Gaia-X Domain Health

Position Paper Version 1.0 2021

Contents

Contributors

Authors:

The document is co-authored by a large group of contributors from the GAIA-X Health community. Thank you all for your contributions!

Editors:

Jaana Sinipuro (Sitra) and Bert Verdonck (Philips).

Note:

this is version 1.0 of this position paper, which is not a definite nor final version of this document. Please consider it as a baseline for further discussion. We explicitly open this version for further comments from many more stakeholders in the healthcare domain. We anticipate to include further insights in a version 2.0 with a target deadline of June 20, 2021.

Executive summary

Gaia-X[1] creates the foundation for a sovereign, federated, open data infrastructure based on European values. This document proposes a collaboration between health domain stakeholders to build out a Health Data Space using the Gaia-X framework and aligned with the European Health Data Space initiative.

The initiative brings technical and semantic interoperability that is essential to unlock the power of health data. It takes the burden away to build a trustworthy and compliant data service stack, to enable the scale that is essential for research and innovation to thrive within Europe. Note: Gaia-X focuses on standards and technical frameworks: this initiative will discuss but not solve all the organizational and legal challenges related to this ambition.

A Health Data Space can be composed from a rich set of shared capabilities on top of foundational cloud infrastructures. These allow creating a number of shared and federated health data spaces, where data is granularly and selectively accessible in line with Europe's privacy provisions and other applicable laws. This implies that federated data spaces will exist on regional, national and European level.

The Health Data Space should contribute to the care delivery processes (primary use of data) for the individual patient or resident as well as to secondary use of data on a cohort or population scale. The data space enables a data value chain between data holders and data users, across the broad and complicated health domain ecosystem.

This document invites for collaboration with the existing health data initiatives and for deploying a number of concrete and valuable use cases. We describe 5 archetypical use cases and reference a large set of additional candidates. These examples illustrate the essential and common enablers that become the basis for a development roadmap. We propose an initial "Hello World" use case to validate the Gaia-X concept and architecture in a relevant, end-to-end and real-world implementation before the end of 2021.

This paper sets a baseline for further discussion: with the Gaia-X technical community to validate the standard and architecture proposals, with the health domain stakeholders to adopt the learning from existing data initiatives, with additional use case owners, with contributors of requirement specifications and solution components, and with funding partners to support the implementation of these ambitious plans.

Thank you for your support on our mission to access and to share health related data securely and confidently for the benefit of our European patients, citizens, residents and societies.

---

[1] Gaia-X: a federated data infrastructure for Europe: https://www.data-infrastructure.eu/

Mission and Goals of the Gaia-X Health Data Space

Gaia-X creates the foundation for a federated, open data infrastructure based on European values, a credible basis for the Health Data Space. No single country, no single company can master this challenge. This document proposes a collaboration between health sector stakeholders to build out a Health Data Space using the Gaia-X infrastructural framework. The Gaia-X Health Data Space initiative holds the potential of achieving the scale required to access and to share health data securely and confidently, governed and controlled by each Member State in a consistent manner.

Note: we use "data space" in the singular form throughout this document, not assuming there will be one single, monolithic "space" but a large collection of new as well as existing data sharing and data collection initiatives. We anticipated there will be multiple data spaces, all connected in a federated approach. We use "ecosystem" for the large and rich set of stakeholders in the health sector. This ecosystem will use multiple data spaces to manage and exchange health-related data between stakeholders.

The initiative brings technical and semantic interoperability that is essential to unlock the power of health data (refer to European Interoperability Framework[2]: the focus is on technical and semantic interoperability, and legal and organizational interoperability is out of scope in this paper). It takes the burden away to build a trustworthy and compliant data service stack, to enable the scale that is essential for research and innovation to thrive within Europe. And it brings a framework compliant with Europe's privacy provisions and other applicable laws to unlock data for primary and secondary use that is secure yet easy to use.

Mission: access and share health related data securely and confidently.

Goals

- Provide the means to link currently isolated data and disparate applications, between citizens, care providers and other stakeholders, within countries and across borders, in a transparent manner, adhering to international interoperability standards.

- Provide a framework to implement the Health Data Space at scale, in a compliant, secure, and trustable manner.

- Enable the storage and access of personal and non-personal health information in trusted and collaborative cloud infrastructures, with elasticity to scale and with a proper legal basis (consent, anonymization, etc.).

- Implement clear governance for the use of data on personal, regional, national and European level and for delivery of care, for research, for commercial and governmental use. Keep control with the citizen at all times, only except when the data is used for the public good according to GDPR[3] or national legislation.

---

[2] European Interoperability Framework (EIF): https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

[3] General Data Protection Regulation (GDPR) article 6: lawfulness of processing: https://gdpr-info.eu/art-6-gdpr/

Challenges addressed

The challenge of managing **chronic disease** burden and the care of an **ageing population** is daunting. The health sector is lagging in digitization and is missing the integrated and longitudinal views on a single patient and on patient populations. The health ecosystem is **complex** with many stakeholders, many regulated processes and many sources of public funding, which **slow down** transformation and innovation. The COVID-19 **pandemic** has exposed these issues even further and has demonstrated the fragility of healthcare systems worldwide.

Most patient data is collected and stored by **disparate systems** that do not necessarily interoperate: Practitioners have difficulty to access and exchange information; medical decisions cannot benefit from a patient centric, longitudinal view and these decisions are not optimized on a regional, national, let alone a European scale. Access to health-related data is distributed between many stakeholders such as health care providers, insurances, but also companies in a secondary health market and individual patients. And citizens have no clear view and control on their health-related data. Tomorrow's healthcare requires data from various sources to be available, combined as well as processed, while meeting the highest standards of privacy and security and keeping transparency and control with the citizen.

European healthcare systems should ensure that all citizens and residents have (remote) access to essential, high quality, affordable healthcare services, in line with the United Nation's Sustainable Development Goals. The simple, performant, secure and affordable exchange of personal health information is an essential enabler for those goals.

The current **highly fragmented and heterogeneous EU market** limits the ability to roll out digital health innovation at any sufficient scale. Numerous ambitious data driven initiatives exist across Member States. However, few to none of these have ever achieve scale with societal impact. EU privacy legislation (GDPR) is world-leading, however stakeholders in the health domain struggle with interpretation and with local legislative variations. The Gaia-X initiative cannot resolve these challenges on its own, but it can advocate these with policy makers.

Without a more open market for health data in Europe, innovative companies are forced to focus on China and the USA as their lead markets, because that is the only way to achieve scale. The current **complexity** to deploy a data management solution *across* European borders limits the ability to aggregate health data for science and research (for innovation stakeholders). A solution for patients and their care providers that works in one country is difficult or expensive to deploy in another country (for care delivery stakeholders). Without a more seamless access to health data, both academic and commercial R&D are forced to turn to other geographies to accumulate data for innovation and validation, and they will be hesitant to deploy solutions to smaller countries.

The recovery plan of the EU[4] calls for a green, inclusive, resilient and digital recovery. The European Health Data Space is an essential part of this strategy according to the European

---

[4] EU recovery plan: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_940

Commission[5] and the European Parliament policy documents[6] plead strongly for the digital sovereignty of European citizens. The funding priorities in the Digital Europe Programme[7] indicate the same: *1) to ensure that citizens have control over their personal data across borders, 2) to make available better data for research, disease prevention and personalized health and care, and 3) to make digital tools available for citizen empowerment and for person-centered care.* A common, federated European Health Data Space will foster better exchange and provide access to different types of health data (e.g. electronic health records, genomics data, data from patient registries, patient reported outcomes data, clinical trial data, epidemiological vigilance data, and more).

The Health Data Space should contribute to the care delivery processes (primary use of data) for the individual patient or resident (for healthy living, prevention, diagnosis, treatment and homecare, leading towards value-based healthcare) as well as to secondary use of data on a cohort or population scale (for research, innovation, crisis preparedness and crisis management, public and population health). Today, there is an enormous duplication of efforts in the management of data in primary versus secondary use cases, whereas the technological enablers are largely the same (e.g. to deploy and maintain a secure and scalable data lake, to register patient consent, to transform free text notes into standardized and structured data). Achieving interoperability among these systems is critical for clinical care as well as clinical research. This opens up a market for virtual healthcare services (e.g. e-consultations, e-interventions, telehealth, tele-radiology, remote care management, and other aspects of tele-medicine), as well as for digital health science using both trial- and real-world-data (e.g. clinical data trials, trustworthy and ethical artificial intelligence).

The Health Data Space needs to respect the sovereignty of Member States in the healthcare domain and to allow full control on the data exchange across borders with the Member States and their citizens. Current solutions are either **too permissive** (e.g. exposing data to the public domain or transferring data usage rights to a single commercial company) or **too restrictive** (e.g. study-specific point solutions or local-for-local solutions without opportunities for reuse).

Gaia-X develops a standard based, technical framework[8] to implement distributed health data systems in all European countries in a legally secure manner and enabling compliance with GDPR and other health data regulations.

Finally, there continues to be too much **variation in the adoption of open, international standards** for personal and clinical health data (silo-ed adoption and excessive local variation). Many international and widely accepted standards exist (IHE, HL7 FHIR, DICOM, SNOMED, LOINC, etc.), but they are deployed at various levels and with many local variations. The level of interoperability (technical, semantic, organizational) is expected to improve when people can collect and exchange personal health data on a much larger and cross-national scale.

---

[5] EC strategy for the European Health Data Space: https://ec.europa.eu/info/sites/info/files/soteu_2020_en.pdf

[6] EU parlement policy document: https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf

[7] Digital Europe Programme: https://data.consilium.europa.eu/doc/document/ST-6789-2020-INIT/en/pdf

[8] GAIA-X concept, standards and technical architecture: https://www.data-infrastructure.eu/

Solution: Data space description in a holistic view – detailed view on the endeavor

Partners in the health domain

The health domain comprises of four essential contributors:

1. Data holders & data users

2. Application and service providers

3. Data space governance and operating entities

4. Cloud service providers

Data holders & data users

The end users of the healthcare applications and services are the holders (providers) and the users (consumers) of personal health information[9]:

- Citizens/residents/patients who are the subject of prevention and care
- Informal caregivers: parents for their children, partners, caregivers for the elderly or for the disabled, etc.
- Care providers: therapists, nurses, general practitioners, specialists, etc. and their organizations and associations
- Payers and insurers
- Industry: pharmaceutical, medical device, healthy living and prevention, digital content and therapeutics, administrative and financial services, etc.
- Academic and research institutions: universities, institutes, etc.
- Government agencies, non-governmental organizations or charities

All of these end user stakeholders will benefit from data-driven applications and services built on a trusted, safe and secure cloud infrastructure.

Some of these stakeholders will become Gaia-X participants when they choose to make use of the Gaia-X cloud and data services in order to provide services to specific end users.

The breadth of stakeholders and application areas is illustrated in Figure 1**Error! Reference source not found.**, with a non-limiting overview of data holders and users within the major

---

[9] We use the terms data holder and data user as defined in the EU Data Governance Act: 'data holder' means a legal person or data subject who, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal or non-personal data under its control; 'data user' means a natural or legal person who has lawful access to certain personal or non-personal data and is authorised to use that data for commercial or non-commercial purposes. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767

application areas. These stakeholders are active in the research and development phase, they are supporting healthy living, prevention, diagnosis treatment, and they manage healthcare on a population and societal level.
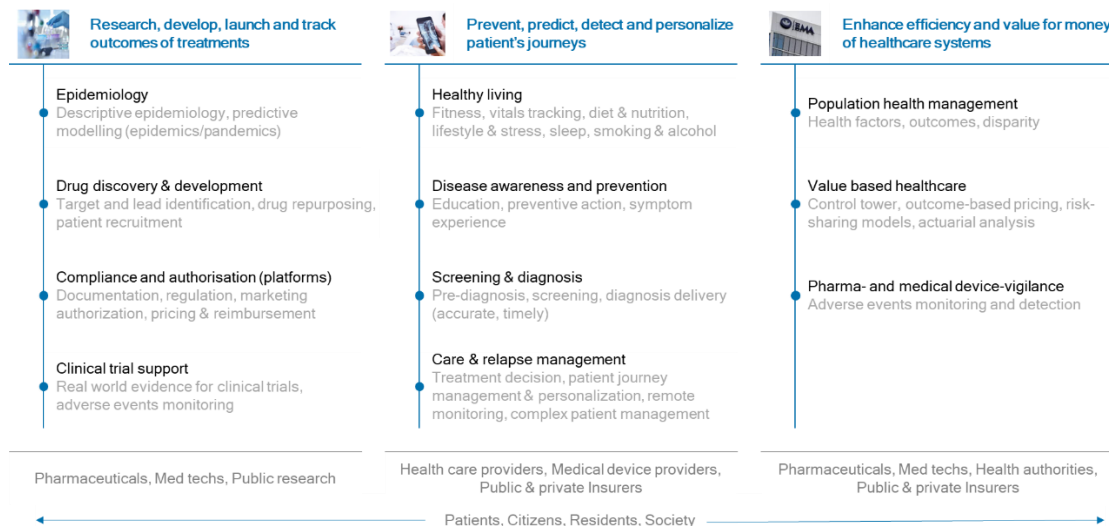

Research, develop, launch and track outcomes of treatments	Prevent, predict, detect and personalize patient's journeys	Enhance efficiency and value for money of healthcare systems
Epidemiology Descriptive epidemiology, predictive modelling (epidemics/pandemics)	**Healthy living** Fitness, vitals tracking, diet & nutrition, lifestyle & stress, sleep, smoking & alcohol	**Population health management** Health factors, outcomes, disparity
Drug discovery & development Target and lead identification, drug repurposing, patient recruitment	**Disease awareness and prevention** Education, preventive action, symptom experience	**Value based healthcare** Control tower, outcome-based pricing, risk-sharing models, actuarial analysis
Compliance and authorisation (platforms) Documentation, regulation, marketing authorization, pricing & reimbursement	**Screening & diagnosis** Pre-diagnosis, screening, diagnosis delivery (accurate, timely)	**Pharma- and medical device-vigilance** Adverse events monitoring and detection
Clinical trial support Real world evidence for clinical trials, adverse events monitoring	**Care & relapse management** Treatment decision, patient journey management & personalization, remote monitoring, complex patient management	
Pharmaceuticals, Med techs, Public research	Health care providers, Medical device providers, Public & private Insurers	Pharmaceuticals, Med techs, Health authorities, Public & private Insurers

Patients, Citizens, Residents, Society

Figure 1: the health domain stakeholder landscape

Application and services providers

The providers of data management and data processing applications and services include creators and operators of such applications and services for a specific target audience, a group of users, a cohort or a population:

- Application software providers
- Data service providers
- Data and service catalogue providers

The Gaia-X Health Data Space enables an equal playing field for application and service developers, taking the burden away of building a trustworthy and compliant data services stack. Small, medium and large companies, together with governmental and non-governmental institutes can take the lead in developing an agile, innovative and front-running market that has the potential to transform healthcare in Europe, and to set an example globally. The value created from access to European health data is significant and ranges from developing new applications, to training AI algorithms, to providing better patient experience, to realizing operational efficiency gains. Seamless access to data also facilitates new business models: Platform, Software, Analytics, Data, and more *as-a-service* models. It will enable European players to scale, and ultimately, to compete globally.

The future landscape for innovative applications is incredibly rich and may include: life style improvement using various tracking devices and (non-)human coaching, symptom and vital sign tracking, continuous outcomes measurement, real-world-data trending and analysis, decision and referral guides, appointment scheduling, clinical data trial management,

advanced image and genomic data analysis, health resource and capacity optimization, emergency and triage management, and many more.

Data Space governance and operating entities

The Health domain has many requirements: some are essential but not sector specific (electronic identification, data access, permission and consent management, auditing, etc.), others are essential and healthcare specific (clinical coding systems and terminology servers, clinically semantic interoperability, clinical data and image standards, anonymization of personal health data, etc.). All data applications and services require a number of these common standards and common components. Therefore, two different actors need to participate in each data space initiative:

- Data governance: the entities defining the governance and the rules of the data space, the common data standards and the minimal requirements for the secure storage, processing and exchange of personal health information. This will typically be a combination of governmental, non-governmental and industry actors issuing laws, guidelines, standards and code of conduct. The resulting framework of standardized agreements could be a digital 'soft' infrastructure[10] or it could be established by 'hard' law[11]. This framework enables semantic interoperability, it guides the ethical use of data and controls the commercial or non-commercial exploitation of such data, it enables data sovereignty and a level playing field for data sharing and exchange. The Gaia-X Health Data Space initiative intends to co-create such framework with the Member States and with the European Commission.

- Operating entities: they operate within the soft infrastructure guidelines and they run the mechanics of the data storage, access management, data processing and exchange. They run the "connectors", linking currently isolated data in a transparent, secure and audited way, and providing the services that enable data consumption by authorized consumers. A number of industry consortia can build out such operating entities at scale.

A Health Data Space is composed from a rich set of shared capabilities on top of foundational cloud infrastructures, where data is granularly and selectively accessible in line with Europe's privacy provisions and other applicable laws. This implies that federated data spaces will exist on regional, national and EU level. In essence, the Health Data Space enables digital patient rights management, and last but not least data standardization and data normalization.

Cloud service providers

---

[10] Soft infrastructure introduced by Innopay & Sitra: https://www.innopay.com/en/media/616/download

[11] European health data space – proposal for a regulation: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12663-Digital-health-data-and-services-the-European-health-data-space

A Health Data Space is created on a foundation of cloud services and edge components. Gaia-X will create an ecosystem of trusted, safe and secure cloud and edge infrastructures in Europe to allow health data to flow securely and in line with Europe's privacy provisions and values. The federated cloud and edge ecosystem will enable the creation of a competitive marketplace for cloud services while avoiding dominant vendor 'lock-in'. European cloud users need freedom-of-choice to select local or international providers, depending on the use case, in a transparent way, with documented and certified compliance to Gaia-X policies and standards.

The data value chain

All stakeholders in the health domain interact in a broad and complex ecosystem. Figure 2 illustrates the consolidated data value chain in a desired target state. The data holders create and aggregate data at the left of the figure. The data users consume data at the right, in line with the access permissions they have been granted. In most situations the data holders are data users at the same time, and they combine data to improve the services they provide. The central part of the figure lists the common enablers of a Health Data Space, the core of the Gaia-X federated services architecture.
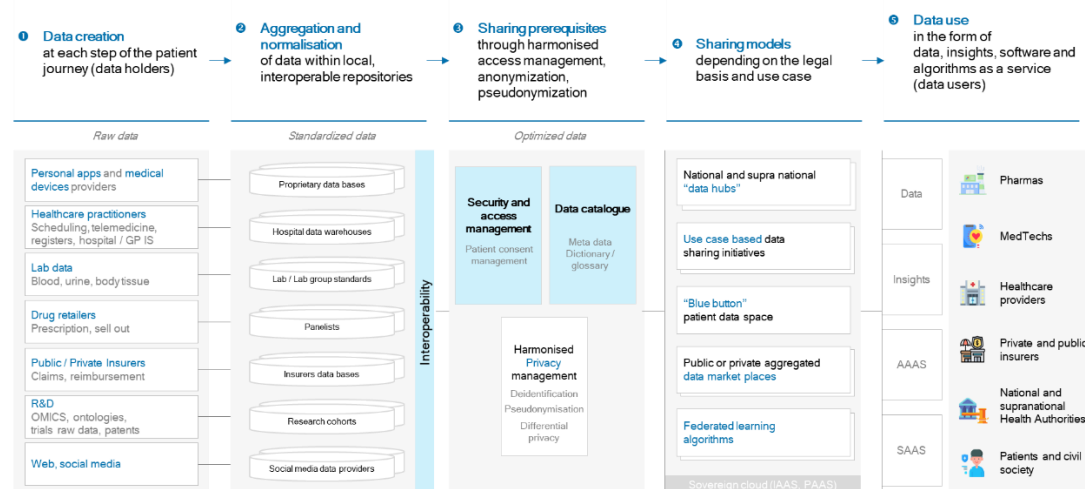
Figure 2: the data value chain in a desired stateUse-cases (scenarios) within the data space and the current state of development

The health sector is a complex landscape with many stakeholders. National and European authorities regulate the sector strictly, to ensure the quality, safety and effectiveness of care delivery as well as the protection of the privacy of European citizens. European countries provide healthcare through public funds and finance them largely through private and statutory health insurance. Most of these healthcare policies are established at the level of the Member States, with Europe complementing these national policies to achieve common objectives (e.g. EMA[12], ECDC[13], EHDS[14]).

---

[12] European Medicines Agency (EMA): https://www.ema.europa.eu/en
[13] European Centre for Disease Prevention and Control (ECDC): https://www.ecdc.europa.eu/en
[14] European Health Data Space (EHDS): https://ec.europa.eu/health/ehealth/dataspace_en

In this context, it is no surprise that health data initiatives exist in many countries with many different variations. There cannot be one single data space for the entire health sector across all Member States. This initiative can strive towards a better integration of existing and future data space initiatives, towards coordination across a federated landscape, and towards a smarter reuse of concepts, infrastructure and common components.

For the scope of this positioning paper we have described a small, representative set of health sector use cases. Many more use cases are available, and some have been well documented in the Gaia-X use case gallery. An overview with a short description of these is summarized in appendix A in this paper.

Major European initiatives are active in the Gaia-X-Health user group. With the meanwhile 22 use cases with content related to healthcare, research and the healthcare industry, the diversity of the healthcare domain is already covered to a large extent. For the basic development of the Gaia-X infrastructure, the domain Health has been significantly involved in all technical working groups from the start of Gaia-X, to work out the basic requirements and possible solutions for the complex and federated health data domain.

These cases are similarly valuable and each deserves a chance for further development. However, in order to keep this paper focused on the most common enablers and characteristics, we decided on the current subset. Further use cases will be added to the roadmap while the initiative makes progress.

For the five use cases described in this positioning paper we call for volunteers to join these initiatives and to extend the functionality with additional enabling capabilities. The initiative intends to onboard the learnings from past or current other initiatives, to combine existing efforts and to reduce duplication of efforts.

Description Use Case 0: Hello Health!

Solution

This is the "Hello World" application for our Gaia-X Health Data Space, i.e. the most basic feature set that demonstrates the functioning of the data space, end-to-end. "Hello Health" is a template service that demonstrates the basic capabilities of the data space, to serve as a training environment and test bed, and to reduce the adoption thresholds for new stakeholders of this initiative.

The initial solution scope could be as follows:

- Focus on people's body weight curves, and a small set of characterizing attributes like height, gender, age, social context and living location, i.e. a very simple yet relevant set of personal health information
- Allow an individual to register for participation and to make a personal weight table available to the data space (weight in function of time plus a basic set of context attributes: height, gender, age and living location) – deploy this as Gaia-X federated service

- Allow an organization to register for participation and to contribute weight curves of a group of individuals to the data space (similar data as from individuals) – deploy this as Gaia-X federated service
- Manage consent with the data contributors; provide the legal document templates for use (terms of use, privacy policy)
- Pre-processing: anonymization of this data (or restrict access to personal identifiers) and calculate BMI from weight/height, calculate age from birthdate, derive country/region identifiers from address
- Make the data accessible as Gaia-X data assets via a Gaia-X federated data catalogue
- Enable queries of these aggregated weight/BMI curves, with filters on time, gender, age, social context and location, for cohorts large enough to prevent individual re-identification.
- Feedback population statistics to the contributors and show where their individual values fit into the cohort statistics – provide value in return for participation. Provide an API for additional services to hook into this initial basic use case (e.g. provide healthy lifestyle services to specific target cohorts).
- Deploy these services via a Gaia-X federated services catalogue

Problem solved

Provide a working example. Prove that it works across multiple EU countries. Reduce reticence of participants.

There is potential for some side-problems to get solved: invite ecosystems partners to build extensions upon this template service. Use it as a basis for user experience testing. Etc.

Partners/Ecosystem

To be defined among Gaia-X Health user group participants. A small group of volunteers will first refine the exact scope of "Hello Health". Then break it down into components and assign an implementation volunteer per component. Bring a trial implementation to production and search for voluntary contributors (individuals and organizations) to publish their weight curve to the service. Publish statistical results as soon as a minimal sample size is reached.

Main technology/Gaia-X components

This demonstrator would require all of the Gaia-X infrastructural components:

1. Identity & Trust: federated identity management for individuals and organizations

2. Federated Catalogue: to publish the registration, consent and query services

3. Sovereign Data Exchange: to manage consent and usage control

4. Compliance: rights management, onboarding and certification

And it would require very basic health data standards:

- (could be) structured FHIR definitions[15] for: BodyWeight, Gender, Birthdate, Address
- (could be) SNOMED CT coding for semantic concepts[16]
- (could be) cross-sectoral standards for non-healthcare specific concepts[17]

Wherever possible this initiative reuses the concepts and components of the European Health Data Space and more specifically of the eHealth Digital Service Infrastructure (eHDSI)18.

Concrete benefits

Get a demonstrator off the ground quickly. Solve all initial issues with applying Gaia-X concepts and architecture for a real-world health application. Thereafter, use this demonstrator for training and onboarding of new stakeholders.

This could become an early showcase of the iCitizen work package of the TEHDAS Joint Action[19] which seeks to obtain a better understanding of citizens' relationship with health data in the EU, to better inform and sensitize citizens regarding health data and recommends data altruism practices for the EHDS.

Description Use Case 1: Research Platform Genomics

Solution

This solution implements a cloud based genomics platform, compliant with the Global Alliance for Genomics & Health (GA4GH[20]), for the storage and analysis of genome data for medical research. Proven gold standard analysis methods are a key component of a comprehensive solution that is not intended to be established in just one location but can be rolled out and deployed at other cloud locations without big efforts. This is of particular importance, for example, when sensitive data is not allowed to leave the clinic context and an identical workflow for comparable results needs to be ensured. In addition, the platform aims to lead to better prediction of cancer and development of new and improved treatment methods based on advanced data analytics and machine learning.

Problem solved

The following problems can be solved by the platform described:

---

[15] HL7 FHIR standard: https://www.hl7.org/fhir/
[16] SNOMED CT coding: https://www.snomed.org/snomed-ct/five-step-briefing
[17] Example standard for addressing: ISO 19160: https://www.iso.org/obp/ui/#iso:std:iso:19160:-1:ed-1:v1:en
[18] eHDSI: https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHealth+DSI+Operations+Home
[19] TEHDAS Joint Action Towards the European Health Data Space: https://tehdas.eu/
[20] Global Alliance for Genomics & Health: https://www.ga4gh.org/

- provides the required technical infrastructure that both enables secure storage of large data sets and powerful compute architecture and methods for the complex analysis of data at petabyte scale.
- Ensuring that the access is secure and GDPR compliant, with datasets, analytics pipelines, and compute provided by different actors in the health care and research sectors.
- Making available extensive and published datasets in the context of cancer research through secured interfaces.
- Automated deployment and roll-out capability of the interoperable and virtualized methods of other secure top locations in the healthcare sector

Partners/Ecosystem

First examples of existing services and compute services from the de.NBI cloud[21] together with other volunteering partners of the health domain could be used to provide a proof of concept. This first approach is then to be concretized and appended as other providers of the health domain join the platform.

Main technology/Gaia-X components

This demonstrator would require all of the Gaia-X infrastructural components:

1. Identity & Trust: federated identity management for individuals and organizations

2. Federated Catalogue: to publish the registration, consent and query services

3. Sovereign Data Exchange: to manage consent and usage control

4. Compliance: rights management, onboarding and certification

The Federated Catalogue, Data Connector, Identity and Access Management, Self-Description, Standards should be compliant with GA4GH and based on the Task Execution and Workflow Execution Services technology. The deployment could be on OpenStack. It would require data standards for genetic data (i.e. WGS, BAM, FASTQ and others).

Concrete benefits

- The project enables secure and GDPR-compliant access to data of different actors in the healthcare system through compliance with Gaia-X provided policies.
- The ability to integrate providers of powerful (infrastructure) components and high-performance computing and analysis functions via the Gaia-X federation services as well as their accessibility for different users, promises time, cost and efficiency advantages through the use of scaling effects.
- The openness and the resulting flexibility enable the connection of existing (data) platforms to European research and health domains and international initiatives.

---

[21] de.NBI cloud: https://www.denbi.de/cloud

- Standardization of pipeline development based on the Global Alliance for Genomics and Health (GA4GH) and use in European initiatives such as ELIXIR and EOSC-LIFE will lead to larger international acceptance and sustainability of the platform.
- The simple and proven access will enable even greater use of European academic clouds in the future as the host for the platform within the framework of EHDS.
- With its ability to integrate data across individual domains (e.g. image data, clinical information), the Gaia-X network offers the potential to realize more complex integrative analyses within personalized medicine for the benefit of patients.
- Within the federated approach, data exchange and integrative analysis with other omics platforms will be enabled, such as the German Human Genome-Phenome Archive (GHGA)[22] as a secure and trusted long-term archive for human omics data.
- The developed GA4GH compliant analysis pipelines will be used in the Beyond 1 Million Genomes[23] initiative, the pan-European network of genetic and clinical data. Accredited users will be able to use the rolled-out analysis environments in the respective local databases involved for the secure analysis of genomic data.

Description Use Case 2: Image Archiving and Image Sharing for Medical Professionals and the Citizen

Solution

The solution implements the sharing of medical images across a healthcare enterprise and with the citizen, based on an IHE (Integrating the Healthcare Enterprise)[24] profiles known as Cross Enterprise Document Sharing for Images (XDS-i) and Cross Community Access (XCA)[25]. This profile is one component of the larger IHE XDS interoperability profile that describes how to establish true interoperability across healthcare enterprises when exchanging patient information. The solution includes the means to store and aggregate medical images in the cloud for long term storage in cross-institutional data lakes. And it includes provisions to train and to deploy AI algorithms across institutions (federated learning and confidential compute). The result is an approach to cross-enterprise image sharing that brings the right data to the right people at the right time, to support the care process for the individual patient as well as the collection of medical images for secondary use.

Problem solved

Today a mixture of formats is used to share medical images (film, prints, DVD, Email, software portals, messaging services). Healthcare professionals do not have easy nor quick access to medical images of a patient, especially when these are created at another clinic or enterprise. Patients/citizens do not have a single means to access their medical images, independent of

---

[22] The German Human Genome-Phenome Archive: https://ghga.dkfz.de/
[23] Beyond 1 Million Genomes: https://b1mg-project.eu/
[24] Integrating the Healthcare Enterprise: https://www.ihe.net/
[25] IHE recipe for securely sharing health information: https://www.ihe.net/news/ihe-hie-white-paper/

the location where they were made. Researchers do not have an easy/affordable way to collect large sets of digital images for clinical trials, to train and to execute AI algorithms and for research. And authorities lack the infrastructure to deploy large scale image-based screening initiatives.

This capability is expected to reduce the amount of unnecessary images (because a previous image was not known or not available at the point of care), to reduce costs in image handling and image analysis, and to give the citizen control on their medical images and their destiny (e.g. for 2nd opinion, when switching treating physician or when traveling across Europe).

Partners/Ecosystem

The Use Case involves the medical image storage and archive solutions that store the often large images at the source. These are typically maintained by the IT organizations of clinical organizations (radiology clinics, hospitals, etc.), on premise or in a cloud. Image sharing across a region requires a central infrastructure with a central registry and a retrieval/viewing functionality, maintained by an IT organization of a regional entity. The required IHE components are available from several IT vendors, to be implemented by an implementation team. This team works together closely with IT specialists at the local and regional entities.

The ambition of this Use Case is to connect these image exchange infrastructures at a supra-regional or supra-national scale, deploying on the federated Gaia-X infrastructure. Existing XDS-i infrastructures are candidates for adoption as nodes of the Gaia-X infrastructure. To realize this, Philips volunteers to take a coordinating role between a selection of existing XDS-i infrastructures, their owners and respective suppliers, in order to realize a breakthrough in scale and availability. We call for other volunteers to join the initiative and to extend the functionality with additional enabling capabilities (e.g. patient consent management, image anonymization, decentral image analysis services, and more).

An alignment is required with the cross-border sharing of imaging on the roadmap of the European Health Data Space. This use case extends upon that initiative.

Main technology/Gaia-X components

This use case relies on the IHE XDS-i and XCA standards. A first implementation can use existing XDS-i implementations and/or implement one or more additional XDS-i implementations and connect them to one or more patient portals. The standards exist to combine multiple XDS-i infrastructures: cross community access (XCA). The main tasks during implementation at scale include:

- Develop a standard deployment model of XDS-i and XCA components on the Gaia-X infrastructure
- Develop a standard implementation package (software/services) to support easy adoption by image storage and archive systems
- Coordinate the integration of existing XDS-i infrastructures

- Coordinate the targeted cloud deployment of new image sharing and image storage infrastructures (image data lakes)
- Deploy the capabilities for decentral execution of AI algorithms on this infrastructure

The deployment requires identity and authentication management (IAM) for patients and providers, and a provider directory.

Concrete benefits

Concrete benefits include:

- Provide an imaging component to the Electronic Health Record or Patient Health Record: the shared imaging record, in a community, region, etc.
- Effective means to contribute and access imaging documents across health enterprises, for clinical care, for research as well as public health
- Enable sharing imaging documents between radiology or surgical departments, private physicians, clinics, long term care, acute care with different clinical IT systems
- Care providers are offered means to query and retrieve imaging documents (images and reports) of interest using the same mechanisms used to query other documents
- Provide access to images for the patient/citizen from a single overview, even when images were created at different locations
- A unified approach to support patient rights (access to images), clinical care (images at the point of care) as well as research (image collection).

Description Use Case 3: Smart Health Connect

Solution

The combination of health data from primary and secondary health care across a large population is the ultimate goal of this use case. Citizens and patients will be able to add their personal data directly to the new Medical Data Space that is built by this initiative, while keeping control over it through deeply integrated consent management mechanisms and user centric data storage. Approaches in the primary health care market will follow strictly the regulations and standards that were developed within the German medical informatics initiative (MII)[26]. The large collection of health data with patient centric view will guide further prevention strategies and algorithms in order to enhance better public care by using machine learning and artificial intelligence approaches.

---

[26] Medical Informatics Initiative Germany: https://www.medizininformatik-initiative.de/

Problem solved

The following problems are solved by this Use Case:

- Standardized interfaces and semantics to link and evaluate data from many actors, in particular from patients
- Aggregation and evaluation of sensitive data provided by smart wearables in protected environments (clinical edge computing environment)
- Federated consent management and data management to control data aggregation, access, and use

Partners/Ecosystem

Ecosystem partners include: patients, practitioners, clinics, smart wearable manufacturers, service providers, Big IT and Big Pharma. This use case will generate a person-centric ecosystem based on self-governed data collection and use/donation to health analytics applications and research. This separates data and applications, which makes data more universally useful.

Main technology/Gaia-X components

Main technologies are the common data standards from the medical field like HL7/FHIR and IHE for ensuring interoperability.

Pre-existing efforts by the MII[27] will be leveraged to ensure interoperability and compatibility with the data integration centers.

The following Gaia-X Core Services are required for this Use Case: Federated Catalog, Data Connector, IAM, Self-Description and Compliance. Within the context of Self-Description, the Metadata Schema for the Health Domain will be defined.

Concrete benefits

- Provide a component for secondary health data like smart wearable data
- Provide the linkage to the Electronic Health Record or Patient Health Record, also called the primary health data record
- Follow strictly a patient centric approach
- Effective means to contribute and access patients' documents across health enterprises, for preventive health, clinical care, and research
- Enable sharing patients' documents and data between patient and private physicians, clinics, long term care and acute care
- Provide access to health data for the patient/citizen from a single point of view

---

[27] MII consent, data sharing and interoperability working groups: https://www.medizininformatik-initiative.de/en/collaboration

- Enabling secure and GDPR-compliant access to patients' data through compliance with Gaia-X provided policies.

Description Use Case 4: Health Outcomes Observatory

Solution

The Health Outcomes Observatory project (H2O)[28] is a strategic partnership between the public and private sectors to create a robust data governance and infrastructure model to collect and incorporate patient outcomes at scale into healthcare decision making at an individual and population level. The H2O approach gives patients ultimate control of their health data and ensures that only they exercise this control.

This will materialize in a network of national Observatories in four European countries under the leadership of a European Observatory that will allow an ethical and legal use of data to:

- Measure outcomes more effectively.
- Promote value-based approaches in health care systems.
- Empower patients to have better communications with their HCPs.
- Facilitate personalised treatments.
- Enhance health research prospects.

The project will:

- Establish Observatories as legal entities and running services in each of the four participating countries (Austria, Germany, Netherlands and Spain) and in three disease areas initially (diabetes, Inflammatory Bowel Disease, Cancer).
- Collect data and share evidence information on patient-reported outcomes (PROs).
- Empower patients with digital tools data to better manage their health, report their health outcomes, and remain in control of their data.
- Extend the H2O concept to additional data sources and other settings.

Problem solved

With the H2O infrastructure and tools, patients will be able to measure their outcomes in a standardized way, whilst keeping full control of their data. Ultimately, this framework of Observatories aims to foster innovation in health care in Europe and beyond to deliver better outcomes for all.

As and when patients provide consent, this health data can be analysed in multiple ways and allow: patients and their physicians to communicate better, health authorities to make more informed decisions and researchers to advance knowledge and science in health care.

---

[28] Health Outcomes Observatory: https://health-outcomes-observatory.eu/

Partners/Ecosystem

Over 20 partners from 12 different countries participate in H2O. The Consortium brings together 13 academic institutions and small and medium enterprises, 8 pharmaceutical[29] and medical device companies and 2 associate partners. The H2O project is led by the pharmaceutical company TAKEDA and coordinated by the Medical University of Vienna (MUW). A full partner list is on the project website[30]. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under Grant Agreement No 945345. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Main technology/Gaia-X components

The project started before the Gaia-X initiative became concrete. It includes the design and implementation of the technical architecture to support data management according to the project vision. These encompasses the implementation of patient apps for PRO data collection and health care provider portal, and the development of a data integration platform, together with the overall Observatory framework.

Whereas the project will make independent design choices for its initial scope, there is an enormous opportunity to leverage the Gaia-X Health Data Space to support the eventual scale-up of this initiative. This could advance the full Gaia-X technology stack to create:

- A patient web portal and mobile apps for PRO data collection – hosted on Gaia-X compliant cloud infrastructure, utilizing the identity, consent and trust mechanisms
- A health care provider portal – hosted on Gaia-X compliant cloud infrastructure, utilizing the identity, consent and trust mechanisms
- A data integration platform – consistently deployed across the national and EU level observatories, enabling data anonymization and aggregation at various levels, creating federated access to selective data subsets

Concrete benefits

H2O is set to transform the use of patient-reported information in health care, enrich the interaction between patient and provider and, as a result, drive better outcomes for patients. The H2O approach will deliver a positive impact across the health care ecosystem by enhancing health research prospects, promoting the development of new treatments that reflect outcomes reported by patients and sustaining more efficient healthcare systems.

Further benefits are summarized on the project impact page[31].

---

[29] European federation of pharmaceutical industries and associations (efpia): https://www.efpia.eu/
[30] H2O partner list: https://health-outcomes-observatory.eu/partners/
[31] H2O project impact: https://health-outcomes-observatory.eu/impact/

Maturity indication of the data space, health status

Given the extent and the complexity of the healthcare domain, there will not be one single Health Data Space but a multitude of federated data spaces. The current state and the maturity vary from case to case. This initiative does not anticipate one revolutionary redesign, nor one man-on-mars sized deployment, but a gradual evolution where existing data spaces are getting upgraded, connected and federated, and where new data spaces can be designed from the start adopting the Gaia-X architecture.

How is the demand side represented?

The healthcare sector is multi-faceted, with a complex interplay between patients, care providers, insurers, governments and various other stakeholders. We have described this stakeholder landscape in section 3.1.

There is demand from the care delivery processes, for primary use of data for the individual patient and the healthy citizen, and for professional care providers: to deliver healthy living, prevention, diagnosis, treatment and homecare, for transforming towards value-based healthcare. There is demand for the secondary use of data on a cohort or population scale: for research, innovation, vigilance, public health and population health.

A health data space creates an open, transparent, federated catalogue of data sources and data services, where demanding stakeholders can obtain access to select data in a consistent, secure and trustworthy manner.

How is the supply side represented?

The citizen/patient is the original data holder of all personal health information. They consent with the sharing of parts of that information for the (professional) delivery of care, for health insurance and for legal obligations. They may also consent to participate in clinical trials and research projects. And they may share select data with select stakeholders for personal or commercial interests.

In most countries there is strict legislation on the minimal information that needs to be shared in the interest of public health, for medical device regulations, to manage the quality of healthcare delivery and for the functioning of the insurance system, for public and social care.

Hence, in theory there is no shortness in supply. However, in practice there is a heavy unbalance between situations where the citizen is not in control of where the data flows (disproportionate disclosure), and situations where they hold back essential data due to privacy concerns and lack of trust.

A health data space compliant with a strict and open set of policies, rules and standards will improve this current unbalance: giving full control to the respective data holders and creating an essential environment of trust.

Is there an equal representation of demand and supply side to provide a sustainable business model?

In essence, yes. Personal health data is the essential fuel in the healthcare engine. Care is delivered based on observations, measurements, and disease and treatment parameters. Treatment methods, medical devices and drugs are developed using tons of test and validation data.

Each use case needs to be evaluated on its potential impact, values and merits. Most of these use cases are in a too early stage for proper business evaluation, as they will be dependent on changes in reimbursement or funding policies.

Is the story of the data space well documented?

Yes. There are many data collection, sharing and processing initiatives in progress (a representative subset is summarized in Appendix B).

The eHealth Digital Service Infrastructure (eHDSI)[32] is a well-documented attempt to enable health data exchange across European country borders and the Joint Action (TEHDAS)[33] helps the Members States and the Commission in developing sharing of health data for public health, treatment, research and innovation in Europe.

There is more to be learned from ongoing initiatives amongst others but not limited to:

- FAIR principes (promoted by www.go-fair.org)
- The International Data Space Association (IDSA) focused on the proper governance of data exchange (www.internationaldataspaces.org)
- The OPEN DEI project, which is working on a set of design principles for data spaces, aligning reference architectures, open platforms and large-scale pilots in digitizing European industry, including healthcare (www.opendei.eu)

The application of the Gaia-X policies, rules and architecture to the complex healthcare landscape is an experimentation ground that needs to lead to the further detailing of a fully federated ecosystem of health data spaces within and across European borders.

What is the business model and the business mechanics of the data space after the PoC implementation?

There will be different business models and business mechanics for the different use cases. In most cases there will be a need for initial funding, for research and innovation and for public-private collaborations to establish a working architecture, the common enablers, and the initial

---

[32] eHDSI: https://ec.europa.eu/health/ehealth/electronic_crossborder_healthservices_en
[33] TEHDAS Joint Action : https://tehdas.eu/

demonstrators of the federated Health Data Space. Thereafter, usage-based payment models will fund the sustainable operation of infrastructure and services.

Which components will be certified according to the Gaia-X federation services?

All of the Gaia-X federation services need to be in place to enable the described use cases. The participants will need to operate these with the full system of policies, rules, compliance and certification, in order to create trust with all stakeholders in the complex healthcare ecosystem. As personal health data is one the most sensitive and trust-critical information for most citizens, the transparency and confidence in the end-to-end system that handles such data is absolutely essential. No single component can be left out of control, out of sight, out of audit, or without certification.

What is the potential for adoption of the endeavor and for further scaling?

The ambition of these use cases is to become a new standard for healthcare delivery and healthcare research. By lowering the thresholds to adopt and deploy these enabling capabilities, the expectation is to optimize healthcare delivery and to accelerate the aggregation of clinical data for research.

The growth speed of global healthcare data is stunning, according to statista[34], harmony healthcare IT[35], businesswire[36], emerj[37] and other sources that count in exabytes and zettabytes. The application areas of big data in healthcare are developing at massive speed, even further accelerated by the COVID pandemic.

The current state of data collection, exchange and processing in Europe is limited by scale, gaps in standardization and by legal challenges and national boundaries. Gaia-X provides a solid legal and technical foundation to enable these use cases to unlock from these traditional boundaries and to scale within and across the European Member States at the required pace to sustain its healthcare systems and to compete with the other large-population countries (>200M citizens).

How can the commitment of the parties involved be proven?

All the participants and use case contributors have prior experience with the design and setup of data lakes, data space and/or data exchange services. There is a strong drive in the Gaia-X

---

[34] Statista: https://www.statista.com/statistics/1037970/global-healthcare-data-volume/

[35] Harmony HIT: https://www.harmonyhit.com/health-data-volumes-skyrocket-legacy-data-archives-rise-hie

[36] Businesswire : https://www.businesswire.com/news/home/20181126005585/en/Seagate-Launches-New-Data-Readiness-Index-Revealing-Impact

[37] Emerj: https://emerj.com/ai-sector-overviews/where-healthcares-big-data-actually-comes-from/

health user community to team up, to scale and to accelerate an efficient and sovereign data exchange.

Are sufficient resources available to realize the endeavor according to its mission?

Not yet. All of the proposed use cases need additional investment and funding to realize their respective endeavors. Any Health Data Space initiative first needs to absorb and evaluated the Gaia-X concept and architecture and extend it with any additional, essential enablers to allow deployment for health care use. A demonstrator project like Hello Health is very well suited to experiment, to prototype, and to learn about the residual resource needs before deploying the solution to further use cases. Given the current Gaia-X roadmap, this should be possible before the end of 2021.

Many of the currently ongoing data space projects obtain funding in one or other way, such as from the Digital Europe Programme[38], Horizon Europe[39] or national funding programmes. The challenge will be to refocus such projects, with their associated funding parties, towards the Gaia-X concept and design. Eventually a Gaia-X based solution needs to lead to faster implementation times and lower development and operational costs.

---

[38] Digital Europe Programme: https://digital-strategy.ec.europa.eu/en/activities/digital-programme
[39] Horizon Europe: https://ec.europa.eu/info/horizon-europe_en

Evolution of the data space

Roadmap of the evolution

The roadmap can be established within three parallel tracks:

1. Apply the Gaia-X concepts and architecture for a minimal yet relevant healthcare use case, like Hello Health. This serves to demonstrate the functioning and validity of the framework end-to-end. The minimal use case can go through iterative development until it fully meets its purpose.

2. Onboard and transform existing data aggregation or data sharing initiatives, to make them compliant to the Gaia-X framework, to connect them and enable data exchange between them. A dedicated onboarding roadmap can be developed towards for this purpose.

3. Start up new data spaces, for new priority use cases, fully in line with the Gaia-X framework, such that they can benefit from the federated data space landscape from the start. A dedicated roadmap for new data space initiatives can be developed towards for this purpose.

Each of these tracks can add and prioritize new requirements towards the Gaia-X framework and feed into the overall Gaia-X roadmap.

Quick wins (for 2021)

Identify candidates for integration. A multitude of past initiatives, ongoing and future projects exist (initial overview in Appendix B). We intend to complete this overview and to identify those data initiatives that would benefit from integration into the Gaia-X framework.

Identify common enablers. These ongoing initiatives carry an enormous amount of knowledge and experience on the current limitations and bottlenecks in adoption, scaling, compliance and costs. The analysis of the common patters amongst these data initiatives will drive the definition and specification of common enablers

Develop a "Hello World" demonstrator. Hello Health is a fast track development to prove the Gaia-X concept and architecture in a simple but realistic and end-to-end health use case. This will create the enabling work to allow better planning of the implementation work of the subsequent use cases.

Mid-term benefits (2022-2023) building on already-launched or soon-to-be-launched projects

The initiative needs to iterate using a learn-and-adapt approach: start with initial demonstrator projects that can iterate and pivot at a fast pace. Learn, provide feedback and updates to the Gaia-X architecture and services infrastructure. Define and develop common enabling components to enrich Gaia-X towards a valuable health data and health services catalogue.

Then scale up the initial use cases into other regions/countries and into other application areas. Enable the ecosystem of participants to build new applications within the common framework, and demonstrate faster time-to-market, improved compliance and lower cost-to-develop and cost-to-operate.

Each of the proposed use cases has prior work and relevant projects in progress. As soon as the Gaia-X concepts are finalized and initial demonstrators proven, they can pivot their prior designs to comply with the Gaia-X concepts and framework. And other existing data initiatives can adapt and integrate into the Gaia-X framework over time.

Long-term benefits requiring significant investments on the 2021-2025 period

Long-term benefits will only come when these initiatives achieve significantly larger scale, a higher level of semantic interoperability and a high degree of trust within the user community.

In the current landscape a number of initiatives try to organize data and services catalogues (e.g. European Data Portal, European Open Science Cloud, OpenAIRE/Zenodo, EUDAT, ELIXIR, UK healthdatagateway, and more). These catalogues have some level of adoption, but none has reached the necessary scale, and they are not interoperable amongst each other. The Gaia-X concept of federated data and service catalogues should alleviate this problem, by providing a federation of catalogues that can be queried easily across all catalogues.

Many of the existing initiatives struggle with the same or similar basic functionality issues. The Gaia-X federated services should provide some of these basic capabilities in a sovereign and trustworthy environment. The Health Data Space extensions should provide the health sector specific basic enablers. Together, this creates a framework in which existing data initiatives can be integrated (with reasonable efforts and costs), and new initiatives can be established in a much easier way.

These common enablers include (preliminary list and summary):

- **Identification & authentication management:** a federated solution for electronic identity management (eID/eIDAS) for natural persons (physical persons) as well as legal persons (private or public entities) – essential enabler for consistent identity & access management (IAM) across European countries.

- **Authorization & permission management**: a federated solution for attributing, withdrawing and verifying the access rights to specific types of personal data with a specified legal basis. This includes the authorization of others to act on behalf of a specific individual (e.g. parents for children, caregivers for incapacitated people).

- **Access logging**: a federated service to compose and to view an aggregated log of any or all access to personal health information, easily accessible and readable for each citizen.

- A federated **catalogue of professional health providers** including physical and digital addresses and affiliations (could be an aggregation of national health provider catalogues) – essential enabler to identify health providers in a trustworthy manner (for individuals and the organizations they work for).

- A federated **catalogue of data institutes**: statistics, science and research institutes as well as commercial entities. This could be an aggregation of national catalogues – essential enabler to identify these entities in a trustworthy manner.

- A federated solution for **consent management**, as consent requirements vary across European countries. A harmonized service API would be ideal to request and to verify patient consent – essential enabler for consistent application of consent management across European countries.

- **Pseudonymization and anonymization services** for patient health records, medical records, images, monitoring signals, measurements, outcomes data, etc – essential for compliance with the GDPR, privacy and security requirements.

- **Federated catalogues and marketplaces** for health data and health services. They include the means to track provenance of data and the legal ground for sharing and processing such data – essential to improve the ease of data discovery and to stimulate a more data-driven healthcare. This may include a patient-data-index, where an individual can discover all data repositories that contain data from this individual.

- A data **request repository**, i.e. a shared space to publish research data needs and to solicit for data providers to collect and to contribute the requested data.

- A federated catalogue of **data storage, data warehouse and data lake services**, compliant with all requirements for the storage and archiving of health-related in Europe.

- **Terminology registers**: machine-readable resources that include terminology, syntax and coding definitions.

- **Terminology translation services:** to translate from one to another coding or system or format. This is required given the current variation of standards in use (e.g. HL7-CDA, HL7-FHIR, EDIFACT) and to manage language and other local variations.

Actions to be taken and recommendations for industry, politics and society

Actions and next steps:

- Stimulate and grow the Gaia-X **health user community**, within the national hubs as well as across European countries. Suggestion: create an international project community with the global health connector of the European Connected Health Alliance[40]. Include all relevant stakeholders, including: professional care providers, clinical specialties, patient and caregiver organizations, governmental and non-governmental interest groups. Include a broader sample of European countries. Define and communicate the process to onboard new participants properly. Improve the communication of the Gaia-X health user group with all stakeholders and with the Gaia-X national hubs.

---

[40] European Connected Health Alliance: https://echalliance.com/projects/

- Maintain and extend the candidate **use case gallery** (Appendix A): select and onboard additional use cases from all Gaia-X participants across all national hubs. This should be an ongoing activity.

- Maintain and extend the **existing data initiatives** list (Appendix B): add the missing initiatives to the list and formally invite these initiatives to join the Gaia-X user community. New candidates can onboard and integrate into the Gaia-X framework and make assets or services available in the Gaia-X federated catalogues. This should be an ongoing activity.

- Support the match-making between existing data initiatives and the proposed use cases. Suggestion: learn from the twinning mechanism in the Digital Health Europe programme[41]. This is an ongoing activity accompanying the use case proliferation.

- Extend and refine the **common enablers** as initially summarized in chapter 4.4, with inputs from the extended health user community and based on the knowledge and experience from existing health data initiatives. Common enabler requirements will either flow into the base Gaia-X framework or be proposed as essential Health Data Space enablers, requiring dedicated funding and development.

- Define an approach and a plan for the comprehensive and consistent **tracking of benefits** across all Gaia-X health use cases. Suggestion: consider the impact key performance indicators (KPI) as defined by the European federation of data driven innovation hubs (EUH4D)[42].

- Define a number of **charters** for various stakeholder groups, to bring transparency and to improve the buy in and support from these groups: citizens, data users, data providers. The charters explain the essence of data sovereignty, the soft infrastructure and the roles of the various stakeholders.

Recommendations for industry:

- **Converge** the ongoing or projected initiatives towards the common Gaia-X framework (avoid divergence and duplication of efforts).

- Improve transparency of the existing initiatives and deploy a process and tooling to maintain this transparency.

- Build the federated data and services **catalogues** to improve the availability of and access to data.

- Commit to privacy and security and to the principles of data sovereignty **by design**.

Recommendations for policy makers and society:

- **Align the EU initiatives** with Gaia-X: European Health Data Space development and roadmap for cross-border data exchange, the TEHDAS Joint Action, the European Alliance for industrial data, cloud and edge, and the EU Cloud Rulebook with the Gaia-X initiative.

---

[41] Digital Health Europe twinning scheme: https://digitalhealtheurope.eu/twinnings/
[42] EUH4D: https://euhubs4data.eu/impact/

- Obtain buy-in and support from the **Health Ministries** of the European Member States with the Gaia-X framework and architecture – invite them to integrate national initiatives with Gaia-X and to allocate appropriate funding for data initiatives that subscribe to the principles proposed in this position paper.

- Solve **electronic identity management** across Europe: continue and accelerate the deployment of federated eID solutions (eIDAS) for natural and legal persons.

- Agree on **data trust and data governance**: where/who defines the standards and the guidelines for ethical use of personal health data. Harmonize these principles and guidelines across EU Member States as much as possible and where needed make the local variations explicit and transparent.

- Provide adequate level of **funding** (from both EU programmes and national plans) to stimulate adoption of the Gaia-X framework and to kick start the deployment of a Gaia-X compliant Health Data Space.

- Encourage the European Data Protection Board[43] to provide clear and updated guidelines on the concepts of **personal data and non-personal data**, and on **anonymization techniques.**

- Reduce **fragmentation of local conditions** on data processing for scientific research purposes, given that Member States have leveraged art. 9 (4) GDPR to introduce further limitations to the processing of health data for scientific research purpose. Reduce fragmentation of local data protection/healthcare rules applicable to health data, in particular in the field of cross-border transfers of health data within the EU.

---

[43] EDPB work programme 2021-2022:
https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_workprogramme_2021-2022_en.pdf

Appendix A: initial Gaia-X use cases for the health sector

This appendix summarizes the list of Gaia-X provided uses cases for the health sector as documented in the use case gallery as originally collected by the German health user group. This list remains to be extended with use cases identified in other national health user communities.

Berlin Health Data Space	Early detection of acute kidney failure by exchanging and analysing data (e.g. Laboratory values, diagnoses, procedures)
KIKS (Artificial Intelligence for clinical studies)	Gain medical insights through data integration within a Digital Health Ecosystem. Transform healthcare on an international scale. Use of "UNITY"
Smart Health Connect	Integrating long-term follow-up data from hospitals, health insurers, new sensors, and patient-reported outcomes to better identify patients at high risk of disease to reduce mortality from serious heart diseases.
Research Platform Genomics	GA4GH compliant genomics platform for cancer research consisting of the complete ICGC data sets and using proven gold standard analysis methods. These can be rolled out to other cloud sites, if, for example, the sensitive data is not allowed to leave the clinic context.
Future Care Platform	A digital platform of the care sector for digital networking of all stakeholders, data and services. The better digital supply will enable better care even the proportion of people in need of care will be increasing by almost twice as much in the next 30 years.
Surgical Platform for AI-based Risk Identification	Risk Identification for complex interventions with AI due to a European surgical video and data storage
Medizinische Krisenmanagement- und Forschungsplattform "UNITY"	Enables the continuous recording of condition, medication, environmental factors and the consolidation of patient, laboratory and

	radiology means. Based on e.g., automated anamnesis, status recording and decision-making
Patient Empowered, Privacy Secured	builds a blockchain based ecosystem in which the patient's real identity will be held by various organizations like universities, hospitals or other public institutions. Essential is patient trust in the research being conducted with their medical data.
Framework of medical records in Europe	Framework with data pools and blockchain technology for faster medical insights through better patient-based data exchange. Crucial is the data security and data ownership as well as to carefully balance between medical benefit and legal responsibility.
Improve Chronic Heart Failure Patient Management	Value-based data platform that supports the management of patients with chronic heart failure. It captures patient data, costs and other data, and presents this data in several disease management dashboards. The approach strives to fewer hospital re-admissions, lower mortality rates, and cost reduction.
CarePay	Platform for exchange of data, the billing and regulation of services. It connects insurers, clinics/hospitals, third parties and participants. Each participant is connected to the platform through a mobile healthcare wallet, providing mobile access to healthcare financing. It has already been successfully launched in Africa.
Image Sharing for Medical Professionals and the Citizen	Infrastructure for the cross-company exchange of images based IHE standard ("Integrating the Healthcare Enterprise") the "Cross Enterprise Document Sharing for Images" (XDS-i) standard.
EMPAIA Markerquantifizierung	Common infrastructure for data, applications, and commercial services including a data marketplace provides data for the development of AI applications for marker quantification.

COVID-19 Dashboard & Hub	Real-time data dashboard to provide information on the spread of viruses. It is used as the central web portal at the RKI where the reported incidence data submitted by all health departments are displayed per region.
KAMeri – Kognitiver Arbeitsschutz für die Mensch-Maschinen-Interaktion	Improves human-machine interaction and reduces mental stress states. The use of new technologies, such as the continuous recording of EEG brain waves and subsequent evaluation in cloud-based solutions, reduces occupational accidents and promotes the physical and mental health of employees.
Recupera REHA	Improving care at home through telepresence and treatment analysis. The aim is that patients carry out self-determined therapy measures, e.g. with the help exoskeleton systems.
The digital twin	Digital patient twin to assist personalised medicine in oncology will foster cross-institutional standards and processes for data exchange. To allow a coherent overall picture of the patient without great expense.
Differential diagnosis	Supportive platform for clinical diagnosis with neural patient representations. Critical diagnoses could be detected more quickly, and rare diseases were diagnosed in a more targeted manner.
Smart Diabetes Management	Large-scale, digitalized, accessible and semi-autonomous system to treat diabetes. The Smart Diabetes Management interfaces with Continuous Glucose Monitors (CGM), insulin pumps, smart watches and heart rate variability monitors.
End-to-end Cell & Gene Therapy Orchestration	Improving Cell and Gene therapies through digitalization and networking. Wide coordination and automation of information as well as process flows will improve the therapy supply chain by connecting all relevant stakeholders in the production process.

Appendix B: existing health data initiatives

This appendix lists a characteristic set of existing health data initiatives. This is a comprehensive and non-limiting summary which can be extended with additional examples.

EU level health data initiatives:

ECDC - European Centre for Disease prevention and Control - publications & data: https://www.ecdc.europa.eu/en/publications-data

EMA - European Medicines Agency - big data landscape: https://www.ema.europa.eu/en/about-us/how-we-work/big-data

EHDSi - Europe Health Data Space infrastructure: https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHealth+DSI+Operations+Home

National health and statistical data institutes:

Austria - Austria data market: http://www.datamarket.at/

Belgium - Sciensano: https://healthdata.sciensano.be/

Estonia - e-estonia for healthcare: https://e-estonia.com/solutions/healthcare/

Finland - Findata: https://www.findata.fi/en/

France - Health Data Hub: https://www.health-data-hub.fr/

Germany - Genesis-online: https://www.destatis.de/EN/Home/_node.html

Ireland - eHealth Ireland: https://data.ehealthireland.ie/

Italy - Istituto Nazionale di Statistica: health statistics: https://www.istat.it/en/health-statistics

Netherlands - centraal bureau voor de statistiek - gezondheid en welzijn: https://www.cbs.nl/nl-nl/maatschappij/gezondheid-en-welzijn

United Kingdom - the Health Data Research Innovation Gateway: http://www.healthdatagateway.org

European health research projects:

BD4BO - Big Data for Better Outcomes: https://www.imi.europa.eu/projects-results/project-factsheets/bd4bo

EHDEN - European Health Data and Evidence Network: https://ehden.eu/

BigData@Heart: https://www.bigdata-heart.eu/

PIONEER - Prostate cancer diagnosis and treatment enhancement through the power of big data in Europe: https://prostate-pioneer.eu/

HARMONY & HARMONY PLUS - Healthcare alliance for resourceful medicines offensive against neoplasms in hematology: https://www.harmony-alliance.eu/

ROADMAP - Real world outcomes across the AD spectrum for better care: multi-modal data access platform: https://roadmap-alzheimer.org/

FAIRplus delivers guidelines and tools to facilitate FAIR principles to data from IMI projects and datasets - project website

H2O - Health Outcomes Observatory: https://health-outcomes-observatory.eu/

MHMD - MyHealthMyData: http://www.myhealthmydata.eu/

BIGPICTURE - central repository for digital pathology: https://www.bigpicture.eu/

COVID-RED - Remote Early Detection of COVID-19 infections: https://www.covid-red.eu/en/

EU-PEARL - EU patient-centric clinical trial platform: https://eu-pearl.eu/

Digital Health Europe - deployment of digital solutions for person-centred integrated care: https://digitalhealtheurope.eu/

Evotion project - on hearing loss: http://h2020evotion.eu/

BigO project - on child obesity: http://bigoprogram.eu/

iPrognosis - on Parkinson's Disease: http://i-prognosis.eu/

SOMA - on work-related stress: http://www.soma-analytics.com/

BBMRI-ERIC - a European research infrastructure for biobanking: https://www.bbmri-eric.eu/

B1GM - Beyond One Million Genomes: https://b1mg-project.eu/

EUCANCan - cancer genomics and health data sharing: https://eucancan.com/

EUCAN-Connect - aims to promote collaborative and multidisciplinary research in high-value cohort and molecular data on a large scale.: https://eucanconnect.com/

euCanSHare - a joint EU-Canada project to establish a cross-border data sharing and multi-cohort cardiovascular research platform: http://www.eucanshare.eu/

EuCanImage - a highly secure, federated and large-scale cancer imaging platform: https://eucanimage.eu/

ELIXIR - coordinates, integrates and sustains bioinformatics resources vital for research. https://elixir-europe.org/

I-HD - the European institute for innovation through health data: https://www.i-hd.eu/

European data science initiatives:

EDP - European Data Portal with the EDP catalogue: https://www.europeandataportal.eu/en

EOSC - European open science cloud - service marketplace: https://marketplace.eosc-portal.eu/services

OpenAIRE - Science. Set Free. Service catalogue: http://catalogue.openaire.eu/search

ARGOS - an open extensible service that simplifies the management, validation, monitoring and maintenance of Data Management Plans: https://argos.openaire.eu/home

zenodo - a general purpose repository or research data sharing: https://www.zenodo.org/

EUDAT - the EUDAT Collaborative Data Infrastructure (EUDAT CDI) and eTDR (a trusted digital repository): https://www.eudat.eu/

National data initiatives:

BBMRI (Netherlands) - provides access to biosamples, images and data; tools to capture, integrate and analyse data; support on ethical, legal and societal implications: https://www.bbmri.nl/

BBMRI (Germany) - the umbrella organization of university biobanks in Germany: https://www.bbmri.de/

BBMRI (Italy) - an infrastructure distributed throughout the national territory that includes Biobanks, Biological Resource Centers and Collections located in different Italian regions: https://www.bbmri.it/

deNBI cloud (Germany) - federated compute services and data exchange for German researchers, part of the ELIXIR ecosytem: https://www.denbi.de/cloud

PALGA (Netherlands) - The nationwide network and registry of histo- and cytopathology in the Netherlands: https://www.palgaopenbaredatabank.nl/

Global data science initiatives:

Accumulus Synergy - a nonprofit consortium to build a global information exchange platform to transform how drug innovators and health regulators interact to bring safe and effective medicines to patients faster and more efficiently: https://www.accumulus.org/

DHO - Digital Health Observatory: https://digitalhealthobservatory.com/about-us/

ECH Alliance - the European Connected Health Alliance is the Global Health Connector for Digital Health: https://echalliance.com/

FAIRsharing.org - A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies: https://fairsharing.org/*

GA4GH - Global Alliance for Genomics & Health is a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a human rights framework: https://www.ga4gh.org/

IDDO - the Infectious Diseases Data Observatory: https://www.iddo.org/

MIDAS Biomedicina - National Open Access Scientific Data Archive Information System from University of Vilnius: https://biomedicina.midas.lt/

OHDSI - Observational Health Data Sciences and Informatics: https://www.ohdsi.org/

TCIA - The Cancer Imaging Archive deidentifies and hosts a large archive of medical images of cancer accessible for public download: https://www.cancerimagingarchive.net/

TSD - services for sensitive data: a sensitive data storage service programme from University of Oslo: https://www.uio.no/english/services/it/research/sensitive-data/

YODA - Yale University Open Data Access Project at the Center for Outcomes Research and Evaluation advocates for the responsible sharing of clinical research data: https://yoda.yale.edu/

Appendix C: suggestions for future work

- In appendix A & B: split the lists into current and future use cases and initiatives; or indicate this as an attribute with each listing.
- In the use cases: be more specific on what needs to be built within Gaia-X and on top of Gaia-X to realize each of these use cases.
- In the use case descriptions: harmonize the language of data holders and data users, and be more specific on the relevant stakeholders. Clarify the legal basis of each use case more clearly.