X, like the Medical Informatics Initiative [6] advocating for more sovereign data exchange and many far-reaching projects being developed like the European Health Data Space (EHDS) [7] [15] or electronic health records in Germany (ePa) [8] as well as many other non Gaia-x health data exchange projects [16,17]. The health domain was also a strong focus of the German Gaia-X-Hub, an official Gaia-X institution, since it was conceived with the majority of use cases situated there [18] and even two of its lighthouse projects, Health-X [19] and Team-X [20]. Despite the many health-related efforts and the focus of the Gaia-X-Hub, successful Gaia-X-based approaches have not been demonstrated to the same degree as in other domains, and standardized, interoperable implementations have not appeared or been used as widely.

1.4. Structure of the paper

Section 2 presents the Gaia-X-Med project. It builds the framework for the six eHealth use cases which are used to evaluate the Gaia-X approach. The section specifically focuses on the use cases' requirements. Section 3 then introduces the relevant concepts from the Gaia-X specifications and shows how they had to be modified to match those requirements. Section 4 gives an in-depth explanation of the Gaia-X-related parts of the implementation of the use cases. Section 5 presents lessons learned and remaining challenges. Lastly, Section 6 concludes the paper and discusses directions for further research.

2. The Gaia-X-Med project

A total of four university research institutes, four clinics and eight companies were involved in the Gaia-X-Med project. Thematically, the project was divided into six use cases, each covered by one working group. Additionally, another working group focused on the technical architecture and interoperability.

2.1. The six use cases

At the center of each of the six use cases of the Gaia-X-Med project is a medical application. They vary in many aspects and cover a broad spectrum of the eHealth domain. Some of these applications are intended for commercial use, some of them for research use, some are mainly developed by companies, some mainly by researchers, some of them are customer facing, some target medical professionals, some represent business-to-business (B2B) cooperation between companies and hospitals and some service official national institutions to fulfill their monitoring duties. They all incorporate various data sources with differing levels of associated privacy regulations, from non-personal temperature sensor data to personal, biometric data of individual patients. Therefore, several different data protection measures come into play, ranging from custom concepts over broad consent [9] for research use to individual patient consent.

In the following, we briefly present the application of each use case in detail (see also Figure 1):

---

[5] https://gaia-x4ki.eu/en
[6] https://www.medizininformatik-initiative.de/en/start
[7] https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en
[8] https://www.bundesgesundheitsministerium.de/elektronische-patientenakte
[9] https://www.medizininformatik-initiative.de/en/template-text-patient-consent-forms

 doi:10.20944/preprints202410.1403.v1

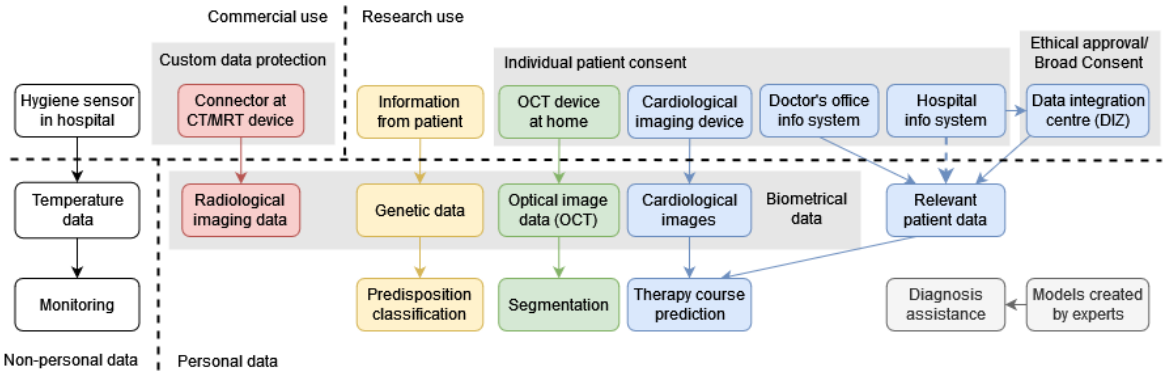

Figure 1. A diagram showing the six use cases of the Gaia-X-Med project, one color each. The top row shows the sources of data, the middle its kind and the bottom its usage.

Use Case 1 – AI-based process optimization for radiology. The exchange of radiologic images has been standardized in DICOM and represents a collaboration opportunity between hospitals and companies to improve diagnosis, possibly enhanced by AI. The seamless exchange of this data technically and organizationally however still remains challenging, thus a device accessing the CT/MRT device producing radiological data is developed to ease this process . This commercial B2B application for hospitals has to respect strict regulations on commercial use of personal or even biometric data of individual patients and thus involves a custom data protection concept, possibly including ethical approval and patient consent forms, as well as medical device development regulation.

Use Case 2 – Integration of patient data for individualized ablation therapy. Ablation therapy is a suggested treatment for certain heart diseases, but its effectiveness varies based on many parameters. To improve its efficacy, three-dimensional cardiologic images of treated patients are collected and integrated with their base data. Based on both, a model is trained to predict expected outcomes for patients with similar characteristics. This prediction is presented as an application for research use by medical professionals relying on personal, biometric data of individual patients, requiring their consent. If the model is to be improved further or has been built upon personal data, consent will and would have been necessary. Here however, routine patient data might be relevant for the model and could be obtained via ethical approval by an associated hospital's data integration centre (DIZ), the data of which is collected with broad consent for research use.

Use Case 3 – AI-based home monitoring of eye diseases. Regularly obtaining OCT images of the eye informs the diagnosis and rehabilitation recommendations made by medicinal professionals. OCT devices provided to patients at home help in obtaining these images and strongly improve patient acceptance, while automatic segmentation further assists in making treatment decisions. A commercial application for medical professionals is developed providing segmented OCT eye images after using a home-based OCT device. The employed data is personal biometric, and its usage requires individual patient consent when using the OCT device and sending the data to be segmented. At the same time, the segmentation process itself might be intended to be improved further via this data or is itself based upon other personal patient data, which in turn will need and would have needed individual patient consent for this type of usage.

Use Case 4 – AI-based causal inference. Classifying genetic predispositions remains an interesting field of research. Connecting the researchers working on classification algorithms and AI models with other researchers able to provide the genetic data sets to train them is facilitated by an open collaborative platform. This application for research use needs data protection measures for each collaboration taking place. If one researcher intends to provide data to another, the initial data protection concept of when the data has been collected has to be checked for whether it permits

 doi:10.20944/preprints202410.1403.v1

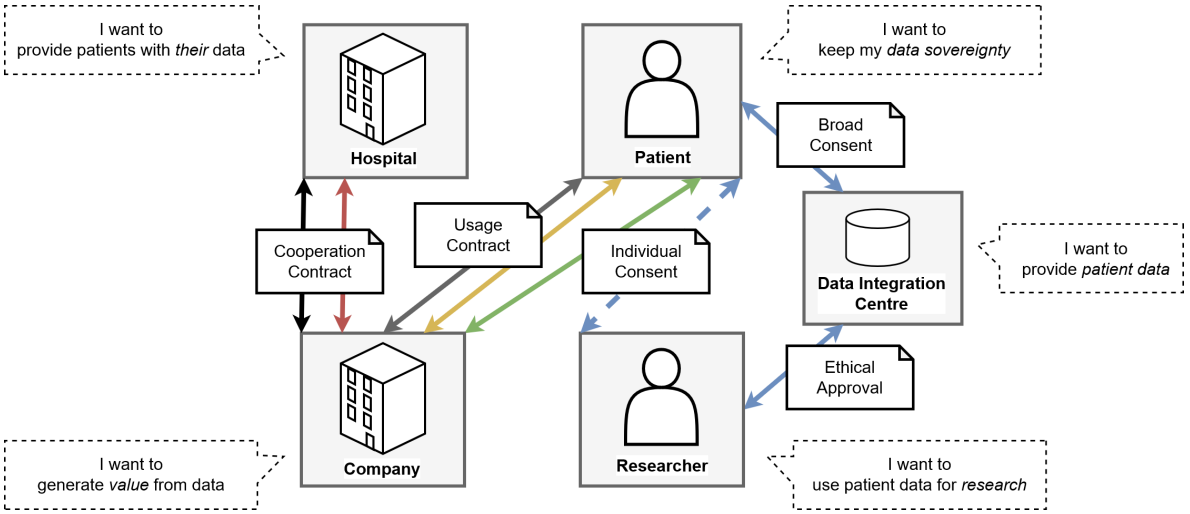

Figure 2. A diagram showing stakeholders in the Gaia-X-Med project, their intentions and agreements between them, which facilitate their interaction while adhering to data security rights. The colors correspond to Figure 1.

this forwarding, as far as it is possible with Broad Consent. Additionally, if classification was to be done on actual patients, additional regulations would apply.

Use Case 5 – HACCP in a Smart Hospital. The state veterinary office has a duty to monitor hygiene processes in hospitals. In order to facilitate this, hygiene sensors—particularly temperature sensors in freezers—are deployed in the hospital and a monitoring application is developed. This commercial application for a state institution has little regulations regarding the privacy of the collected data.

Use Case 6 – Decision Support System. The domain knowledge of medical experts can be formalized as models. Using Large Language Models (LLMs) to generate textual explanations from these models can assist medical professionals during diagnosis. A research use application for medical professionals is developed that does not rely on individual patient data but rather on models hand-crafted by experts, and is thus not subject to the associated data protection regulations.

In each use case, a demonstrator shows the feasibility of the concept, keeping in mind the associated data protection regulations, which typically represent a major challenge during the development of eHealth applications, so as to lay the foundation for later commercialization, if intended.

The fact that all use cases use Gaia-X consistently provides a good picture of how Gaia-X can be used in the medical domain. We therefore now have a look at the use cases' requirements with respect to Gaia-X's services.

2.2. Requirements of the Use Cases on Gaia-X

The use cases are centered around different applications with different requirements each, but also some common challenges, mainly the secure exchange of data between specific stakeholders while respecting data protection regulation. The main stakeholders for all use cases are related to the medical domain and each have their own intentions (see Figure 2):

- Hospitals, as well as doctor's offices, want to provide patients with their data and also use it internally, if possible.
- Companies want to generate value from the data by creating commercial services and products from it.
- Patients want to keep their data sovereignty, including full control over access and distribution.
- Researchers want use patient data for research, and share the results including the data they are based on.

 doi:10.20944/preprints202410.1403.v1

- Data Integration Centres (DIZs) aim to provide patient data to researchers, after anonymization and ethical approval.

Acting upon these intentions requires the involved stakeholders to enter certain agreements, so that their interactions respect the data protection rights of all involved parties. These agreements range from cooperation contracts between institutions like hospitals and companies, to usage contracts with patients as the user of a company's offering, to consent management agreements between patients, researchers and data integration centres.

They all have to be agreed upon beforehand, before any data is exchanged, since many of them represent the consent to the exchange itself. The content of the agreements has to be informative enough for all parties to make an informed decision about consent, especially in the case of individual patients. The given information has to be correct and both parties should be able to verify it, preferably via a neutral third party with sufficient credibility. Underlying all of this is a mutual trust in the other involved parties, which in a digital context has to be based on verifiable identities and information, as well as a legal framework in the case of misbehaviour. So, in summary, stakeholders in the use cases want to be able to maintain each others trust and obtain and respect their informed consent to data exchanges before they take place. In the rest of this paper, we will refer to these requirements as trust and consent. In the following, we will briefly present the scope of the overall project.

2.3. Scope

Of the many features specified by Gaia-X, *trust* and *consent* were determined to be most important and relevant to the use cases and the domain as a whole, since it remains a topic of discussion and research, with frequent regulatory changes still happening, like the German "Gesundheitsdatennutzungsgesetz" (engl. Health Data Utilization Act) [10] and the AI Act [11]. This in particular implies that we do not unify the use cases in terms of what Gaia-X calls data exchange, policies, labels and data governance, all of which are related to the actual exchange of the data after trust and consent have been established, specifying data formats and standards for the exchange, access rights afterwards, life cycle management and more. [21] We chose to keep these aspects outside of our scope, since the individual use cases have little in common in that regard, as some are proprietary and commercial, some are for research, some are based on open standards like DICOM [12] and HL7 FHIR [13], and some are not. We further do not demand any software to be hosted in a centralized location, we aim to be as decentralized as possible and enable the use cases to make those decisions themselves.

3. Gaia-X

Having presented the use cases' requirements on Gaia-X and our focus, namely trust and consent, we next show the Gaia-X concepts relevant to that end.

Gaia-X specifies a framework for service-based ecosystems that enables the sharing of data while maintaining trust between involved parties and retaining the users data sovereignty, thus respecting data protection regulations. It does so in a federated manner, rather than a centralized, cloud-like manner. [22]

To deconstruct this statement and check against our requirements, Gaia-X aims to enable our use cases to share data among their stakeholders and maintains the trust between them while doing so. Further, it allows each of them control over their data, including distribution. This is at the core of adhering to many of the strictest data protection regulations that the use cases have to respect and

---

[10] https://www.bundesgesundheitsministerium.de/ministerium/gesetze-und-verordnungen/guv-20-lp/gesundheitsdatennutzungsgesetz.html
[11] https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
[12] https://www.dicomstandard.org/
[13] https://www.hl7.org/fhir/

 doi:10.20944/preprints202410.1403.v1

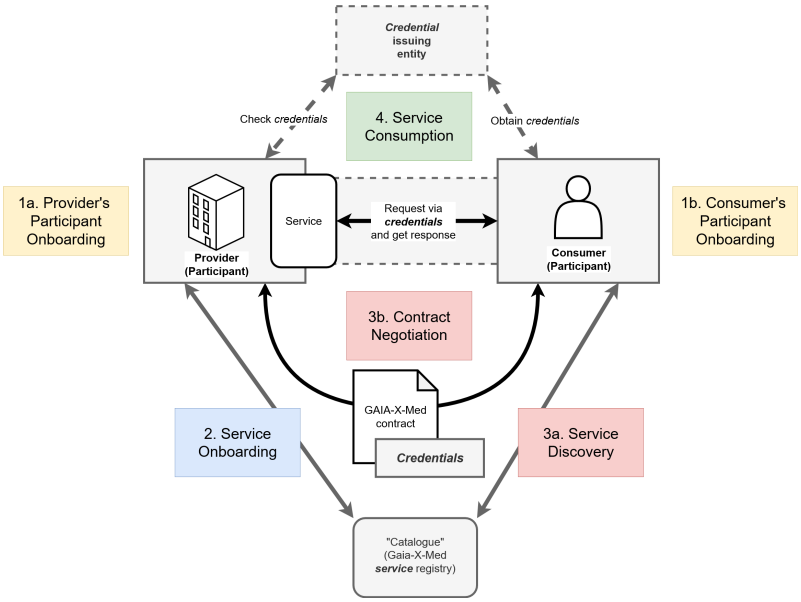

Figure 3. A diagram showing the interactions between the different entities in a Gaia-X dataspace and the resulting processes. Providers provide Services to Consumers, which they can find via a Catalogue. Consumers consume the service after negotiating a contract for its usage with its Provider, which contains Credentials for relevant information.

thereby also implies that some form of consent is given before data is exchanged. How this consent becomes informed has to be determined per use case. In that way, it matches quite well to what the use cases need. To actually achieve this, stakeholders intending to exchange data have to become Gaia-X conformant by adhering to several specification documents presenting many concepts, some technical, some more organizational. [23]

Many of these concepts directly relate to trust and consent and are therefore directly relevant here, while others are out of scope. Some of them are specified in great detail, some are just mentioned, some needed to be extended or adapted to fit the requirements since they were not yet finalized or made compatible to other specifications due to the rapid changes in the Gaia-X specifications overall. To fully confront this, we also decided to freeze the specifications for us in order to be able to implement the use cases on a stable basis. Most notable here were the *Gaia-X Architecture Document* [24] and the *Gaia-X Trust Framework* [25]. Aiming to implement, we therefore modified, extended and adapted several concepts after thorough analysis, basing them on the frozen specifications and trying to remain as close as possible to the original specifications, yet also coherent and useful to our project. In the following, we present the concepts as modified and implemented by us and used throughout the Gaia-X-Med project.

3.1. Architecture of a Gaia-X data space

Becoming Gaia-X conformant by adhering to the specifications starts by entering a *Gaia-X federation*, foremost a collection of potentially data exchanging *Participants*, which serves as the technical and organizational basis for and enables associated data spaces, where these participants come together and actually exchange data. The actual borders of federations and data spaces are somewhat fuzzy, since one can participate in multiple federations simultaneously, enabling different, possibly overlapping data spaces. For the purposes of Gaia-X however, the focus is on the individual exchanges between participants in a federation, rather than associated data spaces as a whole (see Figure 3). Such an exchange is characterized by a *Service* enabling it, with the requesting party being the *Consumer* and the responding party the *Provider*. To be precise, a service in a digital environment necessarily deals with the exchange of some kind of data, but can also include an exchange of physical

 doi:10.20944/preprints202410.1403.v1

goods or the provision of resources. Most of the services relevant for our use cases were completely digital, however. As a core concept, before a service can be consumed, often simply meaning that the data in question can be exchanged, a *Contract* for the consumption has to be negotiated between consumer and provider. This contract represents the consent of both parties to precisely specified future data exchanges and thereby fits our requirement of obtaining consent before data exchanges take place. Consent is also implicitly respected since data exchanges are only to happen if corresponding contracts were negotiated previously. To help the consent become informed, Gaia-X presents the concept of *Credentials*, information about participants given out and certified by *Credential issuing entities*, mostly institutions external to a particular contract. All parties in a contract can then request these entities to verify the information certified by them beforehand, serving as a source of truth about what they negotiate over, addressing the need for correctness in agreements, including informed consent.

While the previous concepts are fundamental to Gaia-X, the following are less precisely specified and modified and adapted by us.

Addressing the requirement for maintaining trust between data-exchanging parties, participants in a federation are required to obtain a particular kind of credential, certifying their identity and their participation in the federation itself. This credential can then be used in the contract and during the later service consumption. This setup enables mutual trust when combined with corresponding real legal person and organization identities checked during onboarding of participants, that are thus subject to law and regulation. In order to facilitate this, several *federation services* are conceptualized, which are supposed to exist once per federation, but not directly associated with any one participant, enabling participants to enter the federation in the first place and form contracts. Helping to find potential providers to form contracts with, participants can use one such a federation service, the *Catalogue*, or service registry, where providers register their service and consumers can browse them. This concept gives rise to the service ecosystem idea, where contracts are not only negotiated between parties previously aware of each other for services with one intended predetermined use, but rather in an open, sharing marketplace of services. We will present more federation services focused on more technical aspects in a later chapter when discussing the implementation.

To be able to specify the interactions necessary to go from initial entering of the federation to the forming of the contract to the eventual consumption of the service, we specified four subprocesses, which we will also give more detail on later. They are the Participant Onboarding, the Service Onboarding, the Contract Negotiation and Service Discovery and lastly the Service Consumption. These processes will also involve further federation services.

3.2. Trust inside a Federation

A core motivation for Gaia-X as a European initiative is to do away with the reliance on large non-European cloud providers, which are not inherently subject to European law, particularly its data protection regulations. When talking about trust, this manifests itself mainly in a decentralized approach with many equal parties and limited managing entities with limited authority or responsibility, as opposed to a few large highly centralized cloud providers. This approach is detailed in the Trust Framework document, an uncharacteristically technical Gaia-X specification, that specifies the involved federation services and even prescribes W3C specifications for several technologies to be used [25].

In the following, we present a small summary of the concepts presented there, which we largely adopted, but also modified in some cases.

Handling trust inside a federation as portrayed by Gaia-X subdivides into two primary aspects—trusting the *identity* of Participants and Services acting in the federation, as well as the information given about them (also called *Claims*.) Both identity and Claims have to be verifiable by the other parties in an interaction for them to be deemed trustworthy. However, it is not enough for Participants to simply present their own Claims without any kind of vetting process to establish

 doi:10.20944/preprints202410.1403.v1

trustworthiness. Therefore, for every kind of possible Claim, independent entities are determined that are capable of certifying the correctness of the presented information, which can then be requested for this verification. Gaia-X calls these third parties *Trust Anchors*. These Trust Anchors could be official institutions certifying the country of residence of a Participant or the location of an involved web server of a Service, thus clarifying the kind of laws potential interactions are subject to. To connect to previous concepts, each certified Claim corresponds exactly to a Credential, and the Trust Anchors to the Credential issuing entities mentioned above. These Credentials then also are substantial parts of the contracts between Participants, which are the basis for their interaction, as previously stated. The Gaia-X Trust Framework prescribes that these Credentials are to be implemented via the W3C specified *Verifiable Credentials* standard [26].

As the second main aspect of trust, one necessary kind of information to be certified in this manner is the identity of a Participant, which needs to be verifiable immediately upon joining the federation. While the identities themselves are to be realized in a decentralized manner through W3C *Decentralized Identifiers* [27], the initial certification still involves a global managing entity responsible for verifying the identities of the legal persons behind Participants before certifying their participation in the federation.

Gaia-X further aims to establish a system of trust on an operative level and defines several further managing entities, collectively bundled under the label of a *Gaia-X Digital Clearing House*. Arguably the most central one is the *Compliance Service*, which is where Credentials are sent to be validated and officially "signed" for compliance—beyond performing routine checks like syntactic validation of the Credential data, it also verifies that the submitted Claims have been signed by any of the registered Trust Anchors in order to maintain a level of credibility. The *Registry Service* exists alongside the Compliance Service inside a Digital Clearing House and contains a list of known Trust Anchors for this purpose. The Digital Clearing House is decentralized in the sense that multiple official instances exist hosted by various third parties which are all running the same versions of the Compliance and Registry Services, setup for redundancy purposes. However, all of these instances are signed off and monitored by the Gaia-X Association to ensure the system of trust is maintained.

However, the established clearing houses at the time we froze the specifications and modified the concepts of Gaia-X were already implementing versions of the Gaia-X concepts without our modifications and were thus unfortunately not directly usable from our point of view. This also manifested itself as many different implementation level incompatibilities, like all Participants requiring VAT numbers, which we will discuss later in more detail, or schemata for certain Credentials not being modifiable to our needs. To remedy this, we decided to setup our own Compliance and Registry Services based directly upon the open source code provided by the Gaia-X federation itself, which we were then able to modify to fit our needs. This in turn enabled us to setup our own standalone federations by including these services in our federation services.

3.3. Addressing conceptual shortcomings of Gaia-X

While many of even the strictest requirements of our use cases, like those related to trust and consent, are adequately addressed by Gaia-X's core and our modified concepts, many of the modifications deemed necessary from our point of view share a common thread and we would like to address this in the following.

Gaia-X aims to provide data sovereignty and individual control over data exchange decisions, which is a requirement of all of our use cases. In some of them, this is conceptually adequately addressed, while some require the data sovereignty of individual persons not associated with an organization, such as patients. They change hospitals, doctor's offices, insurance companies and places of residence and have to be able to make the data exchange decisions themselves, since as legal persons under the GDPR their data sovereignty rights imply individual control at all times. [28]

The Gaia-X specifications do not address this specifically and even conflict with this requirement, since Participants are specified and first and foremost considered to be organizations or companies

 doi:10.20944/preprints202410.1403.v1

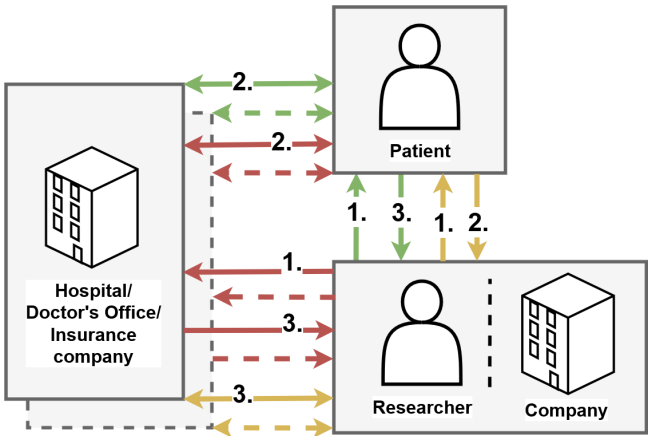

Figure 4. A diagram showing three different flows of patient data from hospitals, doctor's offices, insurance companies, as well as patients themselves, to interested researchers and companies, while keeping the patient in control, one color each.

instead of natural persons—for example, a Participant is required to supply a valid VAT number during onboarding. Originating from the focus on B2B, individual people, thus patients, are considered end-users associated with an organization, which they are either employed by, a customer of or in another kind of business relationship. This is not the case for patients, since they have relationships with many organizations at the same time. This conceptual shortcoming gives rise to the idea of integrating *Personal Data Wallets* responsible for allowing individual people to make the data exchange decisions themselves.

This general shortcoming and the idea of integrating data wallets have been proposed and discussed before in the Gaia-X community, and not only in the medical but also in other domains. This led to the forming of a "Wallet Task Force" [29], a committee to discuss integrating this into the official Gaia-X specifications, which to our knowledge has not happened so far.

While we commend the pursuit of this approach, as it seems fruitful and in accordance with the concepts behind other long-standing efforts like the Gematik [14] and electronic health records (ePa) [**?**] in Germany and other countries like Denmark [15], or Estonia [16], it addresses mainly our most complex use cases with the strictest data protection regulations. Several of our other use cases would, however, be best served by a simpler approach, compromising on the applicability to the complex cases, if necessary.

To motivate our approach in addressing the shortcoming, we briefly present and discuss the options we explored. Figure 4 shows the three different options as flows of data, one color each. In the red flow, we show the most common way in which the problem is currently discussed, where an interested researcher or company first contacts the institutional origins of patient data directly. Patient data in this case is most often considered to stem from hospital or doctor's office information systems, less so from the patients directly, despite them also potentially contributing much medically relevant information directly, via wearables or questionnaires during rehabilitation. Many institutional pathways to accessing the patient data are developed here. For example, Data Integration Centres (DIZ) provide anonymized routine data of multiple patients after ethical approval by its associated hospital of the requester's stated purpose, which has to be for research. This institution can act as the owner of the data in this case because patients signed away their data sovereignty via a broad consent agreement and enabled the hospital to decide over the distribution of the data in question. While a valuable

[14] https://www.gematik.de/
[15] https://healthcaredenmark.dk/national-strongholds/digitalisation/digital-infrastructure/
[16] https://e-estonia.com/solutions/e-health/e-health-records/

 doi:10.20944/preprints202410.1403.v1

pathway for many scientific endeavors, it stands in contrast with the full data sovereignty approach proposed by Gaia-X. To remedy this, instead of the institution making the decision, it would rather have to act as a custodian and ask the patients directly before sending any data. As the first problem with this data flow, this effectively places the burden of consent management on the hospital, doctor's offices, and so forth. Contacting a patient for routine checkups or questions during rehabilitation is already challenging for many medical professionals, having to forward each data request to all previous patients seems organizationally infeasible. Secondly, an interested researcher or company will have to contact many institutions, since individual patients can visit many hospitals, doctor's offices, change insurance companies, move cities, etc. Correlating all this information into a coherent data set is of integral medical use in many cases, however.

The yellow flow aims to remedy these problems by first contacting patients directly and obtaining their authorization to access the relevant, agreed-upon parts of the data they would be able to access at all involved institutions, thus keeping the patient fully in control of the data. We therefore consider this approach to match Gaia-X very well and advocate for its inclusion. However, when integrated via patients effectively becoming trust anchors and certifying access rights for the requesters, this does not solve the problems fully, since patients themselves can also be a source of data.

Lastly, the green flow addresses this last point and is the approach we decided to employ. Patients here are regarded as the single source of their data from the point of view of the interested researchers and companies and they simply contact the patients directly, who then decide directly which data to request from its origin and to distribute to a requester. This is the most straightforward version of data sovereignty and in order to implement it via Gaia-X, we decided to simply integrate individual persons, thus patients, as participants into a Gaia-X federation directly. They go through with onboarding, negotiate contracts and even provide services, if necessary. This way, they retain their full data sovereignty, despite incurring the technological burden of interfacing with the federation services and other participants, typically organizations, on a technical level. The exact technical burden and whether or how it differs from the data wallet approach is an interesting future research question and we will discuss this further, when presenting our implementation.

We also identified many more interesting related research directions, which we present in the following. On a technical note, the yellow flow, with patients as trust anchors, actually likely incurs way less data sent over the underlying communication network and does away with the need for potentially huge storage for patients, in the case of images for example, so should be considered for reasons of efficiency and thus resource footprint, including energy. So enabling both the green and yellow flows would likely amount to the best approach. For simplicity, we started with the green approach, but are also interested in combining them in the future. As an organizational challenge, how to become aware of patients having the data you are interested in, is challenging to many researchers and companies. In the green flow we chose to implement, we intend patients to publish their data offering in the federation's catalog, but other approaches may be more adequate in this and especially other flows of data. Health-X is working on this in a Data Wallet approach. As a final point, the value gained in increased acceptance by the involved patients by utilizing official institutions like Data Integration Centres (DIZ) with their ethical committees, their association to the hospital and their rigid use of anonymisation is not to be underestimated. While we in a principles first manner integrated the patients directly into the federation, there is much space and arguably need for these established pathways of data exchange to still be used and we therefore would encourage them to become integrated and enter as well. DIZs would become participants and all agreements, like broad consent or even ethical approvals, would then become contracts, between on the one side patients, on the other researchers. If similar institutions are on the horizon for commercial use as well, their integration could also be hugely beneficial. Even parts of them, like just registers for certain groups of patients or just anonymising services, could prove to be useful.

 doi:10.20944/preprints202410.1403.v1

4. Gaia-X-Med : Our solutions

Given the task of fulfilling the requirements of the use cases relating to trust and consent and having presented the relevant concepts from the Gaia-X specifications and our adaptations thereof, we next present in this section our implementation of these concepts as it was used by the use cases throughout the Gaia-X-Med project.

4.1. Existing software implementations

We considered several existing implementations of some of the relevant concepts to incorporate into our implementation. We decided on this at the time we froze the specifications for ourselves and adapted the concepts, so it is worth reconsidering them in the future, as well as others that have been developed since.

The Cross-Federation Service Components (XFSC), formerly GXFS, implemented many of the Gaia-X concepts, not just those relating to trust and consent. [12]

While providing a lot, we ultimately decided against utilizing this framework, since it would also incur lots of compexity in aspects that were out of our scope. Further, at the time they underwent lots of change in implementation and documentation due to the rapid changes in the Gaia-X specifications themselves, which often made it hard for us to properly utilize the framework. Despite not using it as is, we closely studied much of the framework and often based our implementation on the same underlying technologies. We also were able to incorporate parts of XFSC with little configuration or modification, like the Compliance service, which we will present later.

The Eclipse Dataspace Components (EDC) represents a data space connector that is mainly concerned with what Gaia-X calls Data Exchange and has its roots in the International Data Spaces (IDS) project. [13]

Since we focus mainly on trust and consent, Data Exchange is out of scope for us and incorporating the EDC fully would add a lot of complexity to our implementation. We have however closely studied its version of contracts, which is a central Gaia-X concept, to inform our implementation. Additionally, as a connector it is server-like and always online, even on the consumer side, which contrasts our asynchronous web technology based approach, that we present in the following.

4.2. Web-based approach

While the existing frameworks obviously all are also based on innately asynchronous, commonly used web technology, we consider our approach to go even further into that direction.

As previously stated, with the goal in mind of enabling individual persons, thus patients, to share data using Gaia-X, while still retaining their data sovereignty, we modified the concept of Participant of a Gaia-X federation to also include individuals, for whom this participation likely represent a higher technological burden than for organizations. We therefore tried to make all processes in which a Participant interfaces with our federation services as accessible as possible and encourage all providers in the federation to do the same with their services. This accessibility manifests itself by not using a connector component that has to be configured and programmed, but rather as the services during the necessary interactions presenting themselves as web pages which are accessed by a normal web browser and structured to guide through the process. We pushed into this direction as far as we could, only compromising in questions of trust, particularly establishing domain-based trust, and consider going even further an interesting research direction. This might even turn out to be a valuable approach to alleviate some of the resource scaling problems on the consumer side that some connector based approaches face, since an interaction initiated by a consumer is handled nearly entirely in the browser and leaves virtually no resource footprint on their side after the interaction, as is the case in most web page usage in browsers.

 doi:10.20944/preprints202410.1403.v1

4.3. Yo-Ga-X Framework

While also implementing all the use cases of course, we deemed a subset of the implementation, particularly features relating to trust and consent, to be useful or interesting enough for general usage in other domains or for testing or learning about Gaia-X. We thus made this code open source as a framework itself, called *Your Own Gaia-X (Yo-Ga-X)* [17]. It includes all components necessary to setup a standalone federation based on our modified Gaia-X concepts. Yo-Ga-X as a framework is also lightweight, since every component is kept simple and thus easy to extend or even replace. It is also easy to setup, as each component can be setup locally or deployed independently and setup guides and introductory examples are provided.

Yo-Ga-X is actually the very framework we used to implement the Gaia-X related parts of the use cases for the Gaia-X-Med project and includes all used federation service components. Instead of the use case specific code however, it includes a general demonstrator for an example service consumption of a Consumer at a Provider's service. In addition to making our code open source, we also provide detailed supplementary documentation, both introductory to the concepts as well as technical for the code and also detailed graphical overviews of the processes or the data flow. Lastly, this documentation also includes introductions to the concepts and how-to guides on how to navigate a set-up federation from the point of view of a participant. We present an excerpt of this documentation in the following.

4.4. The four main processes

To reduce the complexity of the overall process, the Yo-Ga-X framework defines the following four sub-processes (see Figure 5), each focusing only on a few involved participants and Federation Services:

- *Participant Onboarding*, where both Providers and Consumers initially give information about themselves and join the Gaia-X-based federation, which is the base for the Yo-Ga-X-powered service ecosystem,
- *Service Onboarding*, where Providers give information for and register their to-be-offered Service in the federation's Catalog,
- *Service Discovery and Contract Negotiation*, where Consumers browse the Catalog to find Services they are interested in and initiate the negotiation of a Contract with the Provider of a chosen Service,
- *Service Consumption*, where consumers finally consume a Service they negotiated a Contract for, including the authentication at the Provider.

In the following, we present these processes in further detail.

---

[17] https://docs.gaia-med.org

 doi:10.20944/preprints202410.1403.v1

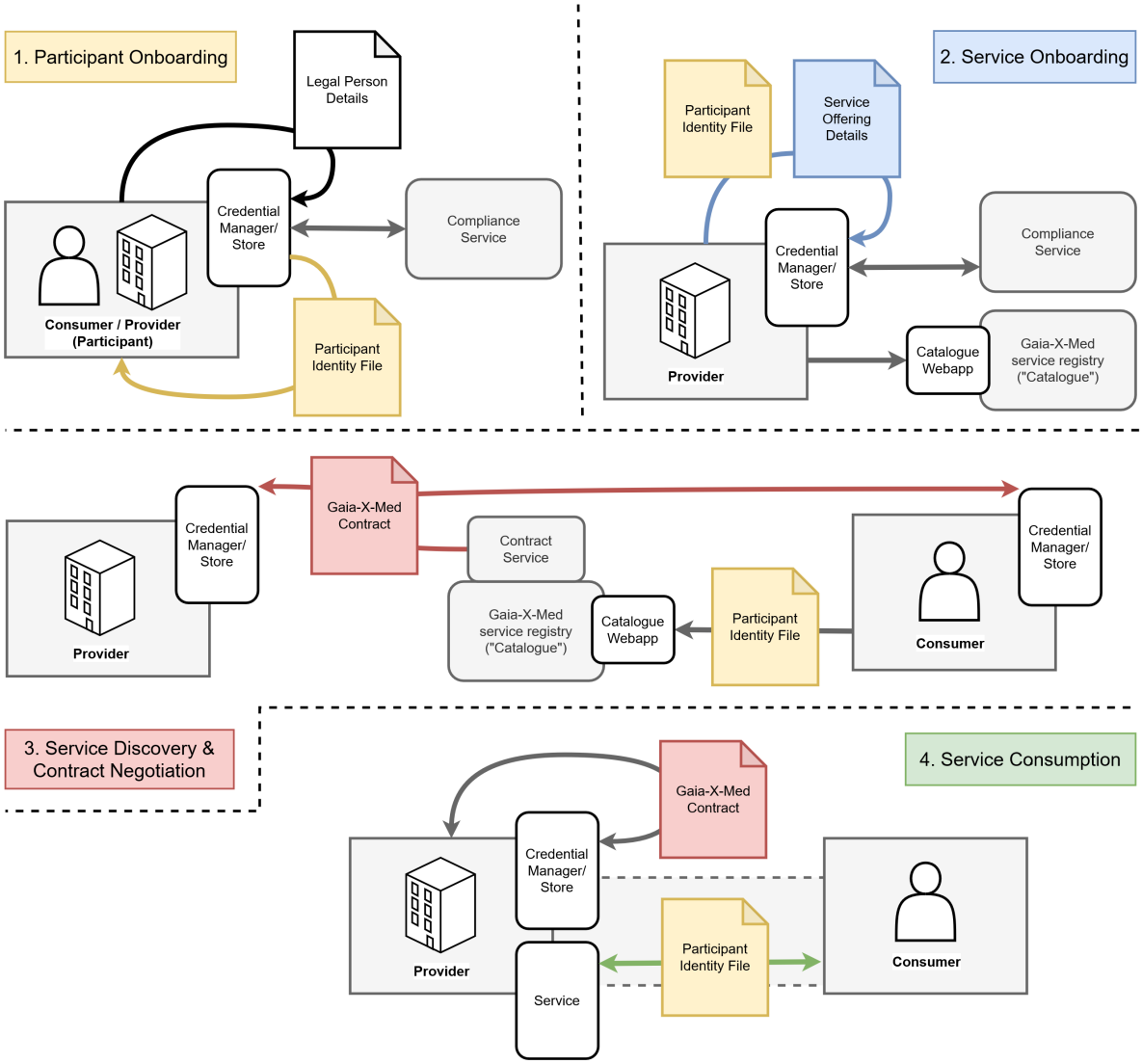

Figure 5. Four diagrams showing each of the four main processes in detail. First, a participant is onboarded, then a service of theirs, next the service is discovered by another participant and both negotiate a contract for its usage, so that finally the service can be consumed.

During Participant Onboarding, all Participants—Consumers as well as Providers—first create a credential containing identifying details about themselves and their legal person which they wish to share with other Participants. The federation's *Compliance Service* then validates the credential's structure and certifies it as being part of the federation. Finally, the Participant receives a *Participant Identity File*, representing his participation in the federation while simultaneously acting as the passkey for authentication.

During Service Onboarding, Providers first create a credential containing details about the particular Service they intend to provide and certify that they are the Participant responsible for its management. Again, the Compliance Service validates the credential's structure and certifies the Service as being part of the federation and that it belongs to the Provider. Finally, the Provider registers the Service in the *Catalog*, another Federation Service, where it can be discovered by potential Consumers.

During Service Discovery, Consumers first authenticate themselves at the Catalog as a Participant of the federation. They are able to browse it and choose a Service they intend to consume. This choice then initiates the *Contract Negotiation* process, during which the *Negotiation Service*, yet another

 doi:10.20944/preprints202410.1403.v1

Federation Service, corresponds with the soon-to-be Consumer and the Provider of the Service to form a Contract expressing the informed consent of both parties to the future service consumption.

During the actual Service Consumption, Consumers attempt to authenticate themselves at the Provider's Service using their Participant Identity File, containing their certificate of participation in the federation. Since this identifying certificate is also part of the Contract the Provider received during the Contract Negotiation process, the Provider can verify the requesting party's identity and provide the Service according to the negotiated conditions.

Having presented the four main processes and how they fit together, we next present how we implemented the systems that enable them.

4.5. Identity and its problems

In order to establish a solid basis for trust in a federation, Gaia-X prescribes the use of Decentralized Identifiers (DIDs), as previously mentioned [27]. As we will see in detail in the following, these identifiers are used widely throughout the federation, for example in Credentials and thus Contracts. As a trust enabling mechanism, they can be verified, clarifying which legal person taking part in the federation is behind the identifier, thus the subject of a credential or one of the parties of a contract. While specifying the technology, Gaia-X still leaves open its implementation of which there are several and we thus discuss the two general approaches we mainly considered, the blockchain and the web domain approach, to motivate our choice.

As per the W3C DID specification, in both approaches the controller of a DID, which here is also its subject, specifies the method of verification of a DID and, for practical reasons, ensures the facilities to go through with it are in place [27]. In the blockchain approach, the DID controller and subject, thus the participant behind the identifier, takes part in a in a distributed ledger storing all its data, possibly utilizing cryptographic Wallets. To sufficiently enable trust and acceptance among participants however, the choice of distributed ledger or blockchain seems pivotal, ranging from individual to federation wide to european to global distribution and from commercially focused crypto currencies like Ethereum [18] to ones supported by official european institutions like electronic IDentification, Authentication and trust Services (eIDAS) [19], European Blockchain Services Infrastructure (EBSI) [20], european Self-Sovereign Identity Framework (ESSIF) [21], aligning more with Gaia-X as a european initiative. While seemingly feasible and often a topic of our discussion with fellow researchers and developers working with Gaia-X, we ultimately decided against this approach, since we determined its integration with the Gaia-X concepts as adapted by us to be too resource-intensive a task for a problem relevant only to some of our use cases and likely a project on its own. Exemplary of this, one of our main adaptations was the integration of individual persons as participants in the federation, where they previously had been end users associated with exactly one organization. While a federation wide blockchain distributed among organizations as the basis of trust seems reasonable, considering this as a form of identity verification for individual people in our estimation has to imply the use of official institutions like eIDAS, EBSI or ESSIF, whose integration with Gaia-X was missing at the time we developed our implementation. Due to its potential in solving similar problems encountered by other projects in the health domain and beyond however, we encourage the pursuit of this approach by Gaia-X itself, including its conceptual integration as a technology into the specifications, alongside our adaptations if needed, and its implementation via XFSC, for example.

While potentially useful, the blockchain approach is not without problems, even mentioned in the abstract of the specifications of the did:web Method [30], the core technology of the alternative web domain approach. As stated there, DIDs that target a distributed ledger, as is the case in the above

---

18 https://ethereum.org/en/
19 https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation
20 https://ec.europa.eu/digital-building-blocks/sites/display/EBSI/What+is+EBSI
21 https://essif-lab.eu/

 doi:10.20944/preprints202410.1403.v1

approach, often don't get adopted in mass due to not accumulating enough meaningful trusted data around the associated identities. Despite likely remedied once the mentioned official institutions are involved, this has to be considered for at most federation wide ledgers. Representing an alternative, the web domain approach uses the did:web Method and prescribes the DID controller, and in this case also the subject, to be the domain directly associated with the participant behind the identifier. Verifying an identifier in this case then amounts to checking whether the correct domain is at the other end of the communication. Whether this association of participant and domain is a sufficient basis for trust depends then on the reliability of the employed Domain Name System (DNS) and the verifiability of domain name to legal person, organization or individual, behind the domain. While a feasible approach for organizations with sufficient technical capabilities, such as web servers to host the systems and possession of unchanging, unique, verifiable domains, this incurs a siginificant technological burden to the average individual person, who typically interact with the internet via unnamed, dynamically changing IP addresses and not hosting long time frame servers, but rather as short time frame clients. Despite thus not fully addressing the requirements of all use cases of our Gaia-X-Med project, we compromised on this second approach due to the first being to resource-intensive to integrate for too little gain for our project specifically. Although we did not employ it however, we outlined the tremendous potential of the blockchain approach in solving challenges common to the health domain as a whole and beyond, we therefore encourage its pursuit in the future. In the following, we present the systems at the core of our implementation that enable the presented four main processes based on the web domain approach to verifiable identifiers just discussed. In doing so, we often contrast it to the unmodified, conventional Gaia-X concepts and their implementations like XFSC/GXFS and EDC.

4.6. Authentification

In a conventional Gaia-X dataspace, Participants are organizations which are identified via signed Gaia-X Credentials that attest their trustworthiness. However, acting parties within a dataspace are usually not the organizations themselves, but rather their employees; individual users that receive authorization to carry out their work. The Gaia-X Architecture therefore proposes an authentication scheme that allows an organization to issue *Principal credentials* to their employees. These credentials link back to the organization's Gaia-X Credential and allow the individuals to act on the organization's behalf. This is solved through two major components; the *Organization Credential Manager (OCM)* and a *Personal Credential Manager (PCM)*, the latter being a smartphone app that allows users to access a blockchain-based wallet housing their credentials, as implemented within the official XSFC toolbox.

Despite its use of relevant technologies, this implementation is based on the unmodified Gaia-X concepts, which we found inadequate and thus modified in order to integrate patients as participants directly, addressing the conceptual shortcoming of patients not always being able to be associated with exactly one organization. Rendering this existing implementation incompatible, we therefore opted to implement a much simpler authentication mechanism, partially inspired by the EDC and based on the web domain approach to identifiers just discussed. The Gaia-X Credentials issued to Participants form the authentication credentials themselves, and they are addressed and uniquely identified through the W3C Decentralized Identifiers that specify their storage location. The Credentials are hosted by the Participants inside self-hosted wallets, and they are accompanied by keyfiles obtained during the onboarding process.

Onboarding

Obtaining Gaia-X Credentials is a relatively involved, multi-step process that includes consuming the API of several Trust Framework services like the Gaia-X *Compliance Service* as well as bundling and signing W3C Verifiable Credentials multiple times. To make this process more accessible, we have developed a web application called the *Credential Manager*. It provides a simple form to the user asking them for their Claim data, then implements the necessary API calls to the Trust Framework services as well as Verifiable Credential creation and signing. As outlined in section 3.2, the Gaia-X

 doi:10.20944/preprints202410.1403.v1

Trust Framework stipulates that Providers host their Credentials in a decentralized manner, e.g. on their own infrastructure. These Credential storage solutions—*wallets*—are to be established through W3C Decentralized Identifiers by encoding the real storage location into an abstract address as a *DID URL*. Through the different *DID methods*, this then supports various kinds of storage configurations, corresponding to the aforementioned verification methods for DIDs.

As per our web domain approach to identifiers, for our implementation, we decided on the `did-web` method, which leverages domain-based trust based on HTTPS and DNS—owning the domain name infers that you own and are responsible for the data accessible through it. [30] DID URLs using the `did-web` method get resolved to plain HTTPS URLs, and web-based storage wallets are simple and straightforward to implement. To that end, the Credential Manager application furthermore works together with another piece of software we developed called the *Credential Store*; a simple *nginx*[22]-based HTTPS wallet solution that Participants can easily self-host on their infrastructure. After configuring the Credential Manager with the Credential Store access keys, the Manager can automatically upload the Credentials to the wallet. At the end of the process, aside from the Gaia-X Credential itself which gets uploaded to the Participant's wallet, the Credential Manager further creates a unique *keypair* and two other important documents:

1. A DID document, which contains the keypair's public key as well as a link to the location of the Gaia-X Credential. It is uploaded to the Participant's wallet, to a location that is resolvable according to the did:web standard.
2. A passkey called the *Participant Identity File*, which contains the keypair's private key as well as their DID URL. It is encrypted with a passphrase chosen by the user, and provided for download.

Authentication

A Participant initiates the authentication process by creating a *Login token*, which is a *JSON Web Token* [23] containing their DID URL and signed by the private key contained within their Participant Identity File. This token is sent to a verifying party; for example, the authentication layer of a Provider Service.

By resolving the DID URL contained in the token into an HTTPS URL, the verifying party can then download the DID Document from the Participant's Credential Store. Since it contains the public key corresponding to the private key used to sign the token, the token's authenticity can be verified by simply checking the signature against the public key.

In a next step, the verifying party follows the link contained in the DID Document to the Participant's Gaia-X Credential, which they can then validate according to the rules defined by the Gaia-X Trust Framework; this includes validating the Credential against its schema, and verifying its Gaia-X Compliance. Once both the token and the Gaia-X Credential have been validated, the verifying party can trust the Participant's identity as well as the authenticity of their Claims.

In the Gaia-X-Med architecture, this authentication flow is largely implemented as a federation-wide service called the *Authentication Service*, which takes as input the Participant's Login token, performs the steps described above, and returns on success the Participant's Credentials—now considered to be trustworthy. Although the necessary steps are openly documented and could thus be performed entirely on the side of the Provider, we decided to provide this service as a convenience so that Providers don't have to implement the entire process themselves. Figure 6 shows a simplified overview of the process.

---

[22] https://nginx.org/
[23] https://jwt.io/

 doi:10.20944/preprints202410.1403.v1

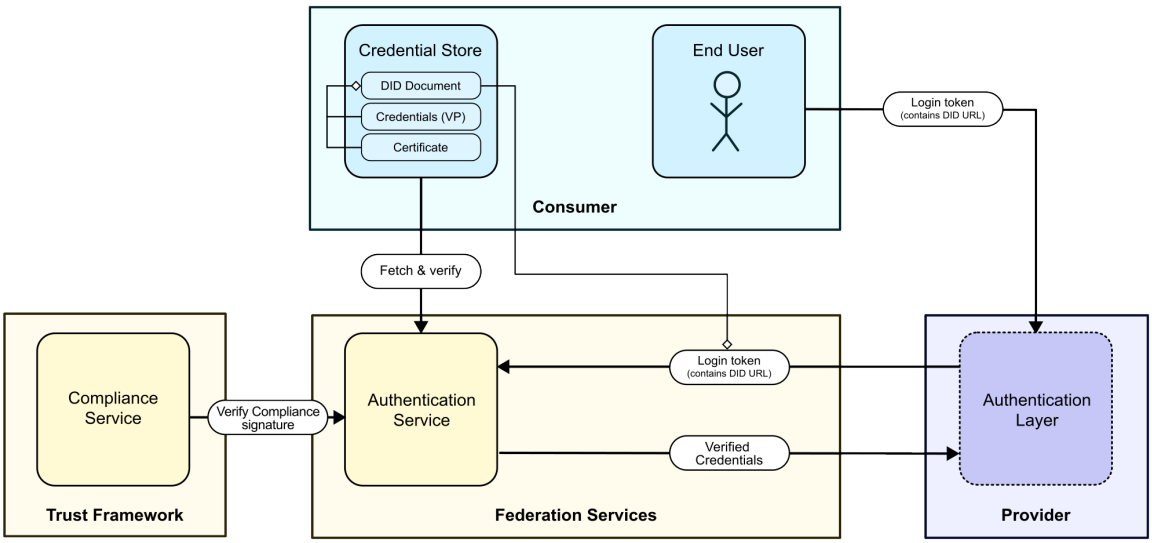

Figure 6. A simplified overview of the Gaia-X-Med authentication flow. A Consumer submits a locally signed Login token to the Authentication Layer, which forwards it to the Authentication Service. After fetching and verifying the Consumer's identity and checking their Compliance Credential signature, the Authentication Service responds with the now verified Credentials, which can then be trusted and used further by the Provider.

As for integrating this authentication flow into the backend architecture of a Provider Service, we developed two methods, each covering a different use case. The first combines the authentication flow with **OpenID Connect**, a widely-used standard for enabling single-sign on authentication. [31] To realize this, an *OpenID Connect Identity Provider* based on `node-oidc-provider`[24] was developed and hosted as a federation-wide web application. After asking the user for their Participant Identity File and decryption passphrase, the application generates a Login token client-side that is then sent to the Authentication Service. The returned Participant Credential is mapped to the corresponding OAuth2 claims, which are finally forwarded to the Provider Service.

The Provider can use any of the various available OpenID Connect clients to enable this authentication method; popular choices include *Keycloak*[25] and proxy server modules like `mod_auth_openidc`[26] for use with the *Apache HTTP Server*[27].

The OpenID Connect method is particularly suited for easily integrating interactive Gaia-X-Med authentication for web application based services. Figure 7 shows a schematic overview of the OpenID Connect method.

---

24 https://github.com/panva/node-oidc-provider
25 https://www.keycloak.org
26 https://github.com/OpenIDC/mod_auth_openidc
27 https://httpd.apache.org/

 doi:10.20944/preprints202410.1403.v1

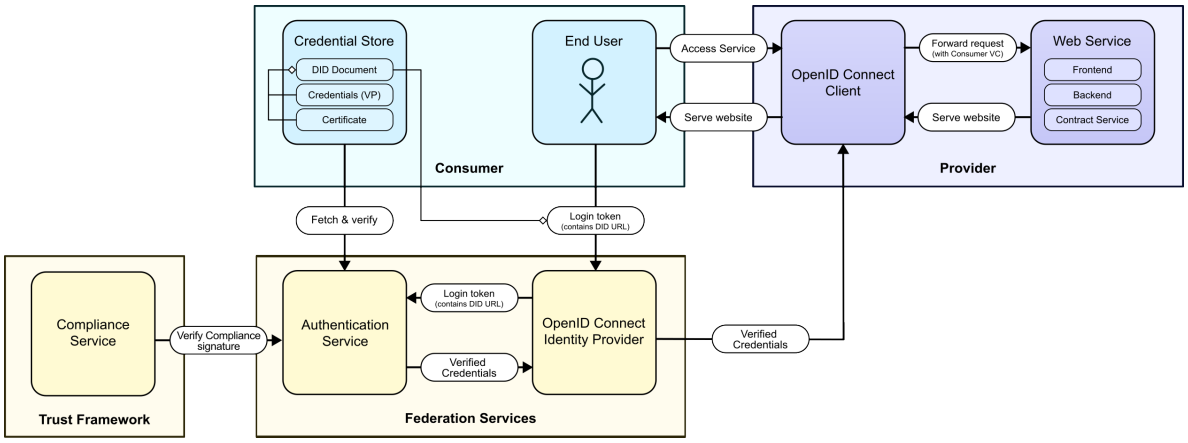

Figure 7. An overview of the OpenID Connect authentication method. A compatible OpenID Connect Client is deployed by a Provider in front of the Service, which redirects the user to the OpenID Connect Identity Provider and forwards the resulting OAuth2 credentials to the Service backend.

The other method is what we dubbed the **Proxy method**. Here, we developed a middleware solution called the *Authentication Proxy*, which is deployed in front of a Provider Service and intercepts incoming requests from Consumers. Before allowing requests to reach the Service, Consumers are required to register an authenticated session with the Authentication Proxy using a Login token. To allow simple interfacing using this method, we have provided clients written in Python and TypeScript that implement the necessary communication protocol with the Authentication Proxy transparently; from a developer's point of view, after instantiating a client using their Participant Identity File and decryption passphrase, it provides a simple HTTP request API that automatically registers a session and uses it to make authenticated requests.

The Proxy method was developed as a very simple, plug-and-play solution for enabling automated requests to API-based Provider Services, as it does not require human interaction as opposed to the OpenID Connect method. (see Figure 8)

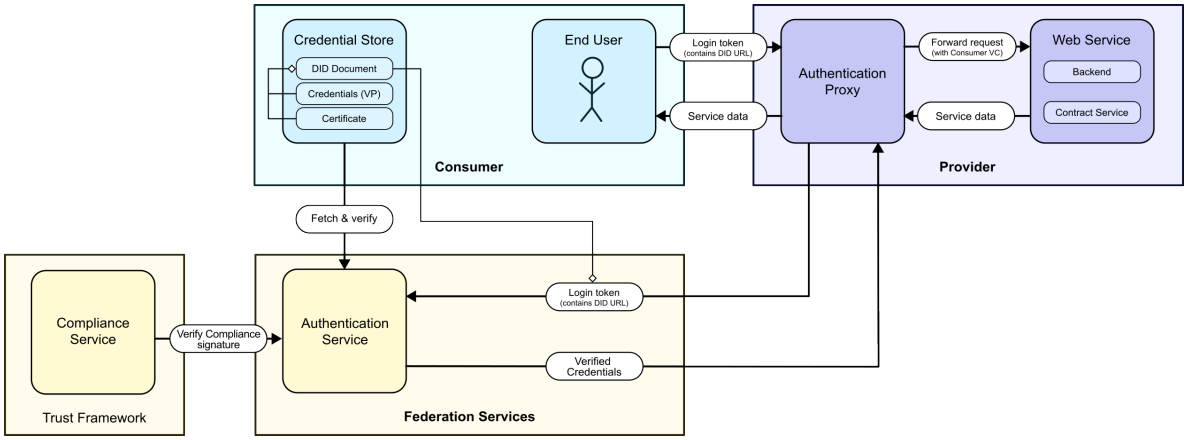

Figure 8. An overview of the Proxy authentication method. The Authentication Proxy is deployed as a middleware by a Provider in front of their Service. It transparently handles authentication and only lets requests with a valid session pass.

4.7. Contract Negotiation

XFSC [12] enables authorization through a *Policy Engine* that, once deployed and configured, allows Providers to set certain rules to handle incoming requests from Consumers. We have decided to approach the topic of *authorization* by using *digital contracts* as a base. Contracts are only briefly touched upon in the Gaia-X architecture itself; stating that while they form an important part of the ecosystem,

 doi:10.20944/preprints202410.1403.v1

their implementation details are out of scope of the architecture. [24] We had three initial requirements for our digital contracts—they should be machine readable, able to be digitally signed by involved Participants, and tamper-proof. We found that the W3C Verifiable Credentials standard, which was already being used in the project for enabling Gaia-X Credentials, fulfill all of these requirements:

- they are formatted as JSON-LD and are furthermore validated against schemata, so they are *machine readable*;
- they can be *digitally signed* with cryptographic keys, which effectively renders them *immutable* as well;
- additionally, they can be bundled by anyone into *Verifiable Presentations*, which can also be signed individually from the Verifiable Credentials within. This provides an excellent method for allowing a neutral third party to "notarize" a collection of Credentials.

A Gaia-X-Med Contract goes through four steps, as shown in Figure 9. The content of a Contract is comprised of two parts; for one, a static, free-form (non-machine readable) non-negotiable text part which the Consumer has to agree to. The other part is defined through the *Contract Template*, which is a *JSON Schema*[28] object embedded inside a Service Offering's Credential. Through the Contract Template, a Provider can offer configurable parameters for the Service to the Consumer, which allows them to negotiate certain aspects of the Contract. For example, a web hosting service could have a configurable amount of data storage.

When visiting the Catalog web application and picking the "Negotiate contract" option on a Service they wish to consume, the Consumer user is presented with the Service's Contract Template rendered into a form as well as the static contract terms to which they have to give their explicit agreement. The data from the submitted form is bundled and signed into a *Contract Offer* (as a Verifiable Credential) and sent to the federation-wide Negotiation Service, which acts as a notary between the Consumer and the Provider. It creates a Verifiable Presentation containing the Consumer's Contract Offer and adds its own signature, creating the *Notarized Contract Offer*, which it then forwards to the Provider. At the Provider's end, a service called the *Contract Service* is listening for incoming Notarized Contract Offers from any whitelisted Negotiation Service. Once an offer arrives, it too performs some routine checks including re-confirming the authenticity of the Consumer making the offer and validating that the offer aligns to the Contract Template. It then forms a response to the offer. By default, all valid offers from valid Participants are accepted; however, a Provider is also able to extend the Contract Service's negotiation routine with programmatic rules which, based on the Consumer's identity and the content of the offer, result in one of two additional outcomes:

- The offer is *denied*—for example, a Provider can make the decision to not accept offers coming from Consumers having a certain Claim; e.g. restricting the Service to Consumers only from research organizations.
- The offer is *accepted under modified conditions*—if the Provider is willing to form a contract but not under the conditions specified by this Consumer. In this case, the Contract Service is capable of responding with an adjusted *counteroffer* that the Consumer can choose to accept instead.

In case the Contract Service accepts the offer, it proceeds to respond with a Contract Offer of its own, carrying the same contents and signed using the Provider's identity. The Negotiation Service finally bundles the Contract Offers from both parties into a signed Verifiable Presentation. Since the Presentation is signed by the Negotiation Service's private key, none of the contained Credentials can be modified or replaced without breaking the signature verification, making it effectively immutable. The finalized Contract is forwarded to both Consumer and Provider. On the Provider side, it is stored inside a database and can then be used to authorize Consumers—after an incoming request,

[28] https://json-schema.org/

 doi:10.20944/preprints202410.1403.v1

the Provider Service backend is required to check whether a currently valid Contract exists for this Consumer, and if there is none, the request is to be denied.

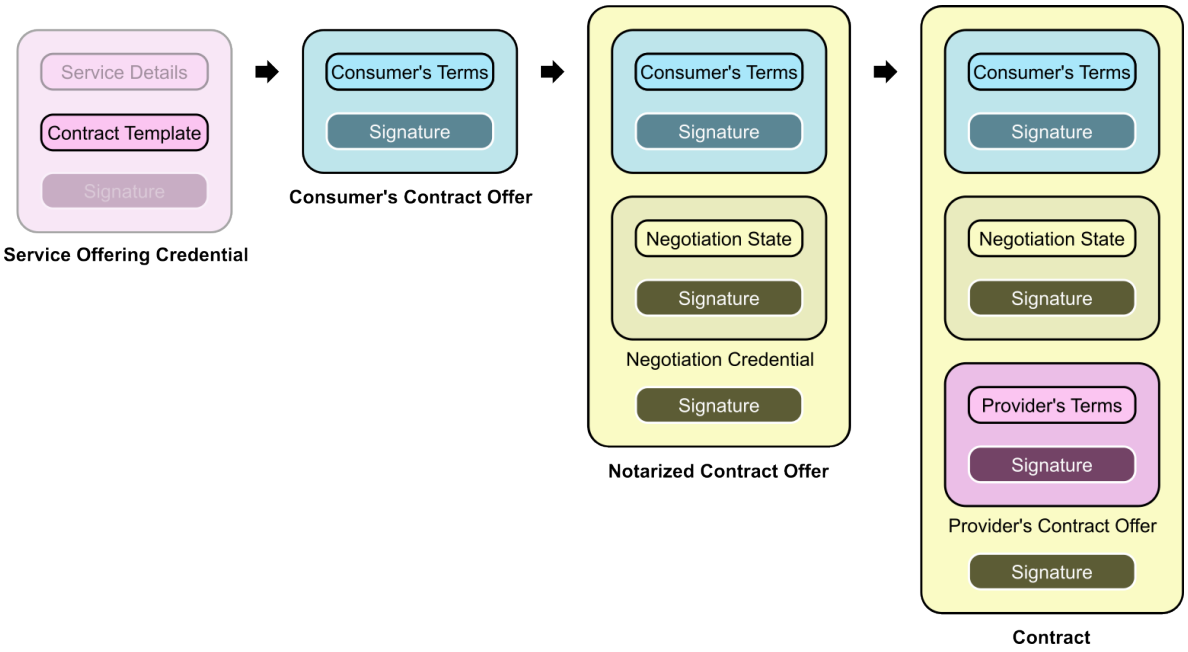

Figure 9. How a Contract is formed in four steps. The signed terms and identities of both Consumer and Provider, as Verifiable Credentials, as well as a *Negotiation Credential* containing negotiation metadata, are bundled by the federation-wide Negotiation Service into a Verifiable Presentation, which is signed as a whole, resulting in the finalized Contract.

5. Evaluation/Lessons Learned

Having presented large parts of our implementation of the presented concepts, we will now discuss whether the use cases of our Gaia-X-Med project were able to utilize it to adequately address the requirements they had on the Gaia-X based approach. Since the use cases cover a large range of possible requirements in the health domain, this, in our opinion, gives valuable insight into the applicability of Gaia-X to the health domain overall. In the following, we present our experiences both common to all use cases and different among them.

5.1. Common experiences among all use cases

Starting technical, all use cases were able to utilize our implementation within their codebase via our Yo-Ga-X framework. They were able to create Participants, onboard them onto a common federation with its federation services hosted by us, publish Service Offerings for Services provided by their Providers to a common catalogue and have their Consumers discover them, negotiate corresponding contracts and finally consume them, all while making use of the appropriate federation services, which checked the compliance of the Contracts and Credentials.

We decided on a common federation for all use cases, despite Yo-Ga-X's capabilities of enabling individual federations easily, in order to test the service ecosystem idea of Gaia-X, where subservices are shared. This has not yet beared itself out, as basically all use cases used exactly their own services, only end users were able to consume the main services across multiple use cases. In our estimation, this sharing would have been much more prevalent if the use case hadn't had such a broad spectrum of applications and had there been even more of them within the same federation.

While a single federation approach was worth trying, we also only provided one Trust Anchor, of which many more would have been beneficial anyhow. A Trust Anchor's responsibility to certify the correctness of information in the form of credentials is generally useful, however finding institutions

 doi:10.20944/preprints202410.1403.v1

with the credibility to do so turned out to be challenging. Most of the time, Participants making claims about themselves were also the most trustworthy source of this information, obsoleting additional certification of said claims. Much of the rest of the time, adequate institutions capable of certifying the information in the context of a research project could simply not be determined.

5.2. Different experiences between the use cases

Considering our focus on trust and consent, the related requirements of each use case on Gaia-X were satisfied to different degrees. While some use cases operated fully within a Business-to-Business (B2B) context with highly capable organizations employing the framework to satisfy their need for maintaining trust in the other participants and using it to obtain necessary consent before each service consumption, thus data exchange, other use cases fell short in enabling all target users of their application to even join the federation, let alone exchange data, with the users in question here being individual persons, thus patients, that were initially conceptualized to provide their own personal medical data in this manner. These users had trouble shouldering the technological burden incurred by being full participants in the federation. This was one of our conceptual changes to Gaia-X, and turned out to be inadequate to satisfy these requirements appropriately, calling into question the applicability of a Gaia-X based approach as a whole in this case. Despite falling short in this regard, some kind of conceptual change to Gaia-X is needed, since there is no organization a patient can be solely associated with, as is the case with customer or employees and businesses, patients can't be end users in the same way, as we discussed earlier, when presenting the modified concepts. We hope to contribute to these efforts further in the future, as many projects, not just from the health domain, are doing.

5.3. Reevaluating our focus

While we still consider trust and consent to have been vital, as they are both central requirements to all of our use cases equally, there are other important aspects we also would have liked to focus on, some of them part of Gaia-X, some not so far.

Of the Gaia-X related aspects, we firstly consider the ones associated with the actual exchange of data, like its format and contents, access rules afterwards, some of which Gaia-X calls Data Exchange, Policies and Labels. In this regard, many other implementations, like the EDC, already provide a lot to their users and we would like to see these efforts, as well as implementations from other large projects like Health-X and Team-X and from other domains and extensive implementations like XFSC and GXFS be combined in a common implementation, leading to interoperable, compatible Gaia-X based federations and data spaces, as is the Gaia-X vision. In order for that to take place and be satisfactory however, we deem it necessary for such an implementation to also consider the requirements outlined in this paper and shared by many projects in the health domain and beyond, leading to the concepts of integrating Data Wallets for individual persons independent of organizations.

Aside from technical implementation, we would also have liked to consider more of the institutional aspects of Gaia-X. Concepts like clearing houses for joining and managing federations and institutions for verifying information, to then be used as Trust Anchors, in our estimation seem invaluable to the longevity of the enabled data spaces and should be in focus in the future, yet also seem challenging to define while satisfying all involved parties.

6. Conclusion/Future Work

Principles like data security, sovereignty and control remain current and important, all the more in the context of a european digital infrastructure, with detailed data protection regulation frameworks addressing them. To handle them properly, solutions are needed to which existing systems should adapt. With the Gaia-X specifications proposing many concepts for doing so, the Gaia-X-Med project has been conceived to test their applicability to the medical domain, where these principles play a central role since both research and commercial use of personal patient data represents a tremendous opportunity to improve patient health, which has to be achieved without compromising on the

 doi:10.20944/preprints202410.1403.v1

aforementioned principles for individual patients. Over the course of the project, we were able to either directly use or adapt many of the Gaia-X concepts, focusing particularly on maintaining trust between parties and obtaining and respecting consent for their interactions. We further developed and implemented them via our open source Yo-Ga-X framework, which allowed us to test whether they could satisfy the requirements regarding trust and consent as encountered in the six eHealth use cases of the Gaia-X-Med project.

We found this Gaia-X based approach to adequately satisfy these requirements in use cases with a business-to-business (B2B) focus consisting only of participating organizations, whereas use cases aiming to incorporate individual persons, in order for them to be able to provide their personal patient data, were not able to do so. They mainly fell short to do so due to the incurred technological burden on individual persons, despite our modifications of the base concepts addressing this, which in those cases called the overall approach into question. Therefore, if a Gaia-X based approach is to be pursued in these cases, some kind of further conceptual change is needed as the base concepts are inapplicable, since there is no organization a patient can be solely associated with, like customers or employees are associated with businesses. Therefore, patients do not fit the concept of end users in the same way.

Looking to the future, we propose for the existing Gaia-X related implementations each addressing some common aspects of the Gaia-X concepts to integrate more closely and to strive to enable fully compatible federations and data spaces, as is the Gaia-X vision. In order to do so however, we deem it necessary to also consider the requirements shared by many projects in the health domain and beyond, such as those relating to individual persons controlling their data without being associated to a single organization, which lead to the efforts of integrating Data Wallets for individual persons. In order for the implementations to harmonize, the appropriate way likely is to adapt the related Gaia-X concepts and adopt adequate new ones. While consent management has been addressed conceptually and many seemingly feasible approaches have been demonstrated for organizations with the technical capability, Gaia-X should also focus individual persons in the future via concepts like said Data Wallets, enabling full control on the side of the individual in combination with full responsibility, possibly offering support with official institutions giving recommendations or approval. Accordingly, trustability involving individual persons should be significantly improved by establishing legally robust digital contracts, presumably involving official citizen IDs via the eID service of the German National Identity Card [29], for example. In order to do so, regulation likely needs to improve to enable this, possibly even establishing official state institutions appointed to handling some of the responsibilities in Gaia-X based federations, like clearing houses for management or trust anchors for verifying certain types of information. Additionally, Policies related concepts should be pushed further to ensure certain federation wide constraints on contracts and service consumptions, possibly even to what happens to the data after. In our opinion, retracability would also be very helpful, meaning independently persisting negotiated contracts and service consumptions by a neutral party to assist with inconsistencies, up to and including potential legal disputes.

Whether Gaia-X will adapt in these ways or not, all of us as a community that implement systems for a European data infrastructure in the future internet that is subject to data protection regulation should certainly look to other countries and solutions they are able to implement, which are far more usable and integrated into the daily lives of people more smoothly. These countries have less strict regulations, but provide much better outcomes for all involved parties, especially in the health domain, be it the individual patient, researchers or healthcare providers and companies.

Author Contributions: Conceptualization, all authors; methodology, all authors; software, B.G. and H.H.; validation, B.G. and H.H.; investigation, B.G. and H.H.; writing—original draft preparation, B.G. and H.H.; writing—review and editing, S.F. and M.L.; visualization, B.G. and H.H.; supervision, S.F. and M.L.; project administration, S.F. and M.L.; funding acquisition, S.F. and M.L. All authors have read and agreed to the published version of the manuscript.

---

[29] https://www.personalausweisportal.de/Webs/PA/EN/business/technology/eID-service/eid-service-node.html

Funding: This research was funded by the German State of Schleswig-Holstein, in the funding framework "Künstliche Intelligenz".

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors wish to thank Dirk Schrödter from the government of Schleswig-Holstein for the continuous support of their activities in investigating AI applications in the healthcare domain, their colleagues in the other work packages of Gaia-X-Med for commenting on the solutions developed, and Dr. Raimund Mildner for being a continuous source of inspiration.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lohmöller, J.; Pennekamp, J.; Matzutt, R.; Wehrle, K. On the need for strong sovereignty in data ecosystems. In Proceedings of the DEco@ VLDB, 2022, pp. 51–63.
2. Jussen, I.; Möller, F.; Schweihoff, J.; Gieß, A.; Giussani, G.; Otto, B. Issues in inter-organizational data sharing: Findings from practice and research challenges. *Data & Knowledge Engineering* **2024**, p. 102280.
3. Deshmukh, R.A.; Collarana, D.; Gelhaar, J.; Theissen-Lipp, J.; Lange, C.; Arnold, B.T.; Curry, E.; Decker, S. Challenges and Opportunities for Enabling the Next Generation of Cross-Domain Dataspaces. In Proceedings of the The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, 2024.
4. Hellmeier, M.; Pampus, J.; Qarawlus, H.; Howar, F. Implementing data sovereignty: Requirements & challenges from practice. In Proceedings of the Proceedings of the 18th international conference on availability, reliability and security, 2023, pp. 1–9.
5. Otto, B.; ten Hompel, M.; Wrobel, S. *Designing data spaces: The ecosystem approach to competitive advantage*; Springer Nature, 2022.
6. Lang, S.; Kneuper, R. Datenschutz und Informationssicherheit in Gaia-X. *Datenschutz und Datensicherheit-DuD* **2022**, *46*, 778–781.
7. Otto, B.; Jarke, M. Designing a multi-sided data platform: findings from the International Data Spaces case. *Electronic Markets* **2019**, *29*, 561–580.
8. Pampus, J.; Jahnke, B.F.; Quensel, R. Evolving data space technologies: lessons learned from an IDS connector reference implementation. In Proceedings of the International Symposium on Leveraging Applications of Formal Methods. Springer, 2022, pp. 366–381.
9. Mügge, J.; Grosse Erdmann, J.; Riedelsheimer, T.; Manoury, M.M.; Smolka, S.O.; Wichmann, S.; Lindow, K. Empowering end-of-life vehicle decision making with cross-company data exchange and data sovereignty via Catena-X. *Sustainability* **2023**, *15*, 7187.
10. Lopes, P.M.; Guimarães, P.; Pereira, T.F.; Machado, R.J. Gaia-X & Fiware: Implementation of a Federated Data Platform in Smart Cities. *Procedia Computer Science* **2024**, *239*, 1506–1515.
11. Bosse, S.; Berns, K.; Bosch, J.; Dörr, J.; Eichhorn, F.C.; Eisert, P.; Fischer, C.; Gassen, E.; Gerstenberger, M.; Gerighausen, H.; et al. Nachhaltige Landwirtschaft mittels Künstlicher Intelligenz – ein plattformbasierter Ansatz für Forschung und Industrie. In *43. GIL-Jahrestagung, Resiliente Agri-Food-Systeme*; Gesellschaft für Informatik e.V.: Bonn, 2023; pp. 41–52.
12. Eclipse Foundation. Eclipse XFSC (Cross Federation Services Components), 2024. Available online: https://projects.eclipse.org/projects/technology.xfsc (accessed on 30-09-2024).
13. Eclipse Foundation. Eclipse Dataspace Components, 2024. Available online: https://projects.eclipse.org/projects/technology.edc (accessed on 30-09-2024).
14. Maia, M.; Jaberansary, M.; Ucer, Y.; Beyan, O.; Kirsten, T. Providing Publicly Available Medical Data Access under FAIR Principles for Distributed Analysis. In Proceedings of the 67. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS), 13. Jahreskongress der Technologie- und Methodenplattform für die vernetzte medizinische Forschung e. V. (TMF), 2022. https://doi.org/10.3205/22gmds071.
15. Kari, A.; Schurig, T.; Gersch, M. European Health Data Space (EHDS), Gaia-X and Health-X dataLOFT, 2023. https://doi.org/http://dx.doi.org/10.17169/refubium-41224.

16. Toma, M.; Bönisch, C.; Löhnhardt, B.; Kelm, M.; Bohnenberger, H.; Winkelmann, S.; Ströbel, P.; Kesztyüs, T. Research collaboration data platform ensuring general data protection. *Scientific Reports* **2024**, *14*, 11887.
17. Arshad, K.; Ardalan, S.; Schreiweis, B.; Bergh, B. Leveraging Clinical Data Treasures: Integration of an AI Platform into Clinical IT. In *Digital Health and Informatics Innovations for Sustainable Health Care Systems*; IOS Press, 2024; pp. 1169–1173.
18. Büchel, J.; Engels, B. Branchentrends beim Data Sharing: Status quo und Use Cases in Deutschland. Technical report, IW-Report, 2022.
19. Boll, S.; Meyer, J. Health-X dataLOFT: A sovereign federated cloud for personalized health care services. *IEEE MultiMedia* **2022**, *29*, 136–140.
20. Meier, J.J.; Hermsen, K.; Bauer, J.; Eskofier, B.M. Digital Responsibility Goals–A Framework for a Human-Centered Sustainable Digital Economy with a Focus on Trusted Digital Solutions. In *dHealth 2022*; IOS Press, 2022; pp. 250–259.
21. Gaia-X Association. *About Gaia-X - Gaia-X: A Federated Secure Data Infrastructure*, 2024. Available online: http://docs.gaia-x.eu/technical-committee/data-exchange/22.10/dewg/ (accessed on 30-09-2024).
22. Gaia-X Association. About Gaia-X - Gaia-X: A Federated Secure Data Infrastructure, 2024. Available online: https://gaia-x.eu/what-is-gaia-x/about-gaia-x/(accessed on 30-09-2024).
23. Gaia-X Association. *Gaia-X Framework*, 2024. Available online: https://docs.gaia-x.eu/framework/ (accessed on 30-09-2024).
24. Gaia-X Association. *Gaia-X Architecture Document - 22.04 Release*, 2022. Available online: https://docs.gaia-x.eu/technical-committee/architecture-document/22.04/ (accessed on 30-09-2024).
25. Gaia-X Association. *Gaia-X Trust Framework - 22.04 Release*, 2022. Available online: https://docs.gaia-x.eu/policy-rules-committee/trust-framework/22.04/ (accessed on 30-09-2024).
26. Sporny, M.; Longley, D.; Chadwick, D. *Verifiable Credentials Data Model v1.1*. W3C, 2022. Available online: https://www.w3.org/TR/vc-data-model/ (accessed on 30-09-2024).
27. Sporny, M.; Longley, D.; Sabadello, M.; Reed, D.; Steele, O.; Allen, C. *Decentralized Identifiers (DIDs) v1.0*. W3C, 2022. Available online: https://www.w3.org/TR/did-core/ (accessed on 30-09-2024).
28. European Parliament.; Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council (GDPR).
29. acatech. Herausforderungen und Lösungsansätze der Gaia-X Förderprojekte (section 2.3), 2023. Available online: https://gaia-x-hub.de/wp-content/uploads/2023/04/WP-Stand-Foerdervorhaben-Gaia-X.pdf (accessed on 30-09-2024).
30. Gribneau, C.; Prorock, M.; Steele, O.; Terbu, O.; Xu, M.; Zagidulin, D. *did:web Method Specification*. W3C Credentials Community Group, 2024. Available online: https://w3c-ccg.github.io/did-method-web/ (accessed on 30-09-2024).
31. OpenID Foundation. How OpenID Connect Works, 2024. Available online: https://openid.net/developers/how-connect-works/ (accessed on 30-09-2024).