



Data Space Business Committee – Position Papers

Consolidated Version for Industry Verticals

Contents

Executive Summary	12
Position Papers	14
Agriculture	14
Data and agriculture, a state of play	14
Challenges addressed	14
Common European Agricultural Data Space	16
Use-cases within the Agricultural Data Space	17
Description Use Case: AgriFood Data Space (Finland)	17
Solution	17
Data governance	18
Problem solved	18
Partners/Ecosystem	19
Main technology/Gaia-X components	19
Concrete benefits	19
Benefits to farms	19
Benefits to all	20
Description Use Case: Agricultural Datasharing Platform (Belgium)	20
Solution	20
Problem solved	21
Partners/Ecosystem	21
Main technology/Gaia-X components	22
Concrete benefits	23
Description Use Case: Data intermediaries: Consent and data exchange in agriculture (France)	23
Solution	23
Problem solved	24
Data exchange platform	24
Partners/Ecosystem	25
Main technology/Gaia-X components	25
Concrete benefits	25
Description Use Case: Agri-Gaia an AI ecosystem (Germany)	26
Solution	26

Problem solved	26
Partners/Ecosystem	26
Main technology/Gaia-X components	27
Concrete benefits	27
Description Use Case: Knowledge Graphs for data integration in the AgriFood sector (Poland)	27
Solution	27
Problem solved	27
Partners/Ecosystem	28
Main technology/Gaia-X components	28
Concrete benefits	29
Description Use Case: Interoperable data exchange for online task management, machine tracking and sensor communication (Belgium)	29
Solution	29
Problem solved	30
Partners/Ecosystem	30
Main technology/Gaia-X components	31
Concrete Benefits	31
Description Use Case: Connected Potato data chain (Belgium)	31
Solution	32
Problem solved	32
Partners/Ecosystem	32
Main technology/Gaia-X components	32
Concrete benefits	32
Education	33
Our vision for a European Education & Skills Dataspace	33
Targeted challenges 2021-2023: list of use-cases	36
Ecosystem partners & stakeholders in France	37
Use-case 1: Exploring AI and data in education and guidance, with researchers	38
Example 1: Modelling trends and guidance pathways to improve tailor-made counselling.	38
Example 2: Developing voice assistants and language learning tools for all CEFR levels and for allophone students	39
Use-case 4: Improving the assessment of digital solutions through harmonised key criteria description and impact measurement.	41

Use Case 7: Tokenization of skills data sharing	42
Use-case 10: Developing a "compliance by design" framework implemented by Gaia-X governance and the Education & Skills Data Space, based on a Code of Conduct, and sovereignty and security conditions.	43
Use-case 14: Development of a one-stop-shop programme for data sharing for research purposes.	45
Energy	48
Mission and Goals of the data space	48
The Gaia-X contribution to the energy Domain	48
Added value to the whole energy value chain	48
Challenges addressed	48
Solution: Data space description in a holistic view – detailed view on the endeavor	49
Description Use Cases Renewables	50
Solution	50
Problem solved	51
Partners/Ecosystem	52
Main technology/Gaia-X components	52
Concrete benefits	52
Renewables – Works risk prevention	52
Problem solved	53
Solution	53
Partners/Ecosystem	54
Main technology/Gaia-X components	54
Concrete benefits	54
Questions/Answers identified at this stage:	55
Renewables – Common taxonomy definition – IEC standards	55
Solution	55
Problem solved	55
Partners/Ecosystem	56
Main technology/Gaia-X components:	56
Concrete benefits	56
Use Cases Nuclear	57
General context	57
Gifem, a key player federating & transforming nuclear industry	57

Identified opportunities within Gaia-X	58
Solution	59
Problem solved	59
Partners/Ecosystem	60
Main technology/Gaia-X components:	60
Concrete benefits	60
Solution	61
Problem solved	61
Partners/Ecosystem	62
Main technology/Gaia-X components:	62
Concrete benefits	62
Solution	62
Problem solved	63
Partners/Ecosystem	63
Main technology/Gaia-X components:	64
Concrete benefits	64
Solution	64
Problem solved	65
Partners/Ecosystem	65
Main technology/Gaia-X components:	65
Concrete benefits	66
Use Cases low-carbon hydrogen	66
Solution	66
Problem solved	67
Partners/Ecosystem	67
Main technology/Gaia-X components:	67
Concrete benefits	67
Solution	67
Problem solved	68
Partners/Ecosystem	68
Main technology/Gaia-X components:	68
Concrete benefits	68
Solution	68
Problem solved	69
Partners/Ecosystem	69

Main technology/Gaia-X components:	69
Concrete benefits	70
Use Cases downstream, energy renovation and electric vehicles	70
Solution	70
Problem solved	70
Partners/Ecosystem	71
Main technology/Gaia-X components:	71
Concrete benefits	71
Solution	72
Problem solved	72
Partners/Ecosystem	72
Main technology/Gaia-X components:	72
Concrete benefits	73
Solution	73
Problem solved	73
Partners/Ecosystem	74
Main technology/Gaia-X components:	74
Concrete benefits	74
Solution	74
Problem solved	75
Partners/Ecosystem	76
Main technology/Gaia-X components	76
Concrete benefits	76
Solution	76
Problem solved	77
Partners/Ecosystem	77
Main technology/Gaia-X components:	78
Concrete benefits	78
Solution	78
Problem solved	79
Partners/Ecosystem	79
Main technology/Gaia-X components:	79
Concrete benefits	79
Use Cases networks	80
Solution	80

Problem solved	81
Partners/Ecosystem	81
Main technology/Gaia-X components:	82
Concrete benefits	83
Solution	83
Problem solved	83
Partners/Ecosystem	84
Main technology/Gaia-X components:	84
Concrete benefits	84
Solution	84
Problem solved	85
Partners/Ecosystem	85
Main technology/Gaia-X components	85
Concrete benefits	85
Use Cases Compliance and standards	85
Solution	85
Problem solved	86
Partners/Ecosystem	86
Main technology/Gaia-X components	86
Concrete benefits	87
Solution	87
Problem solved	87
Partners/Ecosystem	88
Main technology/Gaia-X components:	88
Concrete benefits	89
Solution	89
Problem solved	89
Partners/Ecosystem	90
Main technology/Gaia-X components:	90
Concrete benefits	90
Evolution of the data space	94
Roadmap of the evolution	94
Quick-wins (for 2021)	94
Mid-term benefits (2022-2023) building on already-launched or soon-to-be-launched projects	94

Long-term benefits requiring significant investments on the 2021-2025 period	95
Actions to be taken and recommendations for industry, politics and society	95
Finance	96
Expectations towards a Gaia-X roadmap for the European financial ecosystem	96
2021-2022: “quickwins” rapidly available	97
2022-2023: mid-term benefits building on already-launched or soon-to-be-launched projects	98
2023-2025: long term benefits requiring significant investments on the 2021-2025 period	98
Health	99
Executive summary	99
Mission and Goals of the Gaia-X Health Data Space	100
Mission: access and share health related data securely and confidently.	100
Challenges addressed	101
Solution: Data space description in a holistic view – detailed view on the endeavor	103
Use Case 0: Hello Health!	107
Solution	107
Problem solved	108
Partners/Ecosystem	108
Main technology/Gaia-X components	108
Concrete benefits	109
Use Case 1: Research Platform Genomics	109
Solution	109
Problem solved	109
Partners/Ecosystem	110
Main technology/Gaia-X components	110
Concrete benefits	110
Use Case 2: Image Archiving and Image Sharing for Medical Professionals and the Citizen	111
Solution	111
Problem solved	111
Partners/Ecosystem	111

Main technology/Gaia-X components	112
Concrete benefits	112
Use Case 3: Smart Health Connect	113
Solution	113
Problem solved	113
Partners/Ecosystem	113
Main technology/Gaia-X components	113
Concrete benefits	114
Use Case 4: Health Outcomes Observatory	114
Solution	114
Problem solved	115
Partners/Ecosystem	115
Main technology/Gaia-X components	115
Concrete benefits	116
Evolution of the data space	121
Quick wins (for 2021)	121
Mid-term benefits (2022-2023) building on already-launched or soon-to-be-launched projects	121
Long-term benefits requiring significant investments on the 2021-2025 period	122
Actions to be taken and recommendations for industry, politics and society	123
Appendix A: initial Gaia-X use cases for the health sector	126
Appendix B: existing health data initiatives	128
Appendix C: suggestions for future work	132
Industry 4.0	133
Defining the Data Space Industrie 4.0: Enabling the paradigm shift towards multilateral collaboration	133
Multilateral collaboration and data-sharing – two directions of impact, one target	133
The Data Space Industrie 4.0: A common set of technical, legal, and business foundations	134
Plattform Industrie 4.0: Implementing the Data Space Industrie 4.0 as a joint approach of public and private stakeholders	134
Mobility	137
Expectations towards a Gaia-X roadmap for the European Mobility Dataspace	137

Mission and Goals of the Mobility Dataspace	137
Challenges to be addressed	137
Mobility Dataspace Objectives	139
Public Sector	140
Geoinformation	143
Mission and Goals of the data space	143
Challenges addressed	144
Solution: Data space description in a holistic view – detailed view on the endeavor	145
Use Case 1 – Space4Cities	146
Solution	146
Problem solved	146
Partners/Ecosystem	146
Main technology/Gaia-X components	146
Concrete benefits	147
Use Case 2 – Smart Infrastructure Management	147
Solution	147
Problem solved	147
Partners/Ecosystem	147
Main technology/Gaia-X components	148
Concrete benefits	148
Use Case 3D Planning Platform	148
Solution	148
Problem solved	148
Partners/Ecosystem	148
Main technology/Gaia-X components	148
Concrete benefits	149
Use Case Automatic 3D Spatial Content Generation	149
Solution	149
Problem solved	149
Partners/Ecosystem	149
Main technology/Gaia-X components	149
Concrete benefits	149
Use Case Earth Observation Federate Data Access	150
Solution	150

Problem Solved	150
Partners/Ecosystem	151
Main technology/Gaia-X components	151
Concrete benefits	151
Maturity indication of the data space, health status	152
Evolution of the data space	154
Roadmap of the evolution	154
Quick wins (for 2021)	154
Mid-term benefits (2022-2023) building on already-launched or soon-to-be-launched projects	154
Long-term benefits requiring significant investments on the 2021-2025 period	154
Actions to be taken and recommendations for industry, politics, and society	155
Smart Living	156
Mission and Goals of the data space	156
Challenges addressed	158
Solution: Data space description in a holistic view – detailed view on the endeavour	160
Partners of the ecosystem	164
Use-cases (scenarios) within the data space and the current state of development	165
Use Case Smart Living Assistance/AAL	165
Solution	165
Problem solved	166
Main technology/Gaia-X components	166
Concrete benefits	166
Use Case Smart Living Energy Efficiency of Smart Buildings	167
Solution	167
Problem solved	168
Main technology/Gaia-X components	169
Concrete benefits	170
Maturity indication of the data space	170
Evolution of the data space	175
Roadmap of the evolution	175
Quick-wins (for 2021)	175

Mid-term benefits (2022-2023) building on already-launched or soon-to-be-launched projects	175
Long-term benefits requiring significant investments on the 2021-2025 period	176
Actions to be taken and recommendations for industry, politics, and society	176

Executive Summary

Within the past years data availability and data sharing has rightfully been identified as one of the key dimensions in order to leverage digitalization and generate economic welfare across all industries and for the overall society. Especially business to business (B2B) and business to governance (B2G) data sharing will create tremendous opportunities.

At the Data Space Business Committee (“DSBC”) we accumulate and engage with various experts from a multitude of different data spaces (“vertical ecosystems”) as well as dedicated subject matter experts on relevant dimensions for the creation of data spaces (e.g. Governance, Technic, Legal/Policy/Compliance, Financial). Throughout the following pages the initial ideas, plans, use cases and approaches on how to engage in a data sharing ecosystem based on Gaia-X values are aggregated. Gaia-X creates the foundation for a federated, open data infrastructure based on European values.

The goal of this compilation of dataspace position papers is to provide an idea of the current state in each of the following verticals: Agriculture, Education, Energy, Finance, Health, Industry 4.0, Mobility, Public Sector, Geoinformation and Smart Living. This compilation aggregates information about the data spaces’ current missions, challenges, and roadmaps of planned or previously implemented use cases.

We want to explicitly express our gratitude for the great work that has been done whilst drafting all the different data space positioning papers, which originate from across Europe, and for the sharing.

The following section summarizes the most important aspects in each of the data spaces:

Agriculture: Throughout the entire agriculture chain new digital tools and new data-driven applications are deployed. To scale up and to make optimal use of this data revolution, a key Federation of existing IT infrastructures including platforms that deliver data sharing services (data intermediaries) is needed. A common European agriculture dataspace should ensure data sharing over the full supply chain from Farm to Fork.

Education: Combining the strengths of Edtech’s, governments and research entities, the Education dataspace aims to create an education and skills data space linked to Gaia-X, which is going to foster Europe's competitiveness and to spread of European values improving, simultaneously, research, AI development and innovation, in a legal stable environment.

Energy: The goals of the energy data space are to support and accelerate the energy transition in Europe and to develop businesses at European and worldwide scales. Additionally, new services to European citizens, taking advantage of Gaia-X infrastructure, data protection services and associated value allowed by these services should be provided.

Finance: Banks, insurance companies, financial institutions, Fintech’s, cloud service providers, software editors, academic institutes and governmental agencies have been lacking a trusted cloud service & data space between their internal cloud and data platforms and external cloud platforms. Through the Gaia X initiative, the Finance data space foresee a tangible and close way forward to easily build, assemble and use trusted and value-creating data-based cloud services,

as well as create new products/services and foster new business models that are compliant “by design” with European regulations and values.

Health: Health data is a sensitive topic for European citizens, as well as a strategic asset for governing bodies to understand and protect their population. The four main goals of the Health data space are 1) maintaining European sovereignty on health data, 2) creating trust in personal healthcare data usage, guaranteeing transparency and compliance towards EU and local regulations, 3) improving care provided to patients with big data and AI technologies, and 4) focusing healthcare and health pathways on patients’ outcomes and real-world evidence.

Industry 4.0: Typically, today’s business value in production and manufacturing is generated in a global production value network. This network is built on single, bi- and multi- lateral business relationships. However, this bilateral nature of cooperation hinders multi-stakeholder collaboration down- and upstream the value chain. To realize highly impactful use cases, depends on a fundamental paradigm shift in the global production ecosystem from bilateral to multilateral collaborations, allowing end-to- end data exchange in a fully interoperable and sovereign data-oriented manner: The Data- Space Industry 4.0.

Mobility: By focusing on enhancing user experience in mobility, while creating new business opportunities for the ecosystem’s members, the Mobility data space aims to 1) offer end-to-end seamless and safer mobility for users, 2) increase productivity for the various stakeholders of the mobility and digital sectors, whether public organizations, large private companies, or SMEs, and 3) be fully aligned with EU and National government priorities: Privacy Protection, Digitization, Environment, Industrial development, Territorial cohesion, etc. The emergence of federated service and data platforms which can provide for an easy and trusted exchange of data is thus actively welcomed.

Public Sector: The declared goal of Gaia-X's Public Sector domain is to drive exchange between public sector players and public and private IT service providers as well as platform operators and to make joint efforts to build a digitally sovereign, resilient, and cross-sector data infrastructure and to improve data use overall. Given Germany's federal structure, this also means informing the administrations at all federal levels about Gaia-X even more than before and convincing them to participate.

Smart Living: Given the fact of intensive connections to adjacent domains, such as smart energy networks, smart city, assistance and AAL, etc., the overall concept of Smart Living is “Building as an intelligent Service”. A smart home, a smart building or a group of smart buildings should be able to interact intelligent with smart energy networks, smart city structures, smart mobility offerings and smart assistance services. The mission and goal of a Gaia-X Smart Living Shared Data Space is to enable such cross-domain services and applications.

Position Papers

Agriculture

Data and agriculture, a state of play

Available data are only valuable when shared.

An enormous digitalisation effort is underway in the AgriFood sector. Agricultural and horticultural companies produce an enormous amount of products every day, but more and more also a mass of data. It is generally known that this data only has value if it is shared, brought together and then converted into useful applications or advice.

Therefore, throughout the entire chain from farm to fork, traditional supply companies such as machinery constructors and agrochemists, as well as new players are positioning themselves on this subject, as the data economy constitutes a new growth driver for them. They are deploying digital tools and new data-driven applications, aiming as end-user not only farmers or AgriFood business, but also the consumers. To scale up and to make optimal use of this data revolution in the near future, the willingness to share data between all AgriFood supply chain partners is crucial to keep innovating.

The common European agriculture dataspace should ensure data sharing over the full supply chain from farm to fork. So indeed, the European agriculture data space cannot be an isolated data space as agriculture and food is not an isolated domain but connected to other domains like public sector, energy, mobility, health, and so forth.

This position paper is a first step in identifying the current situation, challenges, and possibilities for an agricultural data space. The next step will be engaging with already existing platforms and their use cases that are already building towards a common agricultural dataspace. To identify what they have in common and to understand their common needs and challenges so we can define how Gaia-X can support them in taking steps toward this common agricultural data space. The paper was produced by ambassadors from the Gaia-X Hubs in Belgium, Finland, France, and Germany.

Challenges addressed

Building trust through consent

The asymmetry of size between the players, where multinationals, SME's, one-man businesses, and family farms collaborate, requires building a climate of trust between the business operators and the farmers. For this to happen, the consent of the data owners on the use of the data seems to be an essential base for the construction of a dynamic ecosystem. A first set of guidelines for this ecosystem was developed with the "EU code of Conduct on

agriculture data sharing by contractual agreement”¹. The Code puts the farmers in control of their data and. At the same time, it promotes the benefits of sharing data and enables agri-business models, including agri-cooperatives and other agri-businesses, but needs to be further developed and integrated into the emerging data sharing initiatives in AgriFood. In this way, mechanisms taking care of possible power imbalances become part of the data ecosystem to guarantee fair compensation to all involved actors and a trustful environment that incentivises participation.

Data Interoperability and portability

A lot of different actors in the Agri-Food ecosystem have developed data driven tools and platforms. The current situation is, that there is **little interoperability and portability** between these different initiatives, which results in the fact that for instance a farmer has to use many different platforms on a daily basis, not being able to integrate the knowledge gained from one platform onto another. Interoperability between these existing initiatives is one of the key barriers for working towards a common agricultural data space.

Access to the digital single market

The spectrum of European farms is wide, as also the selection of available digital technologies. On the one hand, at European scale, it is usually difficult for new start-ups or SMEs with novel applications to **reach viable number of customers** fast enough. On the other hand, farms may struggle with technology solutions that are not fully suitable for them, do not obviously show a benefit or it is not directly obvious, how these solutions can be comparable, or the farmers just do not know that better solutions exist. There is a need for a single market of available technologies/services/data sources and their accessibility in cross-country scale.

Need for sustainable business models

Data sharing initiatives are emerging in the AgriFood system, but all of them **are struggling to find a sustainable business model**. Moreover, when connecting data sharing platforms in a federation, new federated business models are necessary.

Involving the whole chain from farm to fork

When thinking of governance of agricultural data space, technical, legal, ethical, socio economical and business aspects need to be considered. But also, from farmers’ point of view, to fully deploy the possibilities of data sharing, it is important to connect with the rest of the food system. Especially consumers are in the position to impact on ‘data-driven farming practices’ through their informed choices based on rich product data. Rich product data could contain physical quality, origin, responsibility indicators, production methods from farm to

¹ This Code of Conduct was jointly developed by CEETTAR, CEJA, CEMA, COPA-COGECA, ECPA, EFFAB, ESA, FEFAC and Fertilizers Europe and endorsed by Animalhealth Europe and CLIMMAR. <https://old.fefac.eu/files/81630.pdf>

markets. Reaching out to the consumer offers great potential for farmers to be transparent and to show how to produce high quality food in a sustainable way.

Common European Agricultural Data Space

A federation of existing IT infrastructures including platforms that deliver data sharing services (data intermediaries), is needed. The definition and development of a federated governance structure that will define the scope, the openness, the technical interfaces and the 'federated' services is a challenge for the AgriFood community: a governance model with the right balance between sufficiently open so that companies are prepared to come up with joint innovations through data sharing and sufficiently protected so that they still retain control over their own data and intellectual property.

The federated governance structure needs to determine the adequate legal, organizational, semantic, and technical interoperability levels, ensuring innovation and scale-up of data sharing at the EU level, preventing lock-in situations, enabling new ways of coordination and value creation and achieving a sustainable digital and societal transformation.

Thus, based on these observations, the objective is to move towards **a common European agricultural data space** that is dynamic and capable of addressing all these challenges collectively. The infrastructures, which will be developed at a later stage, should be able to contribute to the emergence of artificial intelligence solutions, data analysis, etc., which are useful to the agricultural and Agri-Food sectors.

In agriculture, there are already various data intermediaries and companies that cooperate with them. In a way these data intermediaries act as regional data spaces, and they usually also represent regional conditions, so a future EU-wide data space has to consider these existing regional actions. It is through these data intermediaries that a network can be formed to increase interoperability. These existing data intermediaries could be connected based on the Gaia-X Federated Services. This includes a unified identity management, the sovereign data exchange including data usage control possibilities between the brokers and within the systems, and the Federated Catalogue to identify the actors, services and available data items in the network across intermediaries.

An important aspect of the common European agricultural data space is the enforcement of data sovereignty for the farmer as data originator according to the European Code of Conduct on agricultural data sharing. For this purpose we need to provide the farmer with the necessary tools to give consent on the use of the data by a network of data intermediaries and to stay in full control of his data and, if desirable also add independent, federated storage solution for the farmers.

In addition to the technical layer of this data space, the non-technical layers are equally important. Like a unified compliance framework based on Gaia-X. Hence, all Gaia-X Federated Services are needed and will be integrated to set up the common European agriculture data space.

Use-cases within the Agricultural Data Space

As referenced above, data sharing in the field of agriculture and food is not something new, but a lot of existing data sharing initiatives already exist all over Europe. These existing initiatives are all in a way building towards a common European data space, but often have different architectures, requirements, stakeholder and make use of different standards, which makes it a challenge to connect them at the European level. That is where Gaia-X could bring a lot of value.

Below you can find some of these existing initiatives which could be further enabled by Gaia-X. This list is not exhaustive, but gives a general overview of what is already existing and what the specific requirements would be for Gaia-X to further develop these use cases. The Gaia-X Agriculture community plans to further expand this list and plans to include other use cases as well in the future.

Next to existing data sharing initiatives, some horizontal use cases are also referenced.

Description Use Case: AgriFood Data Space (Finland)

Solution

The use case AgriFood Data Space (AFDS) approaches the data space with a wide scope including the entire food system from field to consumers. It focuses on core functionalities that the data space should provide to all actors, and that are missing in present systems of platform economy. Thus, our use case contributes to data intermediary (federation) services, data usage control and data governance mechanisms.

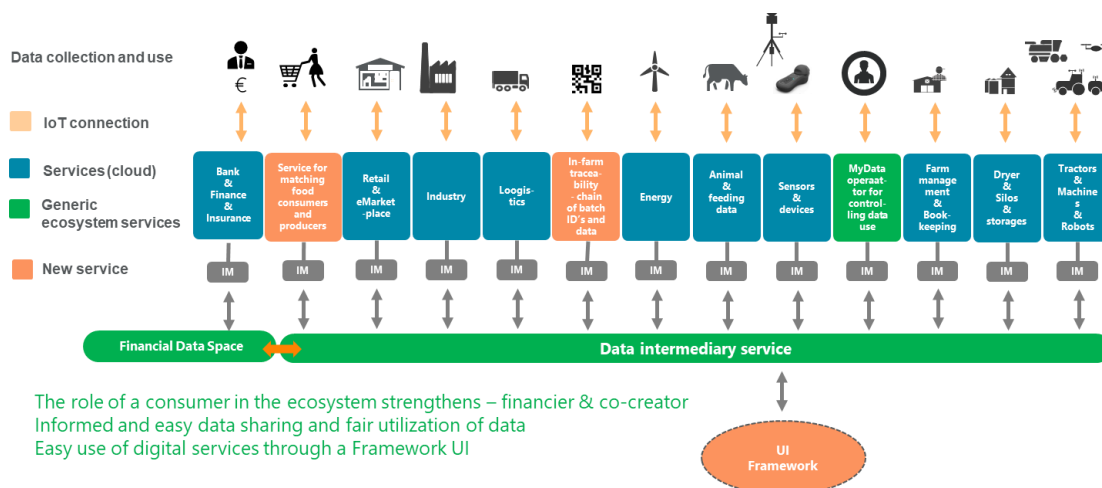


Figure 1: High level illustration of the use case AFDS. Green colored parts represent common ecosystem services needed in the data space and orange parts illustrate new services that implement functionalities that are desired in food ecosystem and made possible

Data intermediary (federation) routes the data, delivers it reliably to the destinations, provides event registering and monitoring. Services are integrated through an Integration Module that encrypts/decrypts the data. Data intermediary routes the data to authorized receivers.

The federation and data usage control services of AFDS are designed to match the principles set for MyData Operators. These Federation Services have two main directions in which they operate:

- **For individuals:** MyData Operators provide transparency, understandability, and convenience to individuals when they share data or receive services using data. Operators provide an aggregated view to data, allow them to control who can use the data and for which purpose, and transparently expose past data use and sharing. Other benefits include intuitive user interfaces, enhanced security, and the tools for managing relationships with different services that process the data.
- **For organisations:** MyData Operators provide easy, legally compliant connectivity to an ecosystem of Data Providers and Data Consumers as well as a relevant base of potential users. Operators facilitate access to high quality, up-to-date data in real time, offer tools and mechanisms for legal compliance such as logging and audit trails of permissions, and offer outsourced tools for complying with data portability requirements.

Data governance

The ecosystem of the use case needs to agree upon the rules that the data is shared and used within the partners. The rights and duties of data source, data provider, data consumer must be clear to all ecosystem participants. Setting the rules and monitoring them in practice require a governance structure that the ecosystem must set up. Data governance must comply with EU and national legislation and it can utilize the ‘EU Code of Conduct on agricultural data sharing by contractual agreement’. The ecosystem may also agree more detailed and specific rules, where the Fair Data Economy Rulebook model provides guidance. The ecosystem of AFDS may include several business clusters, which may have their own specific data governance rules and models within the common AFDS ecosystem rules. For example, diverse data cooperatives could be possible.

Problem solved

The basic data space functionalities enable new services that have been seen vital for sustainable and safe food systems in a manner that are available for all actors, especially farmers. This is due to possibility to create agile low-cost data connections between the actors and the ability to control the sharing and use of their own data by each partner. The needs that our use case addresses are:

- Connections between farmer and consumer: Direct data connection between consumers and farmers, which leads to agile demand-driven production and increased awareness of food and its value.
- Connections between farmers/farms: Services that enable farmers’ organizations or groups to share data in a controlled matter for closer collaboration, e.g. for shared tasks or united market power.
- In-farm traceability despite diverse IT systems: Traceability of food and its journey from field to fork. Not only the traceability of material flow but also the traceability of data is important.

Partners/Ecosystem

The use case has been developed in Finland, and the consortium has 50 + members altogether along the food chain, and includes private companies, research organizations, agriculture cooperatives, and farmers union as core members. The consortium collaborates with AgriFood administration and authority services.

Main technology/Gaia-X components

AFDS uses the Fair Data Economy Rulebook model as its basis to agree upon common business, legal, technical, and ethical practices regarding the AgriFood data ecosystem.

AFDS focuses on the following aspects of the Gaia-X Architecture and Gaia-X Federation Services:

- **Data and Service Catalogue** are service functions to be implemented based on Gaia-X Federated Catalogue features where AFDS resources, assets and participants can be found by potential consumers and end users.
- **User Management and Authentication** is required for the participants in the AFDS ecosystem and for using the Gaia-X Federated Catalogue. These features are built using the Self-Sovereign Identity approaches within the Gaia-X Federation Services.
- **Data Usage Control** is built on the service functions that enable data transactions in a secure, trusted, and auditable way within the data ecosystem. We offer interfaces for the negotiation of data exchange contracts and usage policies for data assets. These functionalities will be compliant with Data Contract and Data Exchange Logging features of Gaia-X Federation Services.
- **Continuous Automated Monitoring** provides transparency to the AFDS members about the compliance of the individual services to common rules. The basis for this compliance is a combination of rules and requirements imposed by Gaia-X itself and AFDS, such as adherence to data policies, cyber security, data privacy or interoperability.

Concrete benefits

Benefits to farms

- Possibility to compose a tailored farm-specific Farm Information Management System from services and data sources within the data space.
- Combining data from diverse sources to calculate carbon footprint or other responsibility indicators for a sales batch of grain.
- Gaining higher price from farm products by connecting data-based product information to the sales batch when taking it to (electronic) markets.

- In-farm traceability of farm produce, even though farm processes are performed using diverse IT and automation systems.
- Possibility to participate in various AI developments by sharing farm data to developers and being able to test and implement solutions in an agile manner.

Benefits to all

- Possibility to create new connections between actors and their data sources cost-efficiently, e.g. among farmers’ groups, farmer – consumer, etc. to understand needs and business opportunities, leading to continuous improvement of each actor’s processes (production, marketing, consuming...).

Description Use Case: Agricultural Datasharing Platform (Belgium)

Solution

As a public private cooperation, DjustConnect offers a regional Flemish agricultural data space that

- (1) makes the data available for the whole sector in one data market,
- (2) ensures safe, transparent, and controlled transport of the data with one open Platform,
- (3) gives 100% control and respect to the farmer who, as the data owner with one dashboard decides when to share his data, with whom and why
- (4) has clear rules in accordance with the legislation and the Code of Conduct on agricultural data sharing in one contract

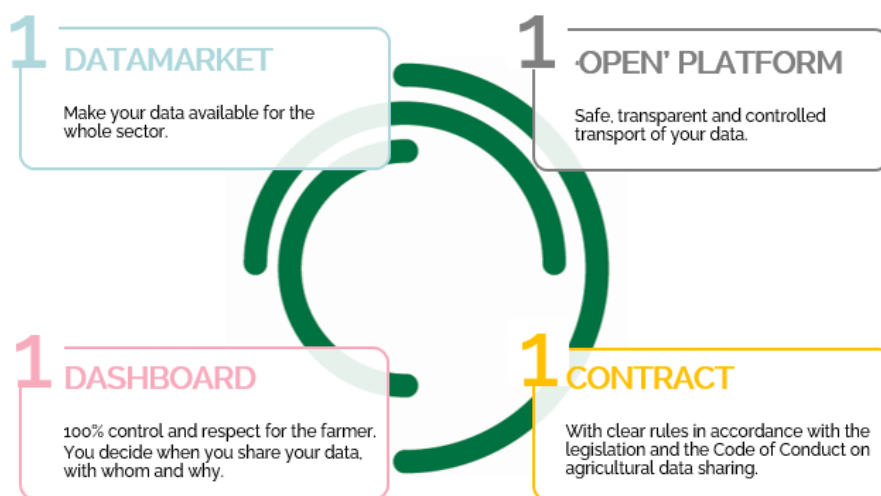


Figure 2

Problem solved

An agricultural business is a rich source of not only quality products, but also data. A lot of these company data are shared with others, such as the government, suppliers, buyers, producer organizations or cooperations. This data sharing has great advantages like smarter digital tools, optimal service or less administration.

However, the Farmer and company manager can't always keep a clear overview of who makes a request or who already has access to which part of his/her company data. All these requests are followed by complex contracts and agreements, which makes the administrative load of the agricultural business even bigger.

Data Providers, like supplier, buyer, producer organizations or agricultural cooperation's, have access to a lot of data. This is either their own company data or they manage the data from their affiliated farmers. So, it is easy to imagine how quickly you would lose a clear overview of all the data transactions/contracts made, with whom and why.

Because there are numerous data sources, it is not always easy for Data Consumers to find the right data for their product or service and to connect with the right data-owner.

Our use case:

- Makes sure that the whole AgriFood chain can be involved, including the small actors: farmers, food businesses
- Demonstrates a sustainable business model for data sharing and shows the way to the digital single market
- Builds trust via data sovereignty for the farmer as data originator
- Delivers data Portability and interoperability

Partners/Ecosystem

DjustConnect is a neutral platform available for everyone who wants to contribute to innovative agriculture through data sharing with respect for farmers. Respecting the rules is the only condition to get permission to use the platform.

DjustConnect grew from the research project 'Datahub for Agrofood'. With the financial support of the EFRO – European Fund for Regional Development and the 5 founding members (AVEVE, Boerenbond, CRV, DGZ and Milcobel) the state-of-the-art infrastructure was built. This infrastructure is now made available for the whole AgriFood sector.

The first applications have been launched. ILVO, in cooperation with the founding companies, is developing DjustConnect further to the needs of the users.

From the research project onwards, we want to make DjustConnect a stable, full public private, go to solution. Expansions, maintenance, and permanent supervision by ILVO are contained within a future plan which goes far beyond the duration of a research project.

Main technology/Gaia-X components

DjustConnect Reference Architecture model and its technological components

- focus on security
- aim to deliver a trusted platform, incorporating crucial building blocks
- prioritize on identity & trust management for authenticating communicating parties
- apply technical enforcement to shape trusted relationships between partners
- apply trustworthy communication based on authenticated and encrypted connections
- apply access & usage control

Specific architectural and technological choices of DjustConnect have been inspired by the International Data Space Association RAM.

- Many similarities exist especially on the Business and Functional layer
- The Business Layer specifies and categorizes the different roles which the participants of the International Data Spaces can assume, and it specifies the main activities and interactions connected with each of these roles.
- The Functional Layer defines the functional requirements of the International Data Spaces, plus the concrete features to be derived from these.

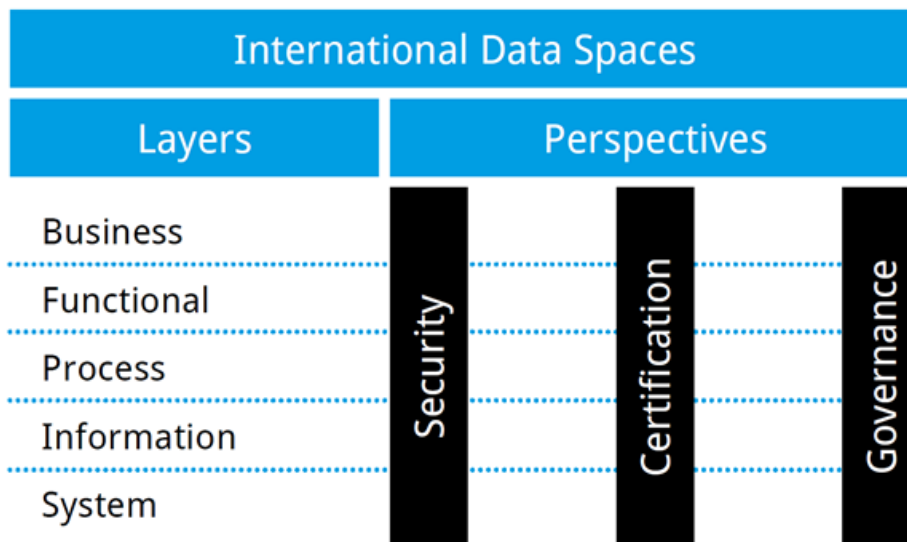


Figure 3

The next step, that will allow us to fully align with IDSA and Gaia-X design principles, is to use the Connector and adopt a more decentralized approach.

The adoption of REST binding by IDSA connectors, is a good step that will support our data providers to maintain the way they expose their data (RestAPIs).

Concrete benefits

Every actor, namely farmers and horticulturists, data providers and data consumers, benefits from the use of DjustConnect.

Farmers and Horticulturists get a lot of benefits by sharing their data, like smarter digital tools, optimal service or less administration. When connecting with DjustConnect:

- you decide who has access to your data,
- you keep a clear overview using your personal dashboard,
- you can change your mind at any time and retract your permission,
- data sharing leads to better advice tools and helps simplify administration.

As Data Provider

- you make your APIs available in our API shop,
- you reach more potential clients,
- you decide who you share with, the farmer decides whether the transaction will take place,
- all transactions are legally covered and replace your individual contracts with farmers and data consumers.

As Data Consumer

- you have a central point for all your data connections,
- you ask the farmer and data provider's permission in one click,
- after gaining permission you have access to data from the AgriFood sector,
- thanks to the shared data you can expand and improve your products and services.

Description Use Case: Data intermediaries: Consent and data exchange in agriculture (France)

Solution

Agdatahub is supporting the agricultural sector by implementing tools to ensure, that data sovereignty can be a day-to-day reality for French and European farmers, as the sector is an attractive cradle of data. Their vision of high-performance and sustainable agriculture, which is open to the world, is firmly anchored in the four pillars of innovation, community, European sovereignty, and fair governance. The use case is acting in creating value in agricultural chains by mobilizing data intelligence for healthy, traceable, and environmentally friendly food, and for technical advice tools to farmers. It gives operational advice on the structuring and exploitation of agricultural data, support, and as well as advice data scientists and marketing managers to communicate about their data offers.

Problem solved

Digital Identity

Digital identity makes it possible for:

- linking the identity of farmers (natural person) with the identity of their farm (legal person) to create a digital agricultural identity recognized by the whole sector.
- securing the information given by the different actors in the chain of consent, from field to plate
- simplify and speed up dematerialized transactions in complete confidence: CAP tele-declarations, contracts for the sale of their products to buyers of agricultural production (cooperatives, traders, industrialists, distributors, etc.).

The benefits of digital identity are the trust in consolidating the relationship between farmers and their partners in digital relationships, the safety in using blockchain technologies which allows to instantly check the origin and authenticity of the identity certificate. Also the digital identity allows the control to provide protection against the risks inherent in data in terms of operations, marketing and identity theft.

Farmer Consent

The main issue of consent is to enable farmers to regain control of their data by having a decentralized and multi-actor consent monitoring system.

This will allow farmers to list, add or revoke consent easily.

Associated with the digital identity device, the device created will allow the user to link his identity as a producer (person) with the identity of the operation (company). He will thus have a dematerialized portfolio in the cloud (wallet) which will be provided by Agdatahub and in which all identities are grouped and secure.

Access to this portfolio is then limited to authorized individuals only. Users will have access to the Agri-consent portal, where they will have a cockpit to manage their identities, authorized persons as well as digital interactions with their partners.

Data exchange platform

As a technical data intermediary solution Agdatahub provides API-Agro, an exchange platform, to offer a functional, technical, business and legal framework for data exchange between the different stakeholders: upstream industries, upstream agriculture, AgriFood companies, digital agriculture, financial services, professional organizations, public sector, education and research, using open and private data and APIs.

Digital identity and consent (in terms of authorizations) for farmers is another approach of Agdatahub. They are implementing a consent router to centralize consents from Farm Management System (FMS) and consent managers that have been developed in the

ecosystem. The router is going to be interconnected to the data exchange platform to check if farmers have consented to share their data executing the exchange.

Partners/Ecosystem

Agdatahub uses its expertise in operational consultancy to support farming industry actors in their exploitation of data and shares its knowledge in different projects. Agdatahub is currently planning to create a dashboard from which the farmer will manage the authorizations that he issues to his employees or partners to access the plan protection products room or use robotic equipment.



Figure 4

Main technology/Gaia-X components

As data intermediary using the same infrastructure as Gaia-X's, the platform can create significant synergies and added value for the entire agricultural industry. Further, both enable a visibility for the digital potential within the domain. The aim is to build a community that acts collectively by collaborating in the development of tools that facilitates the provision, sharing, exchange and interconnection of data in their territory. Gaia-X is conceived to assure data sovereignty. The user decides for himself, in every step.

Concrete benefits

Agdatahub pushes for innovation in the agricultural sector by enabling an easy and secure access to data. This stimulates innovation among industry and digital actors, and anticipates, predicts, and transforms risks into opportunities. It is important to ensure equitable access to data for all stakeholders in the European agricultural ecosystem.

Respecting the privacy by design principles, all solutions have been implemented to ensure high-security level with Know Your Customer (KYC), data encryption and full traceability of all actions. Furthermore, the Agdatahub products have integrated the recommendations from the EU Code of conduct on agricultural data sharing by contractual agreement to empower farmers on their own data governance.

The principles of security by design and privacy by design of Gaia-X added value to the ecosystem and the guaranty of the highest security requirements and the protection of privacy.

Description Use Case: Agri-Gaia an AI ecosystem (Germany)

Solution

The Agri-Gaia project (<https://www.bmwi.de/Redaktion/EN/Artikel/Digital-World/GAIA-X-Use-Cases/agri-gaia.html>) is funded by the German Ministry of Economic Affairs and Energy. It aims for an AI ecosystem for the agricultural and food industry based on Gaia-X. For this purpose, an innovative B2B platform will be implemented, which provides industry-specific adapted AI building blocks as easy-to-use modules and brings together users and developers of AI algorithms in the AgriFood domain. Agri-Gaia closes the circle from sensor data acquisition on the agricultural machine, training of the algorithms on appropriate servers and continuous updating/optimization of the algorithms. Appropriate interfaces and standards are being developed so that a manufacturer-independent infrastructure for the exchange of data and algorithms is created.

Problem solved

AI development in the AgriFood domain is often a B2B process. Even bigger companies in this domain do not have in house AI knowledge and are forced to cooperate with AI companies in order to integrate AI in their products and processes. Agri-Gaia builds an infrastructure for these cooperation processes in order to ease the B2B AI development in the AgriFood sector.

Partners/Ecosystem

Project partners are:

- AgBrain GmbH
- Agrotech Valley Forum e.V.
- Amazonen-Werke H.Dreyer GmbH & Co. KG
- Robert Bosch GmbH
- CLAAS E-Systems GmbH
- DFKI GmbH
- University of Applied Sciences Osnabrück
- Josef Kotte Landtechnik GmbH
- Maschinenfabrik Bernard Krone GmbH & Co. KG
- LMIS AG
- Robert Bosch GmbH
- University Osnabrück
- Wernsing Feinkost GmbH

Additionally, about 30 associated companies from the AgriFood domain are involved. The ecosystem itself addresses the entire AgriFood domain.

Main technology/Gaia-X components

The Agri-Gaia ecosystem will be built based on the Gaia-X infrastructure, which ideally meets the requirements of the industry and the Agri-Gaia ecosystem in terms of data sovereignty, decentralization/multi-cloud and edge support and service delivery. A Gaia-X federator will be established that hosts the Gaia-X Federation Services. All participants in the Agri-Gaia ecosystem use these services for identity and access management, sovereign data exchange or to find other participants and services using the Gaia-X Federated Catalogue.

Concrete benefits

Agri-Gaia defines standards for AI development in the AgriFood domain and enables interoperability of data, trained models and edge devices. It eases the process of getting the needed data for AI development and B2B cooperation for AI development in this domain.

Description Use Case: Knowledge Graphs for data integration in the AgriFood sector (Poland)

Solution

Farm management is a complex process that involves multiple activities carried out by farmers and other stakeholders, who have to manage multiple and heterogeneous data sources collected and generated through various applications, services and devices. The rapid advances of IoT technologies, AI and Big Data, among others, have boosted the adoption of smart farming practices, which emphasizes the use of ICT in the farm management cycle to exploit the available data.

The explosion of data availability, however, has led to new challenges. Data is usually available in different sources, in different formats, and represented according to different models, thereby hampering data interoperability and integration. The lack of integrated data access, in turn, hinders the full potential of value creation based on all the available data, and the development of smart services and applications supporting the decision-making processes. Thus, a key challenge to realise the smart farming vision to its fullest is to combine/integrate those different and heterogeneous data sources in order to support the decision-making processes.

Knowledge graphs provide a flexible and efficient solution to address some of those challenges. In particular, they can provide an integrated view over (initially) disconnected and heterogeneous datasets, through the interlinking of different entities, based on Linked Data principles, and in compliance with any privacy and access control needs.

Problem solved

The following problems are solved by this Use Case:

- Lack of/limited access to (semantically) integrated data, hindering the full potential of value creation based on all the available data
- Limited/difficult data sharing and reusability in the AgriFood sector, including a rare adoption of FAIR principles
- Hidden valuable knowledge in logically related data (e.g., relations between data elements) that could be discovered through the established links and the possibility to exploit semantic inferencing and queries
- Interoperability issues between different services/applications relying on different data models and semantics, which is a key challenge in the AgriFood sector due to the numerous solutions available from different vendors/providers.

Partners/Ecosystem

Ecosystem partners include: agriculture organizations (e.g., advisory centers), AgriFood solutions providers, infrastructure providers and semantics experts. This use case will deliver knowledge graphs providing an integrated view over different and heterogeneous datasets, which can then be exploited by the agriculture organizations or AgriFood solution providers to deliver more efficient and innovative services and applications to the final end-users (e.g., farmers, advisors) supporting their decision-making processes (e.g., benchmarking, yield optimization, improved precision fertilization, etc.). Infrastructure providers may be also involved in order to enable the storage and processing of large amounts of data.

Main technology/Gaia-X components

Main technologies include:

- Standard and/or well-established ontologies/vocabularies in the AgriFood field providing the underlying model to represent data in the knowledge graphs. Such models should be easily extensible to support a variety of data and final applications. One example for this is the DEMETER Agriculture Information Model (AIM) which provides a common vocabulary to enable different applications to interoperate and exchange data that can be understood.
- Linked Data Pipelines services/tools to facilitate the transformation/translation and publication of different datasets as Linked Data. Such pipelines should be easy to reuse, reconfigure and adapt to handle different types of data and formats. This will also include possibility handle both mostly static data and highly dynamic data, where the latter will require the creation of virtual graphs over the original data sources. Additionally, the goal of such pipelines is to automatize as much as possible the linked data generation process, abstracting the different interfaces and specificities of the different tools supporting specific aspects of this process.
- Semantic databases (triple/quad stores) that can easily scale and handle large amounts of data. Such stores should also support efficiently geospatial features to access and query data based on their geographical location/extension, which is very relevant in the AgriFood sector.
- Easy to use and customize APIs to access the integrated data in knowledge graphs.

The following Gaia-X Core Services would (potentially) be required for this Use Case: Federated Catalogue, Data Connector, Self-Description and related Compliance, and optionally IAM. Regarding the Self-Description, a Metadata Schema for the AgriFood Domain will need to be defined, potentially reusing, if possible, existing schemas that are relevant for this domain like the CYBELE common semantic model.

Concrete benefits

- Access to integrated data in the AgriFood sector that could be exploited by different services/applications to support the decision-making processes of relevant actors (farmers, advisors, public institutions (e.g., dealing with CAP), etc.)
- Improved data findability and accessibility by both humans & machines, which can foster and enhance their reusability, in alignment with FAIR principles
- Discovery of new knowledge through the established links and the possibility to exploit semantic inferencing and queries
- Federated data access over multiple sources (endpoints) via federated queries, using a single-entry point
- Improved interoperability (of services/applications) with standardized interfaces and semantics (via the underlying ontologies/vocabularies) to use data in knowledge graphs
- Improved data sharing in compliance with the security/privacy constraints to access the underlying data.

Description Use Case: Interoperable data exchange for online task management, machine tracking and sensor communication (Belgium)

Solution

Online task management

For farmers, equipment operators and agricultural software users, online task management is a solution for real-time data exchange of historic and ongoing tasks in the field, between existing data platforms belonging to several mixed-colour connected machines. The data consist mainly of static data of field activity (done, or to be executed). The solution will support the main agricultural data exchange for managing and reporting activities done in the field.

Machine tracking

An online standard interface between existing telemetry solutions or portals (cloud solutions) allows farmers/contractors to see their whole mixed machinery fleet in one system of their choice unlike today where they are forced to use several OEM portals for getting an holistic overview of their machines in action. For the initial implementation, focus should be placed on limited data sets (dynamic, pseudo real-time, live data), with a high frequency (from 1 to 5 minutes updates or less).

In-field sensor communication

For farmers, monitoring field properties from different data platforms is a solution for real-time infield data analysis including weather data, topography of fields, crop vegetation indices and soil conditions. To this end, this use case sets the foundations for communicating data across different platforms and IoT devices such as infield sensors with the agricultural software solution of choice.

Problem solved

Online task management

Each agricultural machine platform and software supports different APIs. Not all machines connect to all software. Implementation takes huge amounts of time as each connection is custom. Not all functions of the APIs are equally supported by all data players. A standardized API would support software developers, increase quality and open new opportunities for data services. The farmers will have more tools at their disposal resulting in increased value creation for the agricultural world.

Machine tracking

Today each equipment manufacturer has its own telematics platform. Farmers operate different brands of vehicles; they need to consult multiple systems in order to have a complete overview of their running fleet. This is not practical and as such hindering the value creation on real time data such as logistics optimization. Near real-time communication of data coming from moving vehicles in a standardized format will allow visualization of different sources in one integration dashboard.

In-field sensor communication

There is no standardized API to pull data from platforms used in different verticals in agriculture. This should be a universal solution for cross-functional data exchange between different sectors.

Partners/Ecosystem

Views of the personas involved in the use case:

- **Equipment Operator:** I want to exchange task sets and fieldwork documentation with multiple agriculture software solutions to be clear about what fieldwork shall be done and to document what fieldwork has been done.
- **Agricultural software user:** I want to exchange task sets and fieldwork documentation with one or more equipment of a mixed fleet to be able to steer what fieldwork shall be done and to view what fieldwork has been done.

- **Equipment Owner:** I want to be able to manage which of my equipment is usable for the online task management use case in a mixed fleet, want to be able to share telemetric data of my machine(s), so that I can use the software solution of my choice for tracking and want to be able to share sensor data of my infield sensors to several data consumers (agricultural software solution of my choice).
- **Contractor:** As contractor, I want to be able to use any available equipment in my mixed fleet, to be able to steer and document the fieldwork I am paid for and I want the farmer to share telemetric data of their equipment with me, so that I can track progress of the mixed fleet in the software solution of my choice, while doing work on their field(s).
- **Farmer:** I want a contractor to share telemetric data of their equipment with me, so that I can track progress on my fields with the software solution of my choice and I want to be able to share sensor data of my infield sensors with the agricultural software solution of my choice.

Main technology/Gaia-X components

There are different components required to support this use case: sign-on, discovery, standardized communication, data model, field master database, machinery master database, data privacy/ownership/retention rules and contractual agreements to support all of this.

Concrete Benefits

Key benefits for the end user include:

- a) Easy reporting process: Task orders are planned in the software system and are exchanged with the machines. Results (documentation) are then transferred back to the agricultural software of choice.
- b) Reduced time and effort for managing fieldwork.
- c) Real-time documentation of task orders/ task results.
- d) Real-time exchange of infield information with service providers
- e) Informed decision-making leading to optimization of infield operations.

Having a dashboard linking to IoT platforms of different vehicle manufacturers allows real-time optimization using multi-branded fleets, amongst other functionality. With this use case, farmers are supported in fleet management in grain and forage harvest. A new area of development and benefits for farmers is expected in farm and harvest logistics operations.

Description Use Case: Connected Potato data chain (Belgium)

Solution

Allowing a direct data flow between the smart potato harvester, the farm management system and the potato processing company by connecting the AVR connect platform on the machine with the WatchItgrow decision support system and with the Agristo ICT system using the Flemish data platform DjustConnect as a data intermediary. The data can flow, from the machine to the farmer and the processing business with full control of the farmer, as data owners.

Problem solved

Delivering the needed authentication, consent management and enable the data flow in the potato supply chain.

Partners/Ecosystem

The use case partners are AVR, manufacturer of a full range of potato production systems from soil preparation to harvesting and sorting; Agristo NV, a Belgian producer of pre-baked potato-inspired frozen products in the private label market; VITO an independent Flemish research organization in the area of cleantech and sustainable development and ILVO, Flanders Research Institute for Agriculture, Fisheries and Food.

Main technology/Gaia-X components

Services for identity and access management, sovereign data exchange and services using the Gaia-X federated catalogue.

Concrete benefits

To deliver trusted data sharing to the farmers and the agribusiness, with respect for the data owners and a sustainable business model. Having an architecture to make an international setup possible and to take the next step towards creating interoperability between all machinery at farm level.

Education

Our vision for a European Education & Skills Dataspace

This position paper for an "Education & Skills" Dataspace is built upon four priorities:

Strengthening France and Europe's role in digital economy in education

Education is facing a digital revolution and an exponential growth in digital technology use in education, training, and lifelong learning, requiring massive digital skills development within our societies. In 2019, United States and China accounted for two-thirds of the seven billion euros of EdTech fundraising. In Europe, total EdTech investment has just exceeded 1 billion euros, mainly benefiting France, UK, and Nordic countries. Global market for learning management system (LMS) provides an enlightening example of the economic challenges involved: this market passed the USD

13.4 billion threshold in 2020 and is estimated to grow up to USD 25.7 billion by 2025, setting the scene for major developments, driven by the Covid crisis, such as the combination of distance learning processes with personalisation through artificial intelligence (AI) and the implementation of interoperability through the API'sation of services.

Digital education market is currently dominated by Anglo-Saxon and Chinese players: Gaia-X initiative marks **an alternative model to data handling practices, based on a European digital vision**, and makes it possible, and within reach.

Carrying out an ambition which is politically and strategically consistent with the European vision

Our ambition is in line with both the European strategic position for a digital Europe and the European Action Plan 2021-2027 for digital education.

An Education and Skills Datahub will enable to:

- Strengthen EU's critical digital capabilities by focusing on key areas such as artificial intelligence" [...] and "testing and adopting trustworthy AI technologies».
- Pursue an ethical use of artificial intelligence (AI) and data in education and teacher training, by supporting research and innovation under Horizon 2020 programme
- Implement Cloud-to-Edge solutions and promote sectoral data spaces, especially high-value data sets.
- Enable access to large repositories of high-quality data for public and private actors aiming to increase their productivity and competitiveness, as well as pursuing societal improvements in terms of " public services" [...] and "well-being" (for students, teachers, employees, and citizens)
- Foster cross-sector collaborations and new models to enable the exchange of digital learning services and contents, based on common standards, interoperability, accessibility, and quality assurance.

Moving towards a sovereign environment

We fully support the Gaia-X strategy which aims to impose rules leading to a **sovereign cloud environment**, based on existing solutions, but **giving data owners a full control**. That is, to:

- depend on a European jurisdiction
- decide on data location;
- decide who can process data, and for what purposes.

This data owner control needs to be built upon a standardised architecture and governance mechanisms for service providers and users. Regulation on transparency, interoperability, decentralisation, and trust issues will guarantee:

- sovereign, federated, interoperable, public or private, hybrid cloud infrastructure services, supported by current operators and secured from end to end
- **data flow** and access to AI services within the Education & Skills Dataspace

Data Governance Act (November 2020 version) emphasises the principle of subsidiarity to facilitate data flow within cross-sectoral and EU-wide value chains, and of a harmonised legislative environment. This Act establishes a system of interoperable consent, through independent data intermediaries, strengthening data-sharing mechanisms within the different European data spaces.

We intend to seize the opportunity to **boost education data sharing in a trustworthy environment**, respectful of European regulations and national laws; we carry out the vision of **regulated EdTechs relying on an appropriate levels of data governance**. Gaia-X initiative can accelerate the development and the use of reliable data-driven services for education, study and vocational guidance, and labour market. We expect Gaia-X management to establish the broadest governance as possible, to implement a trust-based framework for cloud solution providers and to unlock potential services in education.

We believe that this global strategy, open to all market players, is the most effective way of building on existing solutions, installing trust, and accelerating the development of new services that will enable a comprehensive European offer equal, or even superior, to major non-EU players, in a sovereign, secure and ethical environment.

Responding to our societies' needs

An enhanced dataflow in an ethical, secure and sovereign environment will provide the right framework to develop innovative solutions, setting the stage for **major transformations of educational systems** for the benefit of **the entire educational community**:

- for students and families:
 - a personalised learning environment, adjusted to individual strengths and needs.
 - specific educational pathways, suggested activities, or resources, to maximise individual learner's outcomes.
 - systemic support in the guidance process throughout school and professional life.

- **for teachers**, to be relieved from low value and time-consuming activities, having the possibility to reinvest their time in pedagogical differentiation, thanks to reinforced knowledge of their students' specific needs;
- **for decision-makers**, for a data-driven management of the education systems, exploiting massive data AI-based analysis for evidence-based decisions, evaluating practices and modelling future developments at the French and the European levels;
- **for EdTech companies**, to be able to assess the impact of their solutions and to better identify users' needs, in order to offer innovative and effective products and services, more rapidly, and for a wider community;
- **for researchers**, in all disciplines related to education and guidance, for a better understanding of all the mechanisms related and for evidence-based practices and policies.
- **for future citizens**, to be able to:
 - Identify their skills more accurately;
 - Benefit relevant training and employment opportunities adjusted to labour market demand.
 - Have a single access point to local offer in terms of employment and training
- **for universities, colleges and professional training organizations**, to offer more relevant training in a lifelong perspective and, to students, an easier European and international mobility;
- **for hiring companies**, to increase the accuracy of offer-demand matching, and to facilitate recruitment based on skills-development (internal mobility between subsidiaries and head office, for example);
- **for local communities**, to be able to steer their investments in vocational training, boosting local employment and enhancing local skills development.

Gaia-X is going to provide **a strong commercial advantage for European Edtechs**, in respect of the ethical values advocated by the French and the European community, in an environment of trust that will help companies to work, to create value-added products and services and **to fulfil the promises of AI for learning**.

Until now, we, most of the stakeholders in the French ecosystem (service providers, digital publishers, administrations and EdTech companies) didn't have a trusted cloud service and a data space allowing data sharing between cloud platforms, national or international (operating under extraterritorial laws). **This reality hinders the use and the acceleration of data-driven services** such as advanced student and teacher support services or business intelligence tools. A trusted and sovereign (European) Cloud would support all European educational actors in order to remain competitive and, progressively, gain bigger market shares in digital education.

Thanks to a strong political impetus around Gaia-X in recent months - and we would like to pay tribute to the founding members, as well as the French and German governments – this initiative represents **a tangible way forward to:**

- build, assemble, and use reliable and added-value data-based cloud services,
- create new products/services and a new offer on the market,
- develop a business model that complies "by design" with European regulations and values.

We consider Gaia-X as an accelerator of innovation on a European scale, extending open innovation and co-construction by providing secure data sharing and artificial intelligence services at scale, in line with compliance and security issues.

Combining the strengths of Edtechs, governments and research entities, we aim to create an education and skills data space within Gaia-X, which is going to foster Europe's competitiveness and to spread of European values improving, simultaneously, research, AI development and innovation, in a legal stable environment.

We have described below several use cases to present some practical examples and achievable scopes of action in order to guide future choices of funding by the European Commission.

For the time being, these use-cases have been prioritized with the French digital education actors. However, we have the firm intention **of sharing them with other European members to gather several countries around this same vision and its overall goal of improving education for all.**

Targeted challenges 2021-2023: list of use-cases

EDUCATION AND GUIDANCE

- Exploring AI and data in education and guidance, with researchers:
 - Modelling trends and guidance pathways to improve tailor-made counselling*.
 - Developing solutions to detect learners' shortcomings so as to offer personalised contingency strategies
 - Developing systems to detect signals leading to school drop-out in order to offer personalised tutoring
 - Promoting the use of adapting learning in education products and services
 - Developing voice assistants and language learning tools for all CEFR levels and for allophone students*
- Facilitating individual and class monitoring by teachers and educational teams
 - Strengthening the security of on-line proctoring (remote proctoring, fraud detection etc.) and developing anti-plagiarism devices
 - Improving the assessment of digital solutions through harmonised key criteria description and impact measurement*

- Fostering national and international comparisons between educational systems, for example the strategies for digital skills development
- Strengthening information for families, supporting their assistance in digital use, in a school-family continuum

SKILLS DEVELOPMENT

- Tokenisation of skills data sharing*
- « Blockchain as a service" trust platforms (in line with the European Commission's work on EBSI): trust hosting for blockchain nodes, facilitating the execution of smart contracts, legal translation of foreign diploma and diploma recognition in Europe.
- Personal education files portability for pupils, teachers and citizens

GOVERNANCE

- Developing a "compliance by design" framework implemented by Gaia-X governance and the Education & Skills Data Space*
- Quality labelling for innovation and experimentation platforms, meeting the Gaia-X Standards, to support co-innovation and deployment of trust services.
- Launching a consent and ePrivacy management programme

TECHNICAL ISSUES

- Interoperability and/or Ontology to ensure data flow
- Development of a one-stop-shop programme for data sharing for research purposes*.

* Six use-cases are described in section 4 below. Other use-cases' description is in progress.

Ecosystem partners & stakeholders in France

- [Ministère de l'éducation nationale, de la jeunesse et des sports](#) (MENJS)
- [Académie des technologies](#)
- [Association française des industriels du numérique pour l'éducation et la formation](#) (AFINEF)
- [Business France](#)
- [Club informatique des grandes entreprises françaises](#) (CIGREF)
- [Direction générale des entreprises](#) (DGE)
- [EdTech France](#)
- [France Éducation international](#) (FEI)

- [Institut national de recherche en informatique appliquée \(INRIA\)](#)
- [Office national d'information sur les enseignements et les professions \(Onisep\)](#)
- [aNewGovernance \(aNG\)](#)

(To be completed)

Use-case 1: Exploring AI and data in education and guidance, with researchers

Example 1: Modelling trends and guidance pathways to improve tailor-made counselling.

- Problem:

Guidance practices should be based upon efficient personalized systems: each path, in studies or professional life, is a unique individual experience. Best guidance requires using a wide data-set (local employment opportunities, personal skills and aptitudes, etc.). Most of the recommendation systems used nowadays do not exploit enough data to implement effective systems.

- Solution:

Federate guidance and skills data (student portfolios, badges, extracurricular data) to identify information on pathways, needs, skills and to build intelligent and predictive models.

- Difficulties/barriers

- Lack of infrastructure allowing data-flow, data interoperability, legal and consent management (use-case n°9)
- Lack of standards to share skills data (use-case n°7)
- Lack of ethical framework on predictive models in guidance
- Legal Framework

- Partners/Ecosystem/Synergies:

Schools, Universities, State agency for guidance (ONISEP) and employment (Pôle Emploi), Edtech, local communities, etc. Possibility to capitalize on existing projects as Lyli project (University of Cergy-Pontoise, Académie de Versailles) or European/international projects gathering guidance stakeholders.

- Main technology / Gaia-X:

- Openness and transparency through the identification of established regulatory and data protection criteria and the guarantee of data control over use.
- Use of existing reference architectural principles, particularly those promoted by the International Data Spaces Association (IDSA).
- Interoperability at three infra-structural levels: interconnection, technical and semantic, for network, data and services.
- Facilitate intra- or inter-domain specific data exchange or data-services link beyond vendors and customers' possibilities

- Security and data sovereignty management within a framework based on Gaia-X principles certifying providers, nodes and services, particularly with regard to technical and organizational criteria. The certification will be proven by an independent and trustworthy third party and be based on already established audit and certification procedures (e.g., Minimum standards for external cloud services by BSI, C5, ISO 27001 et Trusted Cloud).
- Expected benefits:
Extend guidance and recommendation possibilities throughout the European territory, and beyond.

For users

- A more precise and targeted guidance advice.
- Better visibility on possible pathways according to anyone’s specific profile.

For institutions and companies

- To propose innovative, sovereign, and trustworthy offers in guidance and assistance
- To promote a know-how on international level.

Example 2: Developing voice assistants and language learning tools for all CEFR levels and for allophone students

- Problem:
Speech recognition quality is progressively increasing and pinpoints real “conversational” opportunities for language learners wishing to develop their speaking skills. At the same time, text recognition, errors analyses, and adaptive evaluation allow language training to be adjusted to each learner’s strength and weaknesses, language competence being a global competence, measuring know-how in different competence areas, as recommended by CEFR. Innovative solutions exist but are struggling to develop because of insufficient language corpus to be trained upon, with a multilingual dimension, like children’s voices pour example. Many researchers or R&D actors in automatic language processing, for instance, have major difficulties in working on most of languages, due to absence of significant data. Sometimes, although existing, data is unavailable because of data owners’ difficulties in releasing (GDPR compliance, voices being personal data, children voices being even more protected).
- Solution:
Use the infrastructure and framework offered by Gaia-X to promote data liberalization and availability and to reassure all actors on regulatory compliance (GDPR, consent, parental authorization etc.) essential to “open” these data.

Main aims pursued:

Voice command (NLU for Natural Language Understanding)

- recognition of instructions given to an application, either for control (enter a command) or for information input (answer a question)

- cleaning the sound message to facilitate understanding in noisy environments.
- voice transcription in applications (direct voice input into a word processor with good quality transcription, to reflect words intelligibility)

Speech analysis (to analyse fluency, pronunciation, grapheme-phoneme correspondence, phoneme correspondence etc.)

- algorithm development according to CEFR levels and mother tongue
 - enrichment of the specialized APIs as SoapBoxLab
 - facilitation of API development by EdTechs
- Difficulties/barriers
 - Difficulty of collecting and processing data (children’s speech data, for example) to develop effective models for language learning.
 - Lack of a sovereign mutualized and secure infrastructure to assemble language corpus and to ensure data flow.
 - Lack of procedures to create language corpus and to deal with their governance.
 - Legal Framework
 - Contractualisation and consent management
 - Partners/Ecosystem/Synergies:

Schools, EdTechs, language testers, universities, teachers, cultural cooperation networks, language centers, individuals

(to be completed)

- Main technology / Gaia-X:
 - A transparent, regulatory, and protective framework facilitating data release and control over data use.
 - Interoperability at three infra-structural levels: interconnection, technical and semantic, for network, data, and services.
 - Structuring governance to ensure control on data use, particularly for R&D purposes.
 - Security and data sovereignty management within a framework based on Gaia-X principles.
- Expected benefits:
 - Reference language corpus and governance procedures federating all potential data owners.
 - Emergence of new adaptive and personalized language solutions offers.

- Promotion of European multilingualism and plurilingual language skills.

Use-case 4: Improving the assessment of digital solutions through harmonised key criteria description and impact measurement.

- Problem:

Digital education offer, public and private, is extremely rich and heterogenous. Digital demand, for products and services, is as specific as the offer. COVID-crisis has once again shown the urgency to capitalize on what exists, to make education systems more resilient and to preserve pedagogical continuity. Education practitioners and stakeholders (teachers, but also educational managers), and families, have real expectations of offer qualification, in terms of educational impact and targeted skills. On their side, solution publishers are struggling to gather sufficient knowledge on the impact of their products, as they have no visibility on extra-curricular continuum of each student. Any data use for a specific solution can't provide enough information to assess the impact in terms of quality of learning, without a more holistic view on on-line activity and learning "traces" (for students, for teachers etc.)

- Solution:

Gather data on students' learning process, collected throughout all the solutions and services used, in a school & extra-curricular continuum, to analyse the impact of each solution and identify room for improvement.

- Difficulties/barriers

- Lack of infrastructure for dataflow and lack of interoperability
- Consensus to be found on standards to describe digital solutions for education, and skills that can be developed (use-case n°13: "interoperability, ontology")
- Legal framework (learning analytics) and consent management

- Partners/Ecosystem/Synergies:

Schools, universities, state administration and operators (CNED, CANOPÉ, FEI) Ed techs.

- Main technology / Gaia-X:

- Establishment of a regulatory and protective framework allowing data circulation and use.
- Security and sovereignty management
- Supra-national framework allowing offer standardization and, ultimately, its evolution towards better quality for users.

- Expected benefits:

- Provide the entire educational community a better knowledge on digital education offers and recommendation systems for the most suitable resources and services.
- Boosting a European offer of digital educational solutions, based on European values, guaranteeing data valorization and protection.

- Develop European EdTech offer towards interoperability and educational quality criteria, making it more competitive on the international scene.
- Empowering Europe to be more resilient, in case of crisis, and to support other countries in its international cooperation action.

Use Case 7: Tokenization of skills data sharing

- Problem

Today, there is no platform working as a single point of entry, aggregating European citizens' skills: each training organization, employer, and higher education institution proposes its own reference system. Moreover, this data on competences is fragmented between the actors of education, training, employment and guidance without any possibility to easily identify, circulate and recognize these competences. This leads to delays and lack of relevance of recommendations and policies.

Many innovations have aimed to equip individuals with competency portfolios based on "open badges" that allow the individual and organizations to view competences from different sources. However, these experiences have not been widely adopted for several reasons:

- **Interoperability:** A major change in the stakeholders' information systems is needed to accommodate open badges, it is therefore important to implement an automatic way to manage this interoperability.
- **Recognition:** Open badges are not governed or officially recognized; it is necessary to set governance and data traceability to ensure that each stakeholder can contextualize and have confidence in the data they receive
- **Data use:** there is no incentive to share and circulate open badges under the person's control. It is important to develop access management functionalities, nevertheless the means, the solution or the place of data storage.

- Solution

Produce different mechanisms based on artificial intelligence, usable by all stakeholders, upon skills data:

- competence "extractors": algorithms capable of extracting competences from raw data on the basis of common reference systems (ROME/ESCO/RECTEC...);
- Inter-repository translators: algorithms capable of semantically linking competence repositories together and linking them to central repositories;
- skills- tokenizers: based on the European blockchain (ESBI), all the mechanisms and standards for recognizing and tracking badges and skills validation.

- Barriers

- Excessive data spread: skills data are not always explicitly described and can be scattered over several storage locations and between different data owners.

- Lack of the right amount of data to make the appropriate correlations between ontologies and skills' profiles.
- Data space should provide sufficient data to train algorithms properly.
- Partners/Ecosystems/Synergies
OpenClassrooms, Simplon, Paris 1, University of Lille, Université Catholique de Lille, Hauts- de-France, Provence Alpes Côtes d'Azur, University Cergy-Pontoise, Academy of Versailles. And internationally: Europass, Tampere university, Frankfurt Universities, Finish Defense Forces, Swedish Employment Agency.
- Main technology/components Gaia-X
 - Sovereign data exchange
- Concrete advantages/expected benefits

This type of solution will enable education, employment, training and guidance stakeholders to:

- Determine European citizen's skills;
- Be able to align their respective competency frameworks to promote matches and thus to accelerate training and return-to-work trajectories.
- Implement lifelong learning and recommendation process, based on better defined data

Use-case 10: Developing a "compliance by design" framework implemented by Gaia-X governance and the Education & Skills Data Space, based on a Code of Conduct, and sovereignty and security conditions.

Problem:
European education community (parents, teachers etc.) is strongly concerned about personal data protection and data use in a sovereign and ethical framework. In order to overcome any psychological or behavioral barrier and to implement a Data Space at European level, it is necessary to create a trust framework and new opportunities of products and services, based on educational data.

Solution:
GDPR and Data Governance Act, together with the idea of a European sovereign cloud, within Gaia-X community, offer countless perspectives to reconsider place and role of data in education systems.

Our strategy is based on a sovereign and trust framework built at European level with education and lifelong learning sectors, relying on a European **"compliance by design" framework** put in place by Gaia-X governance and adjusted, by the Education and Skills Data Space members, to the digital education ecosystem needs.

This "compliance by design" framework requires to use a next generation linked data infrastructure that meets the highest requirements in terms of digital sovereignty and that

promotes innovation based on European values, such as those provided by Gaia-X initiative, related to a **Code of Conduct** of EdTechs. This Code represents a GDPR transposition to digital education sector; it is going to be proposed by EdTechs to European authorities, to give companies precise compliance guidelines and to foster a unique vision, for all European EdTech providers, on the conditions for deploying digital products and services in education.

The Data Governance Act, which is currently being finalized, is intended to enable the implementation of new rules via a common data sharing infrastructure. Some of these rules can be developed in a cross-sectoral basis, others are sector-specific. This “compliance by design” framework will allow to experiment these rules and to reinforce learners’ data security and safety.

In general, the building blocks of the future European regulation are:

- Consent management: common infrastructure and tools for managing consent on data sharing
- Identity management: mechanism linking different identities together to ensure correct identification of individuals within different services
- Governance management: infrastructure to automate compliance with governance, ethical, code of conduct and regulatory rules on the flow of education, skills, employment and training data
- Contract management: mechanism and infrastructure for generating and tracking contracts for data exchanges
- Data interoperability management: APIs and data interoperability standards
- Trust, certification and security management: certification and authentication infrastructure for organizations exchanging data
- Management of value exchanges on data: management and traceability of monetary or value exchanges linked to data exchange.

This future “compliance by design” framework will have to be followed by an implementation strategy, the adoption of sectorial standards and architectural basis to put these requirements into practice, creating products and services, in the education and skills area, and make them available for all users.

- Difficulties/barriers
 - Consensus to be found on a common Code of Conduct, and adoption by EdTechs in EU.
 - Adapting the compliance by design framework to the needs of the whole digital sector, coherently with the European data governance framework (Data Governance Act)
 - Clarify and promote the possibilities offered by data flow within a trust infrastructure.

- Partners/Ecosystem/Synergies:

European Ed techs, state administrations, G29, a New Governance.

- Main technology / Gaia-X:

- Openness and transparency through the identification of established regulatory and data protection criteria and the guarantee of data control over use.
- Use of existing reference architectural principles, particularly those promoted by the International Data Spaces Association (IDSA).
- Interoperability at three infra-structural levels: interconnection, technical and semantic, for network, data and services.
- Facilitate intra- or inter-domain specific data exchange or data-services link beyond vendors and customers' possibilities
- Security and data sovereignty management within a framework based on Gaia-X principles certifying providers, nodes and services, particularly with regard to technical and organizational criteria. The certification will be proven by an independent and trustworthy third party, and based on already established audit and certification procedures (e.g., Minimum standards for external cloud services by BSI, C5, ISO 27001 et Trusted Cloud).
- Expected benefits:
 - Capitalize on work already done to develop a Code of Conduct with French EdTechs
 - Transparency and common rules for all European states, based upon Gaia-X technical and conventional basis.
 - Boosting a European offer of personalized digital educational solutions, based on trust and consent, guaranteeing data valorization and protection.

Use-case 14: Development of a one-stop-shop programme for data sharing for research purposes.

- Problem:
 - Research laboratories need to access to dataset to work efficiently, ethically and transparently. Obstacles lie in the difficulty to know whether data exists or can be aggregated, to appreciate their degree of quality, to understand how to use them and under which conditions.

- Solution:

The creation of a platform, gathering educational and skills data, would support most of research programs, facilitating data collect and offering the possibility to train super-algorithms (prospective, AI etc.). To summarize, the platform should:

- Collect data across the whole education and guidance ecosystem.
- Make it available to European actors (publics or private) engaged in R&D and in improving educational innovation
- Clean datasets and transform them in reference data
- Identify, define, document, and share open standards to facilitate data exchange.

In our vision, this platform is not entirely public: part of the data will be open data, other part will be available for interested parties respecting appropriate entry criteria.

Different access modalities can be imagined:

- Restrictive, on selected ad hoc data (like for Health Data Hub)
- Open, with a special profile for licensed professionals
- Subject to consent and agreement, as in “smart contracts systems” allowing:
 - Self-execution of a contract (fully or partially) on the European blockchain ESBI, illustrating how to link pre-determined conditions and to execute contractual dimensions.
 - Data-tracking in the blockchain
- Difficulties/barriers
 - Professionals’ authorization to run their services on platform's data
 - Data use regulation
 - Decentralized data control: large part of data is controlled by third parties independent of the MoE, especially EdTech software publishers.
 - Lack of standardization (data formats differ from one publisher to another)
 - Lack of interoperability: “silo effect” between solutions, preventing cross-sector data use
- Partners/Ecosystem/Synergies:
 - Research laboratories and universities, EdTechs, local authorities, state administrations or operators (CNED, CANOPÉ, ONISEP, FEI)
- Main technology / Gaia-X:
 - Openness and transparency through the identification of established regulatory and data protection criteria and the guarantee of data control over use.
 - Use of existing reference architectural principles, particularly those promoted by the International Data Spaces Association (IDSA).
 - Interoperability at three infra-structural levels: interconnection, technical and semantic, for network, data and services.
 - Facilitate intra- or inter-domain specific data exchange or data-services link beyond vendors and customers’ possibilities
 - Security and data sovereignty management within a framework based on Gaia-X principles certifying providers, nodes and services, particularly with regard to technical and organizational criteria. The certification will be proven by an independent and trustworthy third party, and based on already established audit and certification procedures (e.g., Minimum standards for external cloud services by BSI, C5, ISO 27001 et Trusted Cloud).
- Expected benefits:

- Datasets access to promote public and private research.
- Contribution to the establishment of a “healthy” and sovereign data economy, based on common agreement and Data Space subscription
- Digital services improvement, e.g. algorithm training to develop innovative AI solutions
- Better understanding of what is needed, to guide educational policies.
- Capacity building on the proper use of data
- Producing new data on data use, for research purposes

Energy

Mission and Goals of the data space

The goals of the energy data space are:

- To support and accelerate the energy transition in Europe;
- To develop businesses at European and worldwide scales;
- To provide new services to European citizens, taking advantage of Gaia-X infrastructure, data protection services and associated value allowed by these services.

The Gaia-X contribution to the energy Domain

Energy European stakeholders expect strong business benefits and joint tangible outcomes from the creation of the Gaia-X ecosystem. We actively support the emergence of federated services and data platforms that create value and opportunities for all businesses – and ultimately for European citizens.

Through the Gaia-X initiative, we – at last – foresee a tangible and close way forward to easily build, assemble and use trusted and value-creating cloud services, as well as create new products/services and foster new business models that are compliant “by design” with European regulations and values. For that reason, it is not only about the data but also the opportunity to build a strong European digital system layer on top of the existing energy system layer which should be seamlessly coupled.

Added value to the whole energy value chain

Thanks to data-driven solutions, the energy data space aims to help manage the transition towards decarbonized energy and carbon neutrality. Other objectives are to enable energy efficiency and sector coupling (green energy fluids, integration of mobility and building/heating systems...), and also enable more flexibility and renewable energy integration into the European electric system. Within Gaia-X, energy stakeholders will foster European energy system optimization and competitiveness. The European ecosystem is being identified and organized to design relevant use cases on national and cross-border scales: the governance within this ecosystem needs to be clear to ease and accelerate the energy sector digitalization and its adoption. Thanks to data-sharing possibilities between stakeholders, valuable use cases will be addressed so that new services for European citizens and companies can be deployed.

Challenges addressed

Addressing these use cases requires cross-border common data knowledge representations, semantic data models, data collection and sharing data capabilities. Accessing data in trusted and collaborative cloud infrastructures is compulsory to provide comprehensive offerings under high security standards. The Gaia-X strategy of enabling access to aggregated, federated and interoperable trusted cloud services and AI services through the setting of common policy rules will ensure data, data knowledge representation and services protection, but also interoperability and portability. These are crucial to reinforce trust and transparency in order

to scale, digitalize and provide new services at a European level. Moreover, there is a need to have real-time data exchange in order to improve the system flexibility and leverage the electrification of the system.

Since highly secured data will be processed in the energy data space, dataspace stakeholders will be very sensitive to providers with highly secured standards: labelled Gaia-X.

The challenge is also on governance and organization. Big energy companies – some of them being great competitors – are joining their efforts to provide to their national and cross-border customers new services that will help business growth, social improvement and carbon neutrality reach.

Large amounts of heterogeneous data can be gathered to address the most valuable use cases, therefore, the whole value chain must participate to this joint effort:

- **Energy providers:** for the production and storage of energy for all the different kinds of decarbonized energies, energy efficiency and new services;
- **Contractors, partners and engineering service providers:** all companies and partners to improve energy production and engineering;
- **Gas, power and heat network operators (DSO/TSO/Stadtwerke):** with energy transportation and distribution data;
- **Aggregators:** collecting and packaging data to provide it to consumers;
- **Energy consumers,** as households, local collectivities, industrials (smart metering data...);
- **EV station managers:** with vehicle and charging stations data;
- **Safety and certification bodies:** with safety and certification data;
- **Open services:** with public maps, meteorological, city or transport data.

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

Partners of the ecosystem

There is already a broad representation of Energy Companies in the data space, from all around Europe, especially from Germany, Belgium and France. These companies represent all segments of the energy value chain described above.

- The energy companies that have firmly agreed to work on the data space and have written the document are the following: EDF, Enedis, Engie, E.ON SE, Westenergie AG, Arge Netz GmbH & Co. KG. It is clearly noted that the data space is open to more participants.

The energy companies that are willing to participate but have not yet confirmed are the following: Total, Air Liquide, Elia, Enel

They are already able to develop Smart Services (AI, IOT, Machin Learning, Blockchain...) and to integrate suppliers to develop these services.

- Academic partners already involved: Mines Telecom Institutes
- The technological partners that are willing to cooperate and have confirmed their participation are the following: Capgemini, Atos, Sopra Steria, Advance GmbH, Siemens energy AG, Offis e.V., Fraunhofer e. V, Examesh GmbH

The technological partners that are willing to participate but have not yet confirmed are the following: IBM, HPE

Use-cases (scenarios) within the data space and the current state of development

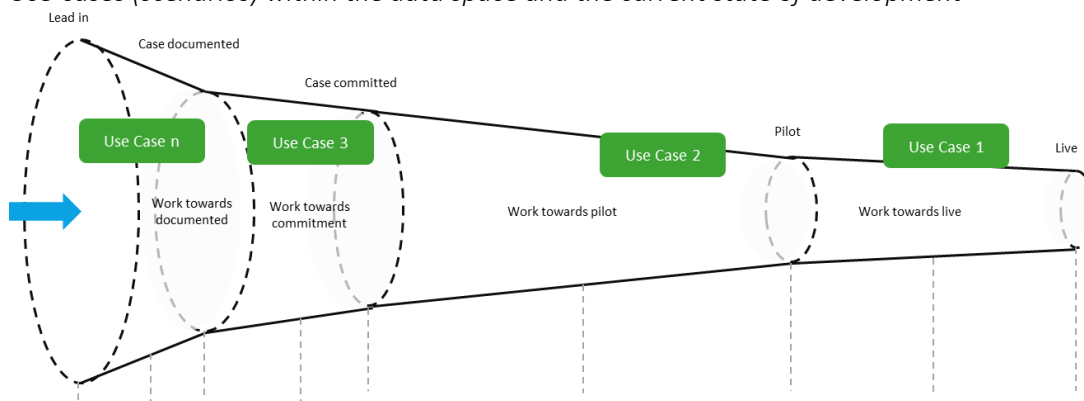


Figure 5

Description Use Cases Renewables

Renewables – Wind and solar asset description model

Solution

The wind and solar description model are the digital backbone to federate all the businesses and its ecosystem around one single source of truth from “DESIGN, MODIFICATION to others REFRESHMENT”. This structured, agnostic asset management description support business process orchestration in a context of new business model refocused on growth of decentralized and distributed energy and carbon neutrality "as a service".

Asset Data Management and valorization lay on a change in paradigm which needs to consider specificities of the data and the complexities of the life-cycle management:

- different technologies offered by equipment manufacturers: Vestas, Siemens, etc.,
- diversity of contracts with equipment manufacturers, data providers,
- strong commitment to third parties on the performance rate of turbines,
- large volume of data, inconsistency making it impossible to share the results and use them (interoperability),
- difficulties accessing data from some farms, ...

The theoretical modelling of assets is a new paradigm all along their lifecycle, enabling to link the two “worlds” as build and operate. Take some following examples to support this new paradigm:

- The development of the renewable assets should define precisely and realistically the layout from trusted and updated technical information. With such asset configuration management, scenario can be defined, simulated by modifying, combining design parameters, analysing components and their assembly to understand and assess the impact of any exogenous effect implied by layout constraints and changes to any layout.
- Thus, it is easier to consider the as-built model and update the final implementation file (Dossiers Ouvrages Executés, DOE), including underground networks and turbines.
- This asset modeling is key to compare the asset production between sites.
- This framework enables to understand how farms are operated and the root cause of observed discrepancies.
- This framework enables to calibrate the operational team on site to better regulate and organize technicians ‘activities according to travel times.

This framework is a new asset centric form of shared master data. Each equipment is linked to a model (technical description of the assets included components, physical model, localisation... as metadata) to ensure coherence and value of data with a “single point of truth” access.

It lays foundations for a Virtual Plant, ensuring continuity of data all along the lifecycle and sharing in ecosystems to drive both traditional and disruptive new business models.

Problem solved

When it comes to manage efficiently and effectively the renewable assets and all the O&M actions for the different actors (Manufacturers, Owners, Developers, Operators, ...), the difficulties to access to data exploitable in a cross functional way by these different profiles of actors and the different digital systems mobilized represents a real challenge.

Being able to share this same abstract representation of data for the Wind and Solar domain would allow a better understanding of the associated operations (asset management, RCA, Structural Analysis, Visual Inspection, monitoring ...) and an obvious improvement of the processes that mobilize the processing of this information.

The availability of standards (IEC 61400-25, etc.) provides a solid reference framework, but these are not transposed into semantic paradigms (ontology based), even if some initiatives exist in this transposition. (1)

Partners/Ecosystem

- Engie
- EDF

Main technology/Gaia-X components

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

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification

Concrete benefits

The asset description model is an opportunity to build comprehensive models, analytics frameworks and improve multiparty collaboration capabilities needed to support digital ecosystems. It is a backbone for renewables operator to ensure continuity of technical data all along the life cycle.

This “single point of truth” is a real accelerator for Greenfields and brownfields assets to deliver more safely, more quickly more efficiently (right the first time) and with a lower Total Cost of Ownership. It lays foundations of virtual plants with the following objectives:

- Projects designed right the 1st time, based on reliable & available data
- Projects delivered for operations, in compliance with specifications
- Enhanced preparation of interventions, using as operated configuration
- Break organization and information systems silos. Each business receives accurate information across the world, which makes it easier to achieve the ambition. Ex: issuing an alert when a design change made in one area may have implications in another.

Renewables – Works risk prevention

Within the domain of safety management, it is common experience that the aim to provide a safe working environment for everyone who enters it, becomes more complex when different actors, belonging to different companies have to operate in the same industrial sites. The key aspects remain that is: sharing risk assessments that are up to date, identifying areas with potential hazards, anticipating what could happened and what to do, and making sure that the information is known and taken into consideration. With a large number of industrial working places/situations for Solar, Wind or Hydro generation, with different hazards due to height, lifting, electricity, fire, pressure, ... safety management is achieved by numerous dedicated works prevention plans that need to be setup after contracting the work. Before starting the works, a joint inspection must be organized that requires different data on the site and the

work, the industrial team and the contractor's team. The efficiency of the plan that is produced depend on the experience on each side but also on the reliability with the exchanged data.

The difficulty to produce paper risk prevention plans for works in industrial environments, leads to the perception of bureaucracy instead of responsibility. New horizons open up by digitalizing the global process with a positive aim to improve the quality, to construct a stronger capacity of hazard identification and control.

Problem solved

Industrial works generate different situations with a high level of risk. In France, an executive Decree (N° 92-158) on specific health and security rules applying to the building, public works and water works sectors describes the required documents and steps that should be produced by the supplier. The key component is the risk prevention plan (RPP) which includes a mutual assessment of the risks.

Similar requirements exist among industries across EU, establishing, updating and reporting on works risk prevention requires a lot of coordination between the two parties and sometimes with the state. The information exchanged carries personal and commercial data.

Digitalizing the process in an extended Enterprise approach requires trustworthy data storage easily accessible to all kinds of companies -especially for different suppliers working on different industrial sites within the EU.

Different software packages are marketed, usually based on one industrial attempt to federate its suppliers. To go one step forward and gain time and efficiency and simplify the process, software solutions for works Risk prevention need a framework in order to develop new processes across the different stakeholders.

Solution

The Energy data space could accelerate the digitalization of the life cycle of risk prevention plans needed to fulfil EU/state requirements in terms of risk evaluation, risk reduction, works authorization.

The contractor and the industry would share data between them and sometimes with the State through digital canals accessible on site and from their respective offices, each party, having access to a specific part of the information.

The Data Space will guarantee that personal/commercial/sensitive data can be processed and shared through different means of communication. It will offer facilities of integration with industrial IT systems as well as easy user interface for small companies.

Partners/Ecosystem

The main partners are

- different industrial actors such as EDF or ENGIE for French Power generation and their equivalents across EU,
- all kind of suppliers that produce different works firstly on industrial sites but non-industrial activity qualifies in a lighter way.
- companies that offer consultancy and/or software packages of risk prevention management
- State representatives of Health and Safety at work
- Companies offering data storage with ISO 27001 and HDS (Health Data Storage) requirements reading to follow Gaia-X framework

The Use case Team would be composed mainly of representatives of industrial Power generation actors and would choose an existing software/Cloud provider involved on Works risk prevention planning in the energy sector.

Main technology/Gaia-X components

At this early stage, different possibilities need to be envisaged for the use case:

- Large companies both on the industrial & suppliers will prefer to interface their own IT solution in order to exchange commercial & personal data through the Energy Data Space or to access dedicated compliant software accessible through a compliant cloud provider
- Smaller companies could login directly on such a software package on a SaaS basis

Hence, Gaia-X components will benefit risk management in general and Works Risk Prevention Planning in particular, through the 4 Federation Services

- Identity & Trust: at an early stage Federated identity Management
- Federated Catalogue: at an early stage Self Description
- Sovereign Data Exchange: at an early stage Security Concepts
- Compliance: at a minimum Relation Between Service Providers and Consumers

Concrete benefits

The main benefits are linked with the reuse of information describing work sites, risks, supplier's data saving time (finding easily existing information) and producing better quality data hence better risk assessment due to the increase of transparency obtained through the gain of control over data sharing. To prove the value of a digitalizing the process, an experiment has been carried out on several HYDRO generation units using Software as a Service with storing data on a small cloud provider. The analysis shows that with 15 000 prevention plans per year on Hydro generation works, EDF has the forecast of a large saving per year (over

500k€) on Hydro Maintenance. The Data Space is a compulsory service for the success of such a project.

The benefits at a large scale would provide gain for suppliers as well as industrial actors. In result, less bureaucracy and more responsibility that will accompany a growing approach over risk management.

Questions/Answers identified at this stage:

- How can actual software/cloud providers involved in Risk management become a validated Gaia X provider? A proof of Value could be carried out at a small scale in order to illustrate how complicated it would be for a given solution to become compliant. An existing Cloud provider (already ISO 27001 and qualified for Health Data Storage) could join Gaia-X AISBL and an existing software provider would implement the needed federation services
- How can I impose the use of the Data Space Energy to my suppliers? Gradually suppliers will identify the benefits for themselves and will choose to comply. Before that stage it will be possible to take into account the added value, like respecting the semantic catalogue, when contracting the works. Data Space Energy requirements would be added to the purchase process.
- How can I buy such a service – do I need to have many solutions in order to establish a competitive bidding? In order to setup a solution to carry out the Proof of Value, an innovation contract will be used through a consortium of industrials ready to make use of the solution on a number of works within a definite period. This will enable suppliers, large or small, to use the solution with no extra cost.

Renewables – Common taxonomy definition – IEC standards

Solution

The objective is to propose an approach and associated services that would allow to do a job or even to provide automatic alignment systems between different information models. For this purpose, we seek to align existing models across sectors and to find transversal vectors of common information model (e.g. CIM, ...) combined with the use of ontology building approaches. Considering the other Gaia-X CUs and focusing on transposing these CUs for the renewable sector in a CTD (Common Taxonomy Definition) would allow to link and valorize the standards around the standard description of use cases, Role Model, Canonical Data Model(s) or even architectures for the energy value chain such as SGAM (Smart Grid Architecture Model).

Problem solved

In the energy value chain, which includes a multitude of actors and processes making it very complex, there are already reference and standardization frameworks that propose common

representations to facilitate the understanding, exchange and operation of associated systems and subsystems. The matrix model associating the major energy verticals (Generation, Transmission, Distribution, DER, Customer) and sectors (Process, Field, Station, Operation, Enterprise, Market) has an IEC core model for the energy and electricity domain, which enables the management of associated information models. CIM Data Model, COSEM Data Model, IEC 61850 Data Model, CGMES, etc... but also models / "Standard "cross sector as SAREF, CIM+, NGIS-LD, FIWARE, ...

Nevertheless, although based on modeling standards such as UML, the semantic links between the different uses of these different standards, information models and ontologies are not a reality. This deeply limits a higher level of transverse interoperability.

In addition, the energy sector can be broken down into several sectors composed of several fluids and the actors and infrastructures that transport them. We are referring here to three main models: electricity, gas and heat (cf. IEC 63200).

We can then understand the extreme difficulty for systems and subsystems (increasingly digitalized) to navigate and mobilize these different representations and the complexity of aligning them.

Partners/Ecosystem

- Engie
- EDF
- Other Utilities (TBC)
- ...

Main technology/Gaia-X components:

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

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification

Concrete benefits

Although the task seems extremely ambitious, namely, to align proven standards (IEC) and Canonical Data Model (CIM...) with ontologies (SAREF, OneM2M, ...) or Linked data formats (NGIS-LD), the opportunity provided by Gaia-X to align actors, CUs and data sets to define advanced interoperability models, opens up real prospects.

It seems imperative to consider the use of ontologies to achieve these cross-sector connections. The benefits are extremely numerous in the long run, because such approaches

would allow to federate different knowledge spaces and representations in the energy domain without reinventing the wheel and consolidate the years of formalization.

Use Cases Nuclear

General context

Essential, for the fight against global warming, the nuclear industry supplies energy in a sustainable and scalable manner:

1. A low carbon energy: France is today one of the countries emitting the least greenhouse gases and can rely on its electricity system to reduce CO2 emissions in other sectors (e.g., transport, industry, construction).
2. A competitive energy which benefits all economic agents: individuals, companies of all sizes...

The French nuclear industry, which controls the entire nuclear energy production chain, from uranium extraction to spent fuel reprocessing, is a benchmark worldwide thanks to its technologies, skills, and employee know-how. It represents more than 2,000 reactor-years of experience.

With its 220,000 employees and more than 3,000 companies, 85% of which are VSEs and SMEs, it also contributes, through its establishments throughout the European territory, to the development of local economies.

Gifen, a key player federating & transforming nuclear industry

Due to the risks of industrial espionage and takeover bids, as well as the need to perpetuate European industrial factories, the future of the sector depends on its ability to ensure its industrial and technological sovereignty. It is critical for French and European energy sovereignty as well as for export capabilities of nuclear industry companies.

On the strength of these convictions, the companies in the French sector are united within GIFEN - by a common objective: to build together the French nuclear industry of today and tomorrow.

Gifen brings together nuclear facilities operators (EDF, ORANO, CEA, FRAMATOME, ANDRA), large companies in construction and engineering (ENGIE, VINCI, BOUYGUES), mid-size companies in construction and maintenance service, SMEs, VSEs, IS/IT providers, software vendors, electronic manufacturers, local and professional trade organizations, and associations ; covering all types of industrial activities (studies, manufacturing, construction, maintenance, etc.) as well as all areas of nuclear power generation (fuel cycle, research, power generation, equipment manufacturing, decommissioning, etc.). At the beginning of 2021, it represents more than 230 companies.

Thus, Gifen addresses all cross-functional stakes in the service of industrial excellence and offers services of common interest. In particular:

- Map the skills of the sector to anticipate future needs,
- Consolidate workload & purchasing forecast at short, mid, and long terms. according to 18 pre- identified business families,
- Analyse the supply chain workload capabilities in response to this forecast in terms of skills and industrial tools and implement actions to guarantee it.
- Put mid-sized companies / SMEs in touch with the major clients, who manage the major R&D programmes in the sector,
- Develop collaborative platforms aimed at facilitating exchanges through the supply chain and increasing the studies and manufacturing quality.
- Support companies towards better safety culture and nuclear quality (e.g.: deployment of ISO 19443 norm),
- Structure French nuclear industry strategy for its international development regardless of the technology.

Gifen works in close collaboration with French & European public authorities.

Identified opportunities within Gaia-X

Regarding presented challenges concerning the nuclear sector, Gaia-X is the opportunity to define and implement technical, functional, organizational and governance solutions allowing the mastery of shared data as well as the necessary support to further deploy the digital transition of the nuclear industry.

The mapping of the sector capabilities using a data-centric approach and the usage of collaborative platforms around key extended enterprise stakes, illustrates the fundamental pillars of this transformation.

Considering this approach, 5 use cases are proposed:

1. Day-to-day collaboration capabilities within the Gifen
2. Nuclear industry observatory: capabilities Mapping & related data analytic services.
3. Usage extension of the *ESPN Digital collaborative* platform for the whole sector
4. Standardization & digitalization of Maintenance Work Packages (eWP / eDRT in French) through a collaborative platform, with opportunities to cover operations in non-nuclear assets.
5. Optimization of nuclear waste management

The use-cases put forward are not an exhaustive list of "nuclear use cases" within the framework of Gaia-X. It is rather a selection of the more advanced and significant cases, calling on different bricks, to experiment and gain maturity:

- on the efficiency of the services being defined by IT providers,
- on the contribution of a common Data Space for Energy players.

Moreover, these use cases have the advantage of being under investigation from a business and IT perspective within Gifen working groups.

Nuclear – Day-to-day collaboration capabilities within Gifen

Solution

To contribute to the operating performance and development of the nuclear industry, it is essential to have collaboration services between actors under the Gifen governance. This means producing, storing, and exchanging information of various formats and criticalities in a secured manner by guaranteeing access control, traceability of exchanges as well as the correct application of the usage rules... It is a prerequisite for the Gifen, in order to succeed in the missions defined by the industrial companies of the sector.

For instance, exchange concerns industrial and technical project data, export market data, industrial feedback, studies, applied industrial and technical standards, etc.

Problem solved

Regarding the Gifen missions, the protection of data and their sharing is a critical issue. This protection covers:

- Know-how or patents (intellectual property)
- Financial interests
- Process continuity and integrity
- People and companies' ecosystem

In addition to these items, the main requirements are:

- Authorization and accreditation management,
- Access traceability
- Storage whatever the format of the information
- Functional modules dedicated to small entities (file sharing, chat, conference call, validation workflows, etc.)

- Interoperability between the information systems of main nuclear players and their partners.

Partners/Ecosystem

All players in the nuclear industry are concerned, regardless of their size, throughout the whole civil nuclear lifecycle (design, manufacture of equipment, construction, operation, maintenance, dismantling and management of the fuel cycle, etc.). It represents around 3000 companies with multiple sites in Europe. The Gifen, enabling French industry transversal governance, has more than 230 French members in January 2021, compared to around 400 companies which represent 80% of the nuclear industry top line. Thus, a pilot should be launched within Gifen governance.

Main technology/Gaia-X components:

The solution supporting the use case would require the following of the Gaia-X infrastructural components envisioned:

- Identity & Trust: federated identity management for individuals and organizations,
- Federated Catalogue: to publish the registration, consent, and query services,
- Sovereign Data Exchange: to manage consent and usage control,
- Compliance: rights management, onboarding, and certification

It would also require these specific items:

- Capability to manage very diverse nature of data (Office documents, video, process and data modeling, document in the form of XML, JSON, etc.),
- Capability to manage visas / signatures across nuclear partners,
- Capability to manage high level of confidentiality with restricted accesses.

Concrete benefits

A common language and associated collaboration solutions represent the basis for better collaboration, creating trust between stakeholders. That makes possible fostering innovation and the development of common projects. The benefits are difficult to quantify at this stage, in particular the indirect benefits linked to common projects that may be put in place. Most of operational benefits are around avoided costs: rationalization of collaborative spaces, consequences of a data leak and data inconsistencies.

Nuclear – industry observatory: capabilities mapping & related data analytic services

Solution

The observatory would include information from the various players of nuclear industry (operator, manufacturer, designer...) allowing consolidation of a detailed mapping aligned with Gifen missions. As instances, it would be valuable to gather in particular:

- Legal form of the actors,
- Workforce in terms of volume, skills, and distribution on the European territory,
- workload forecast at short, mid, and long terms based upon main industrial project assumptions,
- Breakdown over the 18 business families defined at Gifen,
- Information on international critical markets for the nuclear industry,
- Innovation programmes data,

Mapping would be the basis of data analytics-like services enabling to understand, predict and make decisions on the industrial system in response to main nuclear industry stakes.

Notably, the adequacy analysis of the supply chain workload capabilities regarding the operators and main project forecast in terms of skills and industrial tools represents a top priority. It will significantly ease the identification of under pressure job and competencies and ease the definition of the related action plan. It means implementing an interoperable system for collecting, analysing, and sharing this information in a distributed or centralized and secured manner.

Problem solved

The set-up of a "nuclear industry" data space is a prerequisite for several fundamental missions of Gifen:

- Map the skills of the sector to anticipate future needs,
- Consolidate workload & purchasing forecast at short, mid, and long terms. according to 18 pre- identified business families,
- Analyse the supply chain workload capabilities in response to this forecast in terms of skills and industrial tools and implement actions to guarantee it.
- Structure French nuclear industry strategy for its international development, whatever the technology

Gifen has already initiated these actions with its own resources to demonstrate their value. A roadmap for industrialization is under definition. Without a trust & governed workplace nor rules for collaboration and sharing, it remains very difficult to implement this strategy.

In addition, interoperability with the upstream and downstream areas of the energy value chain, as well as related sectors, that are highly capital-intensive and require common skills such as infrastructure construction, is essential at mid-term. Indeed, transversal optimization is part of European Sovereignty, and thus, deserves to be tooled.

Partners/Ecosystem

All players in the nuclear industry are concerned, regardless of their size, throughout the whole civil nuclear lifecycle (design, manufacture of equipment, construction, operation, maintenance, dismantling and management of the fuel cycle, etc.). It represents around 3000 companies with multiple sites in Europe. The Gifem, enabling French industry transversal governance, has more than 230 French members in January 2021, compared to around 400 companies which represent 80% of the nuclear industry top line. Thus, a pilot should be launched within Gifem governance.

Main technology/Gaia-X components:

The solution supporting the use case would require the following of the Gaia-X infrastructural components envisioned:

- Identity & Trust: federated identity management for individuals and organizations,
- Federated Catalogue: to publish the registration, consent, and query services,
- Sovereign Data Exchange: to manage consent and usage control,
- Compliance: rights management, onboarding, and certification.

Concrete benefits

The benefits are difficult to quantify at this stage. They will be detailed for each added value services made possible by this observatory, notably orientations on the nuclear industrial system (manufacturing facilities, design center...) taken following the analysis of the adequacy of charges / resources in the short, medium, and long term.

This will allow to strengthen the industrial policy of the nuclear industry in the medium and long term, in particular by:

- a better identification of project risks,
- an eased consolidation of export opportunities.

Nuclear – ESPN Digital Platform for the nuclear sector

Solution

Background and ambition

The "ESPN Digital" project was launched as part of the EDF SWITCH programme. It is the digital extension of the industrial control strategy of the Nuclear Pressurised Equipment) regulation.

The aim of these regulations is to guarantee a level of control of pressure equipment adapted to the nuclear safety issues at stake.

ESPN Digital is a digital platform built to facilitate the assessment of regulatory compliance of ESPNs. It complies with regulations while making exchanges more fluid between operators, manufacturers, certification agencies and the ASN⁴, enabling them to refocus on their core businesses, collaborate better and become more competitive.

The first version was developed and put into service on a limited number of projects. The ambition is to extend its usage to all the industry players, for the construction and operation of new ESPN equipment intended for nuclear facilities, but also for modifications or repairs of installed equipment.

Service description

The platform allows pooling and tracing all the information provided by stakeholders on ESPNs: history, provided justifications, data, workflow, etc.

Eventually, ESPN Digital will make available guidance to be followed, and enable everyone to draw up with confidence all the documentation required by the regulation and progressively automate its edition, thus facilitating the work of the ASN or the authorized agency (which will eventually lead to a reduction of the time required to obtain the ESPNs certificate of conformity).

Problem solved

The GIFEN wishes to respond to the needs and difficulties raised by the actors of the sector for the application of the regulation through this project. The main difficulties put forward are the following:

- Very long and unpredictable conformity assessment times
- Heterogeneous practices of manufacturers and organizations
- A limited number of actors mastering the ESPN process

Thus, the ESPN Digital platform extended to all the industry should help to meet these challenges by centralizing and making exchanges and collaboration more fluid between all conformity assessment stakeholders. The platform also has the virtue of harmonizing work methods and digitizing all processes to comply with regulations. Finally, the digital tool will offer a consolidated and 360° vision of ESPN Safety Cases and Certification processes to all stakeholders.

Partners/Ecosystem

The actors in the sector involved in the assessment of regulatory compliance of ESPNs are organized into 4 categories:

- The national regulatory authority for ESPNs (ASN)

- The authorized agencies (OH for Organisme Habilité in French): notified and approved organizations and user inspection authorities (Apave, Bureau Véritas, OIU, Vinçotte, GSEN)
- Manufacturers of ESPNs or assemblies containing one or more ESPNs and subcontractors (Framatome, Onet, Sigedi, Endel, Velan, Westinghouse, Mitsubishi, Afcen...)
- Operators (CEA, EDF, FRAMATOME, ITER, NAVAL GROUP, ORANO, TECHNICATOM ...)

Main technology/Gaia-X components:

Digital tool: 3DExperience Enovia digital platform

Concrete benefits

ESPN Digital stakeholders expect many gains through the implementation of this project:

- Direct benefits:
 - 20% of the cost of compliance certifications
 - Estimated savings of 15 M€/year in the current configuration of the EDF fleet + FA3
 - ~80 M€ for a new project (FA3 type) over 8 years
- Major indirect benefits:
 - Better control of project planning and potential impacts on the availability of nuclear facilities
 - Lowering entry barriers linked to ESPN regulations

Nuclear – eWork Platform

Solution

As part of the digitization of the sector, one of the goals of this intermediation platform is to standardize the Work Order (eWork) and its set of associated documents constituting the electronic Work Package. This use case will facilitate B2B exchanges between the various companies performing interventions on nuclear sites.

To facilitate the adoption of this digital Work Order (eWork) by all the companies, the platform must consider the following requirements:

- Guarantee a multitenancy architecture where access to documents and data of each member of the industry is secured.
- Allow access to eWork information through mechanisms adaptable to the digital maturity of each partners

- Implement mechanisms to verify the compliance of eWork with nuclear industry standards
- Provide configurable workflow functionalities,
- Offer visas / signature mechanisms for paper and digital documents such as digital procedures or QA documents.

The nuclear industry partners will have to define and agree on the standards for these digital data exchanges.

Problem solved

The digitalization of the electronic Work Package allows the numerous companies intervening on nuclear site, to have an interoperable and exchangeable format with the eWork. It can be exchanged and enriched by B2B exchanges between the players. These exchanges will be orchestrated to guarantee the best practices in the sector.

This digitalization should enable the sector to have at its disposal several software solutions compatible with the GIFEN recommendations.

For the many companies working on nuclear site, it will be possible to use, when appropriate, their own Work Force Management (WFM) tool well adapted to the company's process, by simply adapting their interfaces. These interfaces will be well defined thanks to this use case.

Partners/Ecosystem

Stakeholders are all the players in the nuclear industry, whatever their size (Operators, Construction and Engineering companies, SME, VSE). They represent the different trades of the civil nuclear, design, equipment manufacturing, construction, operation, maintenance, decommissioning and fuel cycle management, ...

This represents about 3000 companies in France. The Gifen, which would be responsible for the governance of the sector, will have more than 230 members in January 2021, in line with the 400 or so companies that represent 80% of the activity.

Main technology/Gaia-X components:

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

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent, and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification and it would require these specific items:

- a shared third-party tablet signature (or strong authentication) mechanism
- a mechanism to endorse/sign documents and XML data blocks
- an application store to deploy services and applications to partner companies, especially SMEs

Concrete benefits

The digitalization of DRT allows the ecosystem to gain in efficiency by sharing the WO schedules and by digitizing all the data relative to this eWork.

This digitization will allow efficient B2B exchanges between several hundred companies that interact on field interventions in nuclear site.

Digital continuity will reduce the breaks in the chain between the client and the contractor:

- starting field activities quicker by transmitting information via telecommunication networks
- completing field interventions without returning to the office, facilitating the optimization of the planning of all interventions
- suppressing the costs of scanning and printing files; gains on this side contribute to a very short ROI.
- offering the possibility of optimizing maintenance operations by processing data analytics on a greater quantity of data captured.

Use Cases low-carbon hydrogen

H2 – Import/export international routes setting up

Solution

A dedicated low-carbon hydrogen import / export market will connect producers with end-users, similar to today's current gas market. There will be a need for local marketplaces, European marketplaces and international marketplace. These marketplaces will integrate various forms of hydrogen (gaseous, liquefied) and will need to feature the certification of low-carbon hydrogen (see specific use case).

To support the marketplace, a transparent data platform will be required to:

- Map hydrogen production based on location and relevant stakeholder
- Monitor supply and demand
- Monitor prices based on the different products (spot vs. long-term)
- Monitor management rules in case of unused capacity vs. congestion
- Enable market settlement after transactions

Problem solved

Low-carbon hydrogen production could take place next to consumption centers. However, in most cases hydrogen will be produced where it will be the most cost-effective – because renewable electricity is the cheaper, because gas infrastructure can be re-used. We are currently missing the global overview of hydrogen supply and matching demand

Partners/Ecosystem

The marketplace will involve a large panel of partners: producers of low-carbon hydrogen, transportation and distribution operators (including pipeline, vessels, storage), trading houses and end-users (industrials, mobility providers, energy providers).

Main technology/Gaia-X components:

The solution supporting the use case would require the following of the Gaia-X infrastructural components envisioned:

- Identity & Trust: federated identity management for individuals and organizations,
- Federated Catalogue: to publish the registration, consent, and query services,
- Sovereign Data Exchange: to manage consent and usage control,
- Compliance: rights management, onboarding, and certification.

Concrete benefits

According to Fraunhofer ISE¹, a market for hydrogen import / export could be worth between 100 and 700

Billion € per year in Europe.

H2 – Station networks information sharing

Solution

The emergence of these H2 mobility technologies will be linked with the ability of the sector to develop a wide range of H2 stations allowing the user to move without constraints.

We are missing a global view of existing H2 stations, with their characteristics and the global matching H2 demand for mobility, to ensure a quick development of H2 mobility.

To support the development of H2 mobility, a transparent data platform will be required to:

- Map H2 stations based on location and characteristics
- Monitor H2 stations capacities and H2 demand

Problem solved

Mobility is today largely contributing to CO₂ emissions in Europe, and solutions rely on the development of low carbon mobility such as EV and H₂ mobility. EV is developing more and more while H₂ mobility is still in the early phases. The development of H₂ mobility is widely based on the capacity to propose a network of H₂ stations that fits the needs of the consumer.

H₂ mobility represents today a very small part of the mobility market, led by traditional carbonated technologies, and emerging EV.

Partners/Ecosystem

Sharing information on H₂ stations will involve a large panel of partners:

- producers of low-carbon hydrogen
- local authorities and operators H₂ stations

Main technology/Gaia-X components:

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

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification

Concrete benefits

The benefit for such a sharing of information is difficult to assess but it will allow a fast development of H₂ stations across Europe, optimized for the use of H₂ mobility users.

H₂ – Mobility asset monitoring

Solution

The focus of this use case is to build common knowledge representation, collect and share data on operations of Hydrogen refueling stations and hydrogen vehicle fleet. This could include data on main design characteristics, energy consumption of HRS and individual major component, energy efficiency of HRS and individual major components, hydrogen production rate, refueling time, vehicle efficiency, etc. A good basis of work is the listed data in “JIVE and MEHRLIN Performance Assessment” framework and this could be extended to higher collection frequency and a wider range of HRS and vehicle.

Gaia-X enables for this use case to provide a complete functional and technical framework which for this use case will provide the means to make available throughout the value chain and between multi-actors the data and representations shared to build portable and secure services that can address a federated monitoring by the actors concerned while respecting the rights and access to information. Moreover, the availability of these data and representations to allow the creation of new services by third parties who will commit to use these open and shared repositories.

Problem solved

The provision of data depends on the local data management, which is used by the operator on site, although there is no integrated common infrastructure.

It will allow to develop services around maintenance, individual component improvement and development, HRS architecture improvement.

Partners/Ecosystem

The ecosystem is composed of all actors involved in hydrogen mobility:

- DSO
- Hydrogen refueling station (HRS) operator
- Hydrogen vehicle fleet operator
- HRS component manufacturer
- OEM
- Gaia-X services platform

Main technology/Gaia-X components:

This demonstrator would require the following of the Gaia-X infrastructural components envisioned

- Identity & Trust: Enabling the identification of assets and trustworthy interactions between human and non-human actors (e.g., federated identity management, Trust management)
- Federated Catalogue: a description of the existing data and services in a self-description approach (e.g., Self-description)
- Sovereign Data Exchange: Use of technical blocks (data containers) to exchange in a secure way data and representations between different participants and different roles (e.g., Usage Control for data protection, Security Concepts)
- Compliance: use of the Relation Between Service Providers and Consumers components between along the value chain of H2 mobility

Concrete benefits

This demonstrator focuses on the brought usability of vehicles and refueling infrastructure data. We see a huge retention to provide infrastructure data for further purposes due to the risk of being misused. By using Gaia-X and its components of identity & trust, sovereign data exchange and compliance third parties it will allow to develop services around maintenance, individual component improvement and development, HRS architecture improvement.

Use Cases downstream, energy renovation and electric vehicles

Downstream – Energy renovation: map building potential for renovation

Solution

2021 marks a strong acceleration of the energy-efficient building renovation policy, with very ambitious targets in terms of the numbers of buildings renovated. Concrete financing mechanisms across European member states have been put in place to subsidize residential and tertiary building renovations. The platform we would like to develop helps identifying and prioritizing the building renovation programmes to be carried out; obtaining, in a context of multiple levers to master, the right financing; and manage and monitor the energy performance of renovated buildings over time.

Our platform offers automatic data collection and standardization flows from multiple databases, machine learning models and data sharing principles between public and private players in the energy renovation ecosystem, and therefore builds on a comprehensive and continuously updated database. This database would provide a set of detailed information on all the characteristics of the building and benchmarks for building energy consumption. It would make it possible to target the appropriate renovation work, illustrate the possible benefits and advise on the available fundings. This capacity was illustrated in the tRees project, developed by the namR startup, which allowed them to identify all the educational buildings in the Haut de France region in France (19,734 buildings), and to characterize them with more than 150 attributes.

Problem solved

Most players at national and local levels underline a lack of reliable data to define, finance, implement and manage their energy renovation policy and their assets strategy. The pain points identified include:

- A partial vision of their assets, difficult information to collect, limited internal human resources, costly and numerous external diagnostics and audits to implement to gather information
- Difficulty in identifying the renovation work that could be carried out and in estimating the associated financial and energy gains in order to be able to prioritize them

- Multiple funding levers to know and master, complex paperwork to undertake, including intricate analysis to be carried out
- Insufficient energy monitoring of buildings, while local authorities are to achieve ten-year energy reduction objectives and must manage new information on a regular basis
- Segmented industry with low interaction between stakeholders
- Archaic tools for planning, construction, and operation

Partners/Ecosystem

The energy renovation ecosystem is made up of numerous players at both national and local levels, with both private and public actors. The energy renovation value chain has a variety of players:

- Public: ministries, energy and urban planning agencies, local authorities, public funding institutions
- Energy utilities
- Construction companies, Investors and real estates
- Producers of open-source satellite and / or geographic data, such as the IGN for the “land register plan”
- Data specialized startups

Main technology/Gaia-X components:

This demonstrator would require the following of the Gaia-X components:

- Identity & trust management for individuals and organizations
- Federated catalogue: to publish the registration, consent and query services
- Sovereign data exchange: to manage consent and usage control for data protection
- Compliance: rights management, relation between service providers and consumers, rights and obligations of participants

Concrete benefits

Improve the energy efficiency and the energy bill of buildings, in line with the promise of a low carbon balance, to reduce the final energy consumption by at least 40% in 2030, 50% in 2040 and 60% in 2050. This will be enabled thanks to the ability to:

- Obtain / acquire quantitative and qualitative characteristics on buildings, local territory characteristics and the energy renovation sector and be able to compare them with other territories
- Target the buildings to be renovated, the types of renovation to be carried out and simulate the expected gains

- Determine the sources of funding available and simplify requests, in particular within the framework of the recovery plan
- Access a personalized tool library and a network of experts relevant to the renovations to be undertaken
- Estimate and monitor the financial and environmental impact of planned and completed renovations
- Cost-effectiveness ranking of various renovation technological packages

Downstream – Local communities of energy setting up and decentralization

Solution

A local energy manager (LCE) coordinating the local energy efficiency by managing the renewable assets and energy infrastructures, by self-consuming and injecting the extra production to the distribution grids.

The business model will be consolidated by sharing economy principles. A variety of approaches to community ownership, including joint ventures, split ownership and shared revenue could be explored in the project. Enabling clean energy ownership through community enterprise meets the twin objective of decarbonizing with cheap onshore renewables and winning the support of the local community. Paring the constraints and opportunities of energy and real estate sectors, the UC plans to develop a new design/build/own and operate (DBOO) offer for new and renovation districts co-developed with the clients and by sharing investments and profits.

Problem solved

The LCE concept overcomes building energy seasonal peaks and provides interoperable business models and digital services based on trading of various energy carriers between the communities and the gross market and ancillary support to the distribution and transport networks. The essence of the business model is to have long time ambitions and to share a part of the gains with the end users that become prosumers

Partners/Ecosystem

Various players will be required so that the local energy manager has access to both supply and demand data: local energy producers, small and medium consumers, external information based on weather, power prices providers

Main technology/Gaia-X components:

This demonstrator would require the following of the Gaia-X components:

- Identity & trust management for individuals and organizations

- Federated catalogue: to publish the registration, consent and query services
- Sovereign data exchange: to manage consent and usage control for data protection
- Compliance: rights management, relation between service providers and consumers, rights and obligations of participants

Concrete benefits

- Up to 20% self-sufficiency thanks to PV without storage
- More than 70% renewable rate thanks to renewable production, heat recovery, DHC, etc
- More than 30% seasonal peak covered by the geothermal or hydrogen storage

Downstream – Electric vehicles – Energy Roaming

Solution

A platform that allows customers to freely move and charge between their home charger and partnering

public charging networks. Allowing a view of all of their EV ‘fuel’ in one place.

This platform will make use of online maps to help locate chargers, give an availability status, and also allow access to book a chargepoint, start, stop and pay for a charge.

It will provide reporting and data. This will require the access of data from various e-Mobility Service Providers, charge point operators and vehicle OEMs.

Problem solved

The problem to solve is that EV drivers can find charging their cars confusing and worrying. There are many data sources separately held that could unlock solution and value if they were aggregated and centralised for use and analysis.

- Multiple smartphone apps/ RFID cards are needed to access the different Public charging networks
- Varied pricing rates and structure on all networks – difficult to compare competitiveness (connection charge only, connection + unit rate, unit rate only, pay for a time period, etc.)
- Difficult to plan longer journeys across multiple networks you are a member of
- Difficult to know which networks are more reliable/trusted brands for quality & service
- Users frequently find charging points not working or already in use
- EV users must plan their journeys more akin to that of a commuter using public transport. It takes time, effort and can cause worry on the road

- Limited real interoperability of networks

The situation can be exacerbated if a driver does not have a home charger or lives in an apartment block with no EV charging facilities (which is likely at the early stage of EV market development).

Partners/Ecosystem

- Charging point operators
- Mobility service providers
- Vehicle OEMs
- Utilities

Main technology/Gaia-X components:

This use case would require the following of the Gaia-X infrastructural components envisioned:

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification

Concrete benefits

Gaia-X will enable the creation of a new service around the EV's owner experience. This shall occur by producing a set of standards aimed at facilitating the coherent centralisation of data owned by the relevant partners involved (CPOs and EVs).

Downstream – Electric vehicles - New services

Solution

A platform/ecosystem where:

- An EV owner can reduce their total cost of ownership by making available the EV's storage capacity (battery) for a period of time under clear conditions (e.g., minimum guaranteed level of charge of the vehicle and departure time with a sufficient level of charge),
- An operator can aggregate EV individual storage capacities in order to offer services based on a large storage capacity (Virtual Power Plant), with possibly specified grid connection points.
- An operator can actuate/manage the charge of EV currently connected to charging points in order to respond to TSO, DSO or Electricity supplier needs (peak shedding,

ancillary services, voltage management at HV-MV power station perimeter or off-peak management etc.),

- An operator can act as an active agent of the electric market making arbitrage (buying electricity at a cheap or even negative price and selling it when price goes up) or at least optimizing the charging price (knowing the purchase contract of any individual customer and its needs).
- An operator or “smart charging services provider” can also help to optimize the sites’ own consumption by helping it to avoid taking power from the grid when costs are higher and instead taking part of the load from the vehicle batteries (Vehicle-to-Building).

To do so, a huge amount of data needs to be shared among different stakeholders:

- Data of charging EVs: state of charge, time left until next use, minimum state of charge required by driver...
- Charging Point characteristics (location, power, V2G etc.) and status (car connected or not etc.)
- Data of network: voltage and load at HV-MV power stations level (real time and forecast), grid monitoring data, level of congestion of power stations
- Data of production: power produced (nuclear, hydro, coal, PV, solar...)
- EV owner electricity purchase characteristics,
- Market conditions and forecasts

And computing capacity must be available for real time calculations of the “smart charging algorithms”.

Problem solved

This solution brings globally to the electric system the following services:

- Flexibility through load management, network equilibrium, peak shedding, and ancillary services
- Reduce total cost of ownership of EV by offering an EV owner the possibility to generate revenue from the electricity injected or off taken from the grid from the EV
- For electricity suppliers: real time management of their portfolio in order to balance consumption and sourcing of electricity. For example, if renewable energy production is high at a moment of the day when there is not enough demand, it is better to charge the EV than to sell electricity at negative prices

Partners/Ecosystem

- EV owners (individuals and fleets)
- Charging point operators
- Operators, as for example Aggregators
- DSO
- TSO
- Electricity Suppliers

We see in this area a link with the Gaia-X Mobility dataspace.

Main technology/Gaia-X components

This use case would require the following of the Gaia-X infrastructural components envisioned:

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification

Concrete benefits

This use case brings the opportunity to share and to aggregate data from the different stakeholders of the EV value chain. These technical and personal data are the fuel of the services.

By offering this possibility, Gaia-X will enable the creation of new services around smart charging, flexibility and V2G by offering a uniform, secure and real-time access to critical data points.

Gaia-X would therefore play a significant role in the mainstream development of the EV ecosystem in Europe, while contributing towards the optimization of the whole electricity process from production to consumption.

Downstream – Stadtwerke/Local Open Data for Business Models in the Energy Industry

Solution

The focus of this use case is the interdisciplinary challenge of the grid connection process for both customers as well as prosumers, which requires a large amount of both data and information and sub- processes from individual grid operators (typically DSO) as well as a large amount of data from various sources, especially from public geographic information systems (municipal data for the grid connection point and other data from residential registration offices e.g.). While it is possible for a resident to initiate the grid connection process, it is yet not feasible to systematically incentivize e.g. residents that have a heightened potential for

utilizing renewable energy technologies due to a favorable combination of their location, supply contract and other already installed technologies/appliances etc.

Gaia-X enables the central provision of needed data without media breaks and guarantees the quality assurance of the data provided. Thus, not only grid connection processes but also maintenance services as well as other integrated business cases relying on information about technology and usage pervasion can be accelerated.

Especially in the grid connection use case, data required in the grid connection process can be unified and standardized based on a common semantics and access platform with Gaia-X. In addition, the price of the required data for the customer /data receiver is set and there is a role-based access control for various kinds of service levels (e.g., DER contractors, utility, etc.).

Gaia-X offers the required common infrastructure and enables the transparent and traceable transmission of public data from public administration to the grid operators. Gaia-X promotes the digitalization of public administration processes and enables the improvement and added value of data-intensive processes as well as the development of future business models in the energy industry.

Problem solved

The provision of data depends on the local data management, which is used by the offices on site, although there is no integrated common infrastructure. There is currently a technical and organizational challenge to transfer public data in a way that they are digitized as a public administration process and are available to the grid operators even though there is a lack of uniform semantics and access platform.

Further challenges are the identification of the required data sources, the digitalization of analog data sets as well as the quality assurance and development of price models for additional services that can be offered. In addition, concepts for organizational governance and data ownership must be clarified and data maintenance must be guaranteed.

Partners/Ecosystem

Partners participating can be from greatly varying domains as this is a process which has a lot of stakeholders involved. We envision at least:

- DSO
- Contractors
- Utilities (sales)
- Municipal planning and providers (GIS data, land registry) and
- Gaia-X services platform

Main technology/Gaia-X components:

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

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification

And it would require individual municipal data standards:

- GIS standards
- Public procurement standards

Concrete benefits

Open data has great economic potential: For Germany, the economic value of open data is estimated at around 12 billion euros annually. Positive effects of open data result from the use of data in the economy, its potential for innovation, increased transparency and its potential for cost savings. In addition, open data plays an important role for future business models, especially in data-intensive processes such as those found in the energy industry.

Downstream – Infrastructure data for new business models

Solution

Energy infrastructure belongs to the categories of most critical infrastructure in our economy. Critical infrastructures are those that ensure the supply of essential goods and services. They form the nerve cords of our modern society. Due to their importance for the interaction of all segments of society, these infrastructures require special protection. This also applies to data in the energy supply sector, for example, the misuse of which could cause great social and economic damage. For this reason, the data obtained may only be used and processed in a way that is relevant to the matter in hand.

At the same time, digitization has also arrived in the energy industry. It is driving forward the process of restructuring the energy system initiated with the energy turnaround in the form of more efficient processes and new business models. Several hundred energy start-ups are already supplying the energy turnaround in Germany with innovations such as virtual power plants, i.e. physical power plants that are interconnected via a platform. A 'dedicated turnstile' now provides them with even more support. Business models that drive the energy turnaround can and should also be implemented using infrastructure data on a simple and secure way.

Problem solved

The challenge is to reconcile the objectives and principles of using data from critical infrastructures and the need to use data for new business models. To this end, regulatory issues should also be addressed: Is all data classified as critical or is there any differentiation? If so, may only a certain type of e.g. certified data centers be used for processing? Which market players would have access to the data they could use to transform value chains? How would data sovereignty finally be guaranteed?

In short, the provision and secure use of infrastructure data must be clearly regulated. One option could be to use only digital data twins, as proposed in the present Use Case. In this case, it would also be possible for energy suppliers or third parties to develop new business models based on this data. Some operators of critical information already provide such data twins which can be used by others for their own business models.

Partners/Ecosystem

In cooperation with a small group of startup companies a couple of interesting and challenging business models with energy data and data coming from other sectors will be defined. The energy data comes from different partners like grid utilities, wind farms, photovoltaic providers etc. The new business models create new solutions by aggregating and refining those data sources to a new product to be sold on the markets of energy, mobility or other emphasis.

Every partner brings in its main submission. At least we will receive a win-win-situation of parties which provide data for further applications and parties which provide new applications to the market. The platform of Gaia-X provides a perfect Ecosystem for both sides and offers all necessary tools of data security and privacy.

Main technology/Gaia-X components:

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

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification

Concrete benefits

This demonstrator focuses on the brought usability of infrastructure data without endanger the critical infrastructure itself. We see a huge retention to provide infrastructure data for further purposes due to the risk of being misused. By using Gaia-X and its components of identity & trust, sovereign data exchange and compliance third parties will become providers of distinguishing data products and software appliances.



Figure 6

Use Cases networks

Networks – Long term scenarios

Solution

Creating long-term scenarios for the energy transition requires running mathematical optimization models which are to be fed by numerous data.

By creation of long-term scenarios, we understand optimum energy mixes (installed capacity and storages, expansion of networks) from now to 2050. There exist today many different initiatives and projects within the European modelling community, but it appears that models and input data are not fully open, suffer a lack of transparency, which makes the studies' results difficult to understand and analyse. Most of the needed data are often not available, non-consistent (coming from various sources) and not transparent. The openENTRANCE project (started 2019, running until 2023) proposes a first step for more transparency and more data available. This project has the following main deliverables:

- A database offering access to modelling results and inputs
- A nomenclature of data (i.e., an accurate description of all different variables which are available on the database)
- A series of open models (made open during the project), connected to the platform (i.e., able to be fed with data based on a common data format)
- Long-term energy scenarios computed from open data, with an open-source model
- A series of case studies (inputs and results being available on the database) focused on some topics of interest.

The objective of the Gaia-X use-case would be to:

- Share and extend the data nomenclature
- Facilitate access to various sources of data (which would be made consistent with the nomenclature)

- Access to consistent data
- Run Advanced functions:
 - Selection/upload of consistent data
 - Various treatments of data
 - Allow users to run the open models. Those models require IT systems which are often not available for the teams willing to run the models. Gaia-X could offer this facility (containerize models so that they can run on any IT system + rent High Performance computing resources)
 - Benchmark functions making it possible to re-run a specific study with the same model but different inputs /assumptions or the same inputs and a different model
 - Visualisation of inputs and outputs + statistical analysis
 - ...

Problem solved

Decision makers, stakeholders and modelling teams, at European, national and local levels underline a lack of reliable, consistent, transparent data to build relevant long-term energy scenarios. Many studies are published by many different teams which the latter cannot easily interpret nor challenge, due to the lack of transparency both in input data, modelling assumptions and algorithms, and output data. Moreover, modelling team often cannot access to some of the data (including the fact that a high percentage of modelling resources is used to look for data or replace unavailable data) or only have access to low quality data in their studies, and it is very difficult to assess the impact of this lack of quality in data on the results. The approach proposed by openENTRANCE includes most of the advanced functions but relies on a database which cannot be scaled up, as it was designed for small volumes of data. Moreover, modelling teams do not have access to the adequate IT resources to be able to run big cases.

Partners/Ecosystem

- Public actors and operators: EU (DG Ener), national authorities responsible of NECPs
- Energy stakeholders: utilities, network operators
 - EDF
- Energy system modelling teams:
 - EDF, Sintef/NTNU, IIASA, TU-Berlin, (partners of openENTRANCE and many others)
- Mathematical optimisation teams:
 - EDF, University of Pisa
- IT services providers:

- HPE
- Smart services:
 - IA & Optimization methods, algorithms, and solvers
 - HPC & containerization
 - Data analytics: analyses and statistics on models results

Main technology/Gaia-X components:

This demonstrator would require the following of the Gaia-X components:

- Identity & trust management for individuals and organizations
- Federated catalogue: to publish the registration, consent, and query services
- Sovereign data exchange: to manage consent and usage control for data protection
- Compliance: rights management, relation between service providers and consumers, rights, and obligations of participants

What would be included in the catalogue:

- **Data following a data and variables nomenclature** (already implemented) from various data providers, including public sources (ENTSO-e, IPCC, Copernicus, H2020 projects on energy modelling, utilities...)
- **Advanced smart services:**
 - **Data selection, treatments** (e.g., aggregations, disaggregations, scaling both on geography and time, generation of uncertainties...)
 - AI:
 - **Optimization models** (Integrated assessment models, long-term cross sector energy equilibrium models, unit commitment models..., cf list of models connected to the openENTRANCE platform); could include both open and non- open models depending on the business plan
 - **Learning/statistic/IA algorithms** (generation of time series linked to climate change scenarios...)
- Portability/interoperability interconnectivity services:
 - **Databases:** allow hosting high volumes of data with advantages of state-of-the-art database algorithms, allowing easy access for smart services

Access to computing resources: allow users to run tools (models/AI algorithms above) that require HPC on the adequate systems (access to HPC / cloud resources out of user's own environment)

Containerization: enable running the smart services (in particular solving big-size optimization problems or run statistics/IA algorithms on big size data) on any IT system

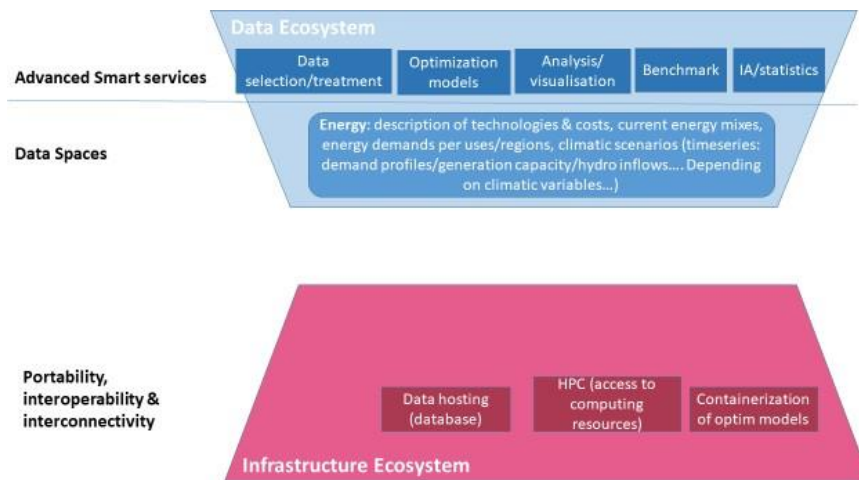


Figure 7

Concrete benefits

- Improve the European ability to build relevant and feasible long-term energy scenarios
- Enhance the reliability of scenarios
- Enable benchmarking of scenarios
- Increase actor's confidence in published scenarios
- Enable different kinds of actors to run their own energy system modelling studies:
 - Actors with access to data but not to state-of-the-art model
 - Actors with access and experience on models but lacking consistent and high-quality data
 - Actors without access to HPC resources....

Networks – EV – CPO and DSO Investment and Planning

Solution

The solution has two parts:

- A technical part with the creation of a platform which can analyse and cross data
- A business part with the sale of studies, which use the data of the platform and personalize the result for the client

Problem solved

The main goal of this use case is to increase the efficiency of the deployment of charging stations in Europe, thanks to the data convergence between the different operators. Finding the best place for a charging station requires having the following cross vision:

- The city's electrical plan to find out if an area can easily accommodate a fast-charging station without doing a lot of long and expensive network work (from DSOs)

- The city development plan to know how the city will evolve (from local collectivities)
- The flow of vehicles in the city; data coming from local collectivities, parking operators and motorway companies. A special interest could be bringing by data about where cars stop and where professional drivers are parking.

Vehicle flow could also be captured by passing through telephone operators, car manufacturers or pure players such as Waze.

Partners/Ecosystem

The ecosystem is made up of 3 major players:

- Local collectivities (cities, ...)
- DSOs
- Charging point operators

Main technology/Gaia-X components:

This use case would require the following of the Gaia-X infrastructural components envisioned:

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent, and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding, and certification Data must be extracted from GIS systems, and visualized.

Concrete benefits

A consolidated knowledge of the data of all the players would allow:

- CPOS to install the most profitable charging stations
- Local authorities to deploy charging stations where it is needed and at with an efficient price
- DSOs to plan in the medium and long term the evolution of their network

Networks - OrtoPhotos

Solution

This use case aims to digitalize the mapping of existing networks to be more precise and have the possibility to create new services and serve new clients on this basis. It would be made possible by acquiring high precision aerial images (from 20 to 5 cm/pixel) and then recover, transform, store and disseminate the aerial images acquired.

This use case is part of the New Large Scale Mapping project from Enedis and is currently facing the following major issues:

- Human factor: Transfer between Regional direction and Supplier by physical exchange of hard drives;
- Integration & Injection into the IS: Data processing for customization;
- Storage: Large volume of data (75 TB);
- Service exposition: Consumption / data visualization / Ortho-photos via APIs.

Problem solved

The current process needs to be industrialized, the exchanges need to be controlled and secured, the volumetry of the exchanges needs to be put under control.

Partners/Ecosystem

Main stakeholders impacted will be DSOs.

Main technology/Gaia-X components

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

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification

Concrete benefits

- Improve the process to secure and track photo embedding
- Set up a scalable, reliable, robust and autonomous orthophoto service
- Ultimately allow new services to be opened and new users to be served

Use Cases Compliance and standards

Compliance – Green certifications

Solution

Within the Gaia-X Energy Space our solution is providing a certification service for green energy. The goal is to issue automated, timely and governmental approved sustainability certificates for Energy, cross- Sector and along the entire value chain. Covering certification of e.g. Electricity, Gas Hydrogen, Green- Fuels and other green goods

The initial solution scope could be as follows:

- Defining and implementing governance structure for partner ecosystem
- Define certificate standards for different energy sources
- Building decentral partner ecosystem and providing decentral digital identities for assets producing or consuming green energy (e.g., Wind turbine, PV-plant, Electrolyzer, Methanol- Synthesis-plant, Steel-plant)
- Connect decentral digital identities with asset sensors
- Implement proven standards into the certification management scheme
- Prove and confirm trustworthiness of implementation
- Issue certificates related to asset sensor data in asset specific wallets
- Transfer certificates between wallets (market participants) based on defined standards
- Devaluate Certificates based on timely (e.g., 15 minutes) asset sensor data
- Defining and implementing automated payment scheme for certificate management

Problem solved

Provide a working example. Prove that trustworthy and automated issuing and management of sustainability certificates works across multiple Energy sources and industrial Sectors as well as across EU countries. Demonstrate that decentral data ecosystems can connect different industries to prove the sustainability of produced goods in regards of energy

Partners/Ecosystem

Define among Gaia-X partners an eco-system for green energy certification. A small group of partners will first refine the exact scope of “Green Energy Certification”. Define the governance and digital infrastructure layer to offer a common structure for building digital services around cross-sector energy certification. Bring a trial implementation to production and search for voluntary contributors (individuals and organizations) to publish their certificated within the eco-system. Publish certificates cross sector to prove the usage of sustainable energy from e.g. wind power plant to end-customer (e.g. green fuels) along the entire value chain.

Main technology/Gaia-X components

This demonstrator would require following the Gaia-X infrastructural components:

- Identity & Trust: federated identity management for individuals and organizations
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification and it would require green energy certification standards:

- Wherever possible we would align with current certification proposals coming from electricity standardization (GO, HKN) gas or hydrogen concepts currently under discussion by CertifHy, TÜV SÜD, AIB, etc.

Concrete benefits

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

Compliance – Existing standards integration to Gaia-X

Solution

Gaia-X will be a non-domain specific platform and will be used in various contexts and use cases. However, those domains have a specific vocabulary as well as corresponding basic communication stacks and data models. The use cases already envision common semantics as data models for communication or automation and control of e.g. DER /DES but there is a strong need to foster the use of domain specific standards whenever possible and be open to incorporate and mediate to those legacy systems and technologies. From the UC point of view, one particular issue will be the convergence of both OT (operational technology) and IT (information technology) in the future Smart Grid. As data from IT and historical archives will be used for real-time optimization in OT (grid control contingency etc), there is the danger of format and semantics transformations which should be prevented. Given approaches like OPC which helps to use domain specific data models as address spaces, such an open approach shall be considered and documented for Gaia-X infrastructure services. Existing standards shall be screened for compatibility with the envisioned Gaia-X infrastructure and fallacies and gaps be documented.

IT/OT

Problem solved

NIST identified 75 existing standards and 15 high-priority gaps in support of smart grid interoperability, in addition to cyber-security issues, as a starting point for standards development and harmonization by standards setting and development organizations (SDOs like IEC and CEN/CENELEC). Sixteen Priority Action Programmes (PAPs) have been initiated by NIST to address areas in which standards need revision or development to complete the standards framework according to their smart grid vision. In addition to the US perspective, the IEC Standardization Management Board (SMB) of Technical Committee (TC) 57 identified over 100 standards and standard parts in a strategic review of power system information exchange. Both of these studies concluded however, that only a small number of standards lie at the core of smart grid interoperability and they can be organized into a corresponding layered reference architecture described in IEC/TR 62357 – the so-called SIA – Seamless Integration Architecture. The evolution of IEC/TR 62357 reflects the broadening scope of TC

57 in step with smart grid use cases from its original charter of “Power System Control and Associated Telecommunications” to “Power System Management and Associated Information Exchange.” Generally, this change reflects the shift in emphasis from lower level interconnection automation OT protocols to abstract information models in the higher levels of the architecture in IT as the number of business functions needing to interoperate with PSAs has increased with smart grid evolution. The TC57 architecture generally follows the form of the GWAC Stack layers 1-7, as it ascends from standards

concerned with communications relating to the connectivity of field devices through to information exchanges to support business processes and enterprise objectives. This reference SOA blueprint shows how these standards relate to each other, require harmonization and presents the gaps where further standards development work is required. In general, all standards settings and development organizations advocate a collaborative approach to the development of open standards for the smart grid, with the reuse of existing standards as far as possible. Gaia-X will be a non-domain specific platform and will be used in various contexts and use cases. However, those domains have a specific vocabulary as well as corresponding basic communication stacks and data models.

Partners/Ecosystem

Partners participating can be from greatly varying domains as this is a process which has a lot of stakeholders involved. We envision at least

- TSO/DSO utilities,
- Regulators and markets (market platform, clearinghouses),
- DER aggregators,
- SDOs from the domain

Main technology/Gaia-X components:

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

- Identity & Trust: federated identity management for individuals and organizations
- Federated Catalogue: to publish the registration, consent and query services
- Sovereign Data Exchange: to manage consent and usage control
- Compliance: rights management, onboarding and certification

And it would require individual functional use cases using standards to be used as demonstrators in order to better understand the integration of legacy data models and systems in order not to cause problems with technical debt when introducing the Gaia-X platform.

Concrete benefits

The cost to fix a software defect varies according to how far along you are in the cycle, according to authors Pressman / Grady. One of the main costs drivers is the integration of components and system from various heterogeneous software or system vendors. Integrating based in standards lowers the amount of coordination and integration tests needed and fosters faster integration with less errors. As integration occurs in the later stages of a software project, costs of failed early interface semantics will cause a high maintenance and integration problem in the later stages.

Compliance – Trusted HUB

Solution

This use case focuses on the design and implementation of a managed service to address the demand of privacy-preserving machine learning and multi-party computation in the Gaia-X ecosystem. Aggregating, combining, and analysing data – including data analysis, machine learning (ML), Artificial Intelligence (AI), and decision making – from different sources are becoming increasingly important almost in each and every domain from Energy 4.0 to Industry 4.0, Mobility, and financial sector. At the same time, this process is often complicated as relevant data is often privacy sensitive and created and owned by many different data owners that do not tend to share with others since when exchanging their data with third parties, they may not only reveal their business secrets but also lose control over their data and for what it is used. Many data owners are aware of the technical and economic potential that is realized by analysing their data, in particular, in combination with data obtained from other data owners. The process of combining and collaboratively analysing different data sources results in new insights, better AI/ML models, better decision-making as well as new/improved data-driven products and services. Considering this fact, this use case geared towards the integration of “Trusted Data Hub,” a Hardware-Software privacy-preserving ML and multi-party computation solution, enabling several different parties to jointly analyse data, just as if they have a shared database without ever revealing those data. In other words, sensitive data sources held by multiple parties can be linked together in a secure manner, while parties gain no additional information about each other’s sensitive data, except what can be learned from the output of data analysis.

Problem solved

AI/ML is widely used in many areas of the energy domain, from energy fraud detection to theft detection, anomaly detection of energy consumption, energy demand prediction, demand response management, renewable energy forecasting, planned/unplanned disruptions forecasting in the power grid, outage detection and prediction, predictive/preventive equipment maintenance, and energy trading, among others. Machine learning and big data solutions enable energy and utility companies to optimize their resources, improve energy flows, manage the grid, schedule energy, and prevent mistakes. Unfortunately, the utility of AI/ML solutions is currently hindered by limited data availability for algorithm training and validation due to the absence of standardized data sharing/exchange as well as the

requirements and concerns to protect the privacy of data owners and parties participating in the energy ecosystem. Although International Data Space (IDS) and Gaia-X Data Space can partially address the first issue, the development of new solutions to concurrently address the demands for privacy and ML utilization is a necessity. Trusted Data Hub aims to bridge this gap by providing a secure, privacy-preserving, and multi-party platform preventing data owner's privacy compromise and protecting data leakage.

Partners/Ecosystem

We envision the following tentative stakeholders participate in this use case:

- Distribution System Operators (DSOs)
- Contractors
- Utilities
- Gaia-X services platform
- Consulting companies and solution providers
- Startup companies interested in new business models

Main technology/Gaia-X components:

This demonstrator would require the following tentative components and interfaces of Gaia-X:

- Identity Provider: federated identity management for individuals and organizations
- Broker: to address the registration of self-description metadata of datasets, connectors, policies, and queries
- Connectors: to support data exchange
- Compliance: rights management, onboarding, and certification

Concrete benefits

Privacy-Preserving and Multi-Party Computation play an important role in the data economy and the spark of innovative new business models. It bridges the gap between the utilization of AI/ML services and the privacy of data owners enabling transparent aggregation, trustworthy refining, and collaborative analysis of data sources to be provided as a new product/application on the energy markets.

Maturity indication of the data space, health status

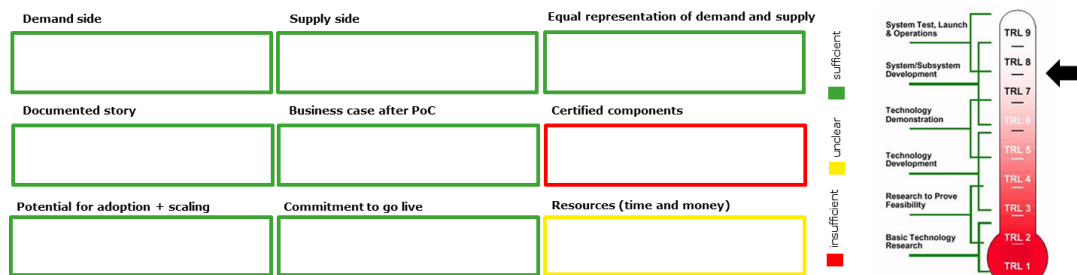


Figure 8

How is the demand side represented?

The demand side is already well represented and the ambition is to expand further:

- There is a broad representation of Energy Companies from Germany, Belgium and France at the moment
- There are companies and use cases in all segments of the energy value chain: production (renewables, hydrogen and nuclear), networks (electricity, gas, heat, etc.), marketing, sales and compliance
- EDF, Enedis, Engie, E.ON SE, Westenergie AG, Arge Netz GmbH & Co. KG
- Elia, Enel, Total, Air Liquide (to be confirmed)

This has to be completed by each Hub / country if necessary.

How is the supply side represented?

These energy companies are already able to develop smart services (AI, IOT, Machin Learning, Blockchain...) and to integrate suppliers to develop these services.

Technology companies able to work within Gaia-X to provide infrastructures and data protection services are also stakeholders of the data space (Capgemini, Atos, Sopra Steria, Advance GmbH, Siemens energy AG, Offis e.V., Fraunhofer e. V, Examesh GmbH confirmed). HPE and IBM to be confirmed.

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

There is a very good equilibrium between demand and supply side in the data space Energy, ensuring the achievement of the roadmap with sustainable business models.

Is the story of the data space well documented?

The use cases are already well documented. The data space energy technical roadmap of Gaia-X infrastructures and services labelled by Gaia-X will be designed during the first semester of 2021.

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

The business model of each use case is or will be described by each use case working group. A pipeline of use cases will be organized with 4 different steps:

1. Use cases project mapping;
2. Use cases projects prioritization;
3. Use cases implementation - phase 1, for use cases already close to implementation;
4. Use cases Implementation - phase 2, for use cases needing R&D.

Several use cases are already detailed and a first prioritization has been done, identifying “quick wins” and “high value use cases”. New use cases will enter this pipeline in an iterative way, when data space’s stakeholders identify new subjects or are joined by new participants and experts.

The business value and the potential for adoption and further scaling are key criteria of prioritization. Before implementation, each use case will be further deepened, especially on two aspects:

- Consortium building if needed, with relationships between all parties definition and necessary resources identification;
- Potential funding identification.

The use cases pipeline will be developed in the “roadmap of the use cases” of the data space energy, defining priorities and timing of implementation of use cases (see 4.1.)

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

All the main component of Gaia-X will be certified:

- Identity & trust:
 - Federated identity management
 - Trust management
 - Federated access
- Federated catalogue:

- Self-description
- Service governance
- Monitoring & metering
- Sovereign data exchange:
 - Policies & usage control
 - Usage control for data protection
 - Security concepts
- Compliance:
 - Relation between service providers and consumers
 - Rights and obligations of participants
 - Onboarding & certification

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

The potential for adoption and further scaling are key criteria of prioritization (see 3.9.5.).

How can the commitment of the parties involved be proven?

Before implementation, each use case will build a consortium if needed, identify the necessary resources and the potential funding, ensuring all parties commitment (see 3.9.5.).

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

Resources are mobilized at 2 levels within the data space:

- **Use cases level:** staffed by companies and business divisions inside companies, which qualify and develop priority use cases. For now, several energy companies from different European countries are working to map and prioritize these use cases, in every segment of the energy value chain. For further use cases implementation, necessary resources will have to be defined (see 3.9.5.);
- **Core Team level:** Participants are, for now, EDF, E.ON, Elia (to be confirmed), Engie, Enedis, Capgemini, Atos, Offis and Mines Telecom Institute. They organize and define the data space vision and objectives and help the use cases working group to remain aligned with the data space ambition and Gaia-X guidelines. This core team also organize the pipeline of use cases. Resources are mobilized by companies mentioned above and are sufficient for the activities described.

Evolution of the data space

Roadmap of the evolution

The roadmap of the data space energy will include 2 roadmaps, that will be built and actualized in consistency:

- The roadmap of the use cases (based on the pipeline of use cases described above) - that will be precised in the next months;
- The technical roadmap of Gaia-X infrastructures and services labelled by Gaia-X, ensuring use cases development; there will be a common technical layer of infrastructures and services ensuring secured data exchange and interoperability.

Quick-wins (for 2021)

- Define the data space governance (how to enter the core team and the data space? what rights and duties of the core team and attendants?)
- Build a community of cross-border European actors
- Define the data space roadmaps, as described above
- Qualify first use cases and build some demonstrators for implementation in 2022
- Lay the foundations around existing shared semantics

Use cases quick wins will be detailed in the soon-to-come roadmap.

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

Among the already or soon-to-be-launched projects, we identified:

- Renewable asset description model
- Renewable works – risk prevention
- Hydrogen station networks information sharing
- Nuclear equipment regulation
- Local communities of energy setting up and decentralization
- EV assets investments
- DSO network mapping
- Energy renovation
- “Green” certifications

List to be completed with the soon-to-come roadmap.

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

- Contribute to the energy transition and carbon neutral economy, through new services and digitalization. Consistency between the Energy data space, the Green Deal data space or others (eg: mobility) will be ensured;
- Contribute to the development of European large cross-border infrastructures and data protection services;
- Develop large cross-border energy services and actors in Europe, able to be developed all around the world;
- Benefit to European citizens with better services, personal data protection, and job development in Europe.

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

Due to the benefits described above, there should be a great interest for EC and Governments to support the data space Energy project.

Finance

Expectations towards a Gaia-X roadmap for the European financial ecosystem

This position paper is based on a collaboration initiated in October between the French and German financial communities within Gaia X national hubs. The intention of this document is to share widely our ecosystem's expectations and to provide a first overview / indication as to a desired 2021-2025 roadmap.

We, banks, insurance companies, financial institutions, fintechs, cloud service providers, software editors, academic institutes and governmental agencies of **the German-French Gaia-X Financial community, expect strong business benefits and tangible outcomes from the creation of the Gaia-X ecosystem, and actively support the emergence of federated service & data platforms that create value for all our businesses– and ultimately to European citizens.**

Up to now, most of **the users among us (banks, insurance companies, financial institutions) have been lacking a trusted cloud service & data space** between our own internal cloud and data platforms and international cloud platforms operating under extra-territorial laws. This has prevented us from accelerating the use of trusted cloud and data-based services and ecosystem-based models that we foresee as providing a competitive advantage to our businesses. And because of the lack of trusted cloud technical standards, **the solution providers among us (fintechs, cloud service providers, software editors) have been hindered in the growing of our cloud business** towards the financial industry, slowing down the necessary transformation of users' operating and business models required to compete more efficiently.

Through the Gaia-X initiative, we – at last – foresee a tangible and close way forward to easily build, assemble and use trusted and value-creating data-based cloud services, as well as create new products/services and foster new business models that are compliant “by design” with European regulations and values, and will fully take advantage of the massive European potential in terms of technical expertise and market size. **We see Gaia-X as a European-wide accelerator of innovation,** scaling up open innovation and co-construction by providing secured data sharing and artificial intelligence services at scale, in a compliant and secured way. By combining the forces of the finance industry with those of governments and academic institutes on the creation of a financial data space within Gaia-X, e.g. building on the Gaia-X „Financial Big Data Cluster“ initiative, the financial sector is contributing to the Gaia-X network **and will foster Europe's competitiveness and financial markets' stability,** through enhancing research, AI development, disruptive innovation and stability in the legal environment.

We fully support the Gaia-X strategy of enabling access to aggregated, federated and interoperable trusted cloud services, data spaces and AI services through the setting of common policy rules that will ensure data and services protection, interoperability and portability. This strategy is to us the most efficient way to build on existing cloud services and accelerate the development of new services that leverage the very dynamic European financial innovation community.

Through many European and national regulations (from the European Commission, the European Central Bank, national regulators in France and Germany, national cybersecurity agencies...), **we are a regulated industry which needs a sound level of data management – this both justifies and requires the Gaia-X initiative** in order to accelerate the development and use of trusted data-based services. **We should benefit from the expected consistency of Gaia-X policy rules with the financial sector regulations; and we expect Gaia-X leadership to rapidly set up a framework that fulfils the requirements induced by the financial regulations.** Only this framework and ultimately its widespread use by solution providers will unlock the full potential of European trusted cloud services towards the financial industry.

Thanks to the incredible political momentum around Gaia-X in recent months – and we would like to acknowledge our tribute to Gaia-X founding members as well as to the French and German governments – **we have been recently gathering our forces in order to share and challenge the actual business benefits that we could expect from the Gaia-X ecosystem that are presented below.**

2021-2022: “quickwins” rapidly available

- **A “compliance by design” framework set up by Gaia-X governance**, in connexion with European and national regulators and governments
- **Secured innovation and experimentation platforms to support co-innovation with multi-entity participation.** Exploring future AI and data-driven use cases, for the financial industry, supervision, competent authorities and academia as well as for cross-industry use cases, e.g. through the start of the European FBDC platform including a demonstrator case.
- **“Compliant by design” electronic/intelligent document & content management cloud services:** a large array of tools has existed for long in Europe, but has been regularly hindered by the lack of an operant framework, especially as to emerging cloud services (i.e. document scanning, text-to-speech...)
- **Authentication, identity management, consent management and ePrivacy cloud services**, which could be cross-dataspace services within Gaia-X
- **Acceleration of the development of trusted open banking services**, capitalizing on the building of European industry-wide standards
- **Trusted Fraud, Risk and Compliance data-based cloud services**, in the domains of fraud detection, fight against money-laundering and terrorism financing, regulatory reporting...
- **Trusted financial risk management cloud services** in the domains of market trading risk assessment, credit risk assessment, ALM simulations... that need massive computation and strong protection of data used

2022-2023: mid-term benefits building on already-launched or soon-to-be-launched projects

- **Stable Supply Chain Finance**, by fostering the synchronisation of physical goods and financial flows along the industry 4.0 supply chains, through cross-industry data sharing
- **Financial cloud services' tokenization platform**, that would facilitate usage-based services' invoicing, secured data access and brokering, down-the-supply-chain contracting and invoicing with subcontractors...
- **Trusted "Blockchain as a service" platforms**: providing trusted hosting for blockchain nodes, smart-contract execution, digital money transfer...
- **Collaboration applications suites trusted cloud services**: leading suites' cloud services, key to financial industry productivity
- **Personal Data digital safe**: at any citizen's initiative and under their control, storage of personal data in a
 - trusted and secured space that could be used by cloud services without copying data
- Building **new services in a directly Gaia-X compliant mode**, such as European Payment Initiative (EPI), Interbank clearing, Clearing house (FPML)
- **Sustainable finance data spaces and innovative AI and machine-learning cloud-enabled services** that analyse the effects of sustainability risks and derive relevant methods, models and database structures for sustainability, climate risk and impact analyses
- **Enhancing market integrity** in the context of market supervision of stock exchanges, thanks to a better detection of market manipulations through new AI methods
- **Improving the data basis for monetary policy decisions**, through the optimisation of the information base for monetary policy (improve the empirical data basis) and the use of AI.

2023-2025: long term benefits requiring significant investments on the 2021-2025 period

- **The emergence of a European financial market-data platform** that would create an alternative to non- European financial market data providers, e.g. by completing the set-up of the European FBDC platform.
- **Trusted "core banking as a service" cloud platforms** that would benefit to the European financial industry through standardization and would strengthen European solution providers' worldwide market position

Health

Executive summary

Gaia-X² 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.

² 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 Framework³: 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 GDPR⁴ or national legislation.

³ European Interoperability Framework (EIF): https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

⁴ 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 EU⁵ 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

⁵ EU recovery plan: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_940

Commission⁶ and the European Parliament policy documents⁷ plead strongly for the digital sovereignty of European citizens. The funding priorities in the Digital Europe Programme⁸ 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 framework⁹ 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.

⁶ EC strategy for the European Health Data Space: https://ec.europa.eu/info/sites/info/files/soteu_2020_en.pdf

⁷ EU parliament policy document: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI\(2020\)651992_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf)

⁸ Digital Europe Programme: <https://data.consilium.europa.eu/doc/document/ST-6789-2020-INIT/en/pdf>

⁹ 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 information¹⁰:

- 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 9 **Fehler! Verweisquelle konnte nicht gefunden werden.**, with a non-limiting overview of data holders and users within the major 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.

¹⁰ 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>

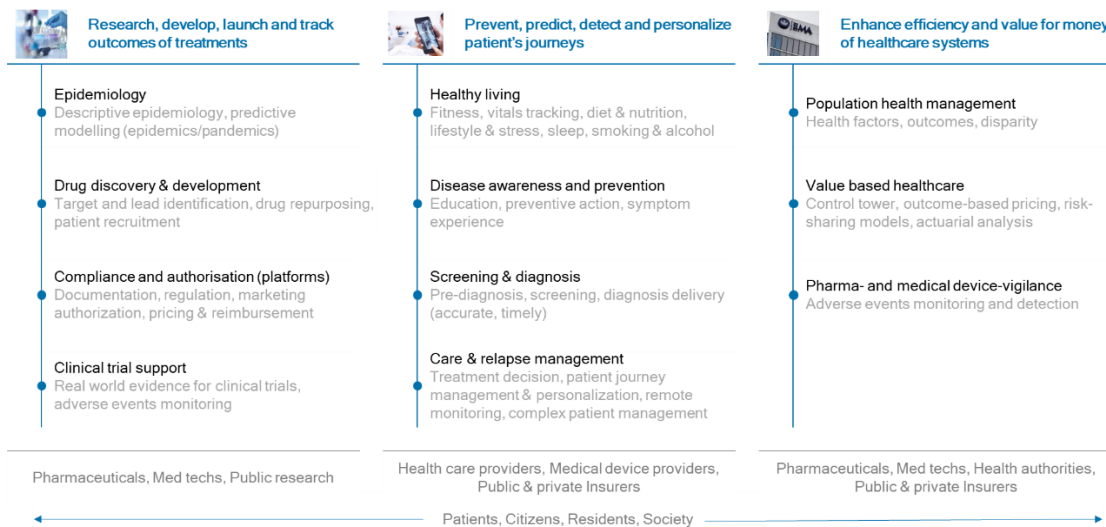


Figure 9: the health domain stakeholder landscape

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’ infrastructure¹¹ or it could be established by ‘hard’ law¹². 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

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

¹¹ Soft infrastructure introduced by Innopay & Sitra: <https://www.innopay.com/en/media/616/download>

¹² 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>

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 10 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.

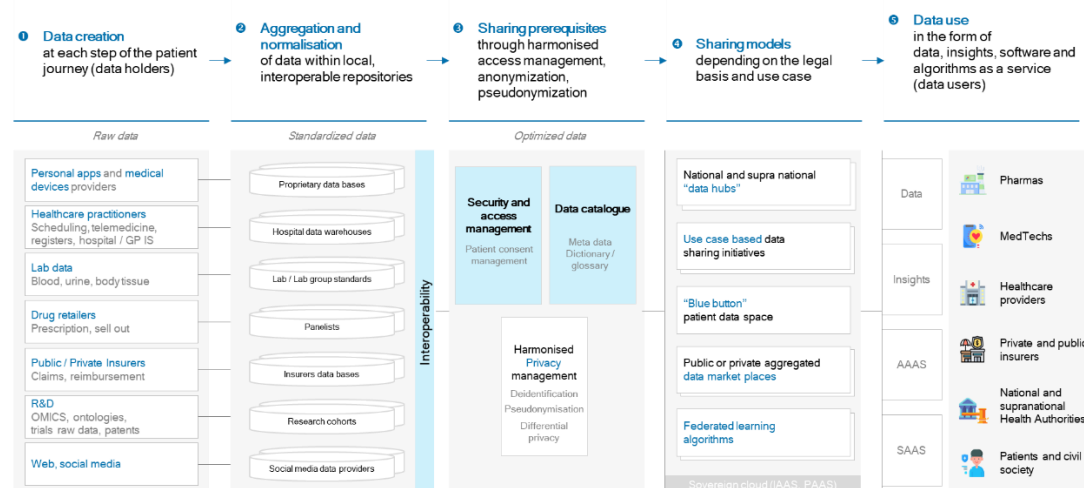


Figure 10: the data value chain in a desired state Use-cases (scenarios) within the data space and the current state of development

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. EMA¹³, ECDC¹⁴, EHDS¹⁵).

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.

¹³ European Medicines Agency (EMA): <https://www.ema.europa.eu/en>

¹⁴ European Centre for Disease Prevention and Control (ECDC): <https://www.ecdc.europa.eu/en>

¹⁵ European Health Data Space (EHDS): https://ec.europa.eu/health/ehds/dataspace_en

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.

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 definitions¹⁶ for: BodyWeight, Gender, Birthdate, Address
- (could be) SNOMED CT coding for semantic concepts¹⁷
- (could be) cross-sectoral standards for non-healthcare specific concepts¹⁸

¹⁶ HL7 FHIR standard: <https://www.hl7.org/fhir/>

¹⁷ SNOMED CT coding: <https://www.snomed.org/snomed-ct/five-step-briefing>

¹⁸ Example standard for addressing: ISO 19160: <https://www.iso.org/obp/ui/#iso:std:iso:19160:-1:ed-1:v1:en>

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)¹⁹.

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 Action²⁰ 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.

Use Case 1: Research Platform Genomics

Solution

This solution implements a cloud based genomics platform, compliant with the Global Alliance for Genomics & Health (GA4GH²¹), 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:

- 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

¹⁹ eHDSI: <https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHealth+DSI+Operations+Home>

²⁰ TEHDAS Joint Action Towards the European Health Data Space: <https://tehdas.eu/>

²¹ Global Alliance for Genomics & Health: <https://www.ga4gh.org/>

Partners/Ecosystem

First examples of existing services and compute services from the de.NBI cloud²² 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.
- 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)²³ as a secure and trusted long-term archive for human omics data.

²² de.NBI cloud: <https://www.denbi.de/cloud>

²³ The German Human Genome-Phenome Archive: <https://ghga.dkfz.de/>

- The developed GA4GH compliant analysis pipelines will be used in the Beyond 1 Million Genomes²⁴ 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.

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)²⁵ profiles known as Cross Enterprise Document Sharing for Images (XDS-i) and Cross Community Access (XCA)²⁶. 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 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

²⁴ Beyond 1 Million Genomes: <https://b1mg-project.eu/>

²⁵ Integrating the Healthcare Enterprise: <https://www.ihe.net/>

²⁶ IHE recipe for securely sharing health information: <https://www.ihe.net/news/ihe-hie-white-paper/>

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).

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)²⁷. 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.

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.

²⁷ Medical Informatics Initiative Germany: <https://www.medizininformatik-initiative.de/>

Pre-existing efforts by the MII²⁸ 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
- Enabling secure and GDPR-compliant access to patients' data through compliance with Gaia-X provided policies.

Use Case 4: Health Outcomes Observatory

Solution

The Health Outcomes Observatory project (H2O)²⁹ 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:

²⁸ MII consent, data sharing and interoperability working groups: <https://www.medizininformatik-initiative.de/en/collaboration>

²⁹ Health Outcomes Observatory: <https://health-outcomes-observatory.eu/>

- 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.

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 pharmaceutical³⁰ 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 website³¹. 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

³⁰ European federation of pharmaceutical industries and associations (efpia): <https://www.efpia.eu/>

³¹ H2O partner list: <https://health-outcomes-observatory.eu/partners/>

- 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 page³².

³² 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)³³ is a well-documented attempt to enable health data exchange across European country borders and the Joint Action (TEHDAS)³⁴ helps the Member 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 principles (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 demonstrators of the federated Health Data Space. Thereafter, usage-based payment models will fund the sustainable operation of infrastructure and services.

³³ eHDSI: https://ec.europa.eu/health/ehealth/electronic_crossborder_healthservices_en

³⁴ TEHDAS Joint Action : <https://tehdas.eu/>

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 of 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 statista³⁵, harmony healthcare IT³⁶, businesswire³⁷, emerj³⁸ 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 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 evaluate 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

³⁵ Statista: <https://www.statista.com/statistics/1037970/global-healthcare-data-volume/>

³⁶ Harmony HIT: <https://www.harmonyhit.com/health-data-volumes-skyrocket-legacy-data-archives-rise-hie>

³⁷ Businesswire : <https://www.businesswire.com/news/home/20181126005585/en/Seagate-Launches-New-Data-Readiness-Index-Revealing-Impact>

³⁸ Emerj: <https://emerj.com/ai-sector-overviews/where-healthcares-big-data-actually-comes-from/>

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 Programme³⁹, Horizon Europe⁴⁰ 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.

³⁹ Digital Europe Programme: <https://digital-strategy.ec.europa.eu/en/activities/digital-programme>

⁴⁰ 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 patterns 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 Alliance⁴¹. 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.

⁴¹ 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 programme⁴². This is an ongoing activity accompanying the use case proliferation.
- Extend and refine the **common enablers** 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)⁴³.
- 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.

⁴² Digital Health Europe twinning scheme: <https://digitalhealthurope.eu/twinnings/>

⁴³ 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 Board⁴⁴ 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.

⁴⁴ 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 use 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 of 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://eucancom.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 ecosystem: <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/): <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.

Industry 4.0

Plattform Industrie 4.0 invites to shape tomorrow's digital ecosystems

In the era of the digital economy, data-based value generation will depend on global data spaces that safeguard and guarantee data sovereignty, data security and data integrity. Plattform Industrie 4.0 drives the formation of the Data Space Industrie 4.0, a data space that accelerates future value creation in industrial manufacturing and production.

Defining the Data Space Industrie 4.0: Enabling the paradigm shift towards multilateral collaboration

Typically, today's business value in production and manufacturing is generated in a global production value network. This network is built on single, bi- and multi- lateral business relationships among organizations in the production ecosystem. It is highly complex, globally widespread and involves multiple players and stake- holders (e.g. OEMs, suppliers, service vendors, etc.). They are cooperating across different regions, markets and increasingly also industries. Flexible production value chains are being built dynamically inside the pro- duction value network based on bilateral agreements to fulfil specific business objectives.

However, this bilateral nature of cooperation hinders multi-stakeholder collaboration down- and upstream the value chain. It lacks the requirements of modern business and regulatory needs: Use cases with both highest business and societal impact. Examples, that rely on full digital transparency and an integral "sovereign flow of data" within the entire data ecosystem are

- efficient supply chain traceability and quality management,
- modular production and manufacturing-as-a-service or
- a sustainable, CO₂-neutral circular economy.

To realize these highly impactful use cases, depends on a fundamental paradigm shift in the global production ecosystem: To satisfy future business requirements, organizations must evolve from bilateral collaborations. In many cases, these are restricted to individual first tier relations along the production value chain. Instead, multilateral collaboration, allowing end- to- end data exchange in a fully interoperable and sovereign data-oriented ecosystem is required: The Data- Space Industrie 4.0.

Multilateral collaboration and data-sharing – two directions of impact, one target

With its 10th anniversary, many manufacturers world- wide already benefit from Industrie 4.0 implemented on the shop floor. Now, established Industrie 4.0 concepts and solutions must be seamlessly integrated to initiate the next stage of the fourth industrial revolution: The implementation of digital, cross-enterprise collaborations within a data-value-oriented ecosystem that provides flexibility and supports dynamic changes.

An "enabling structure" for a competitive digital ecosystem requires business-related mechanisms as a basis to collaboratively share data within an interoperable (scalable by design) and sovereign architecture. This fundamental architecture for collaborative data sharing will not only streamline current business processes and enable future business cases. It will also provide the foundation to

develop new business models and new potential revenue streams for existing organizations and emerging start-ups in the production landscape, e.g. with use of data-marketplaces

ILLUSTRATIVE

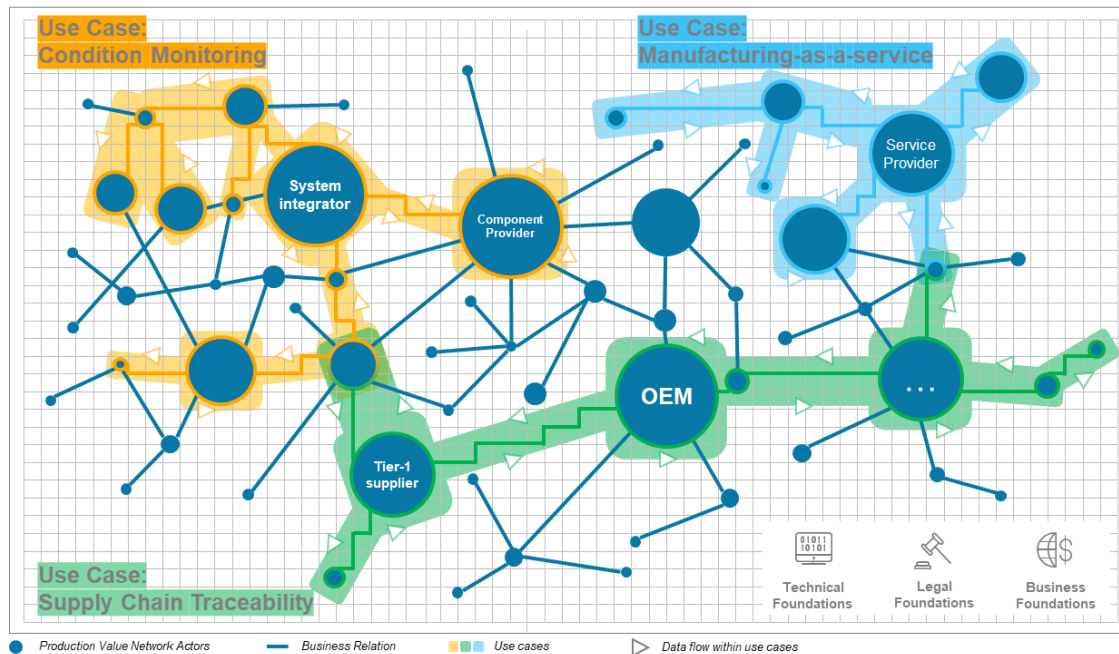


Figure 11: The DataSpace Industrie 4.0 lays a „collaborative mesh“ of technical, legal and business foundations underneath the production value ecosystem. Organizations can use the „collaborative mesh“ as common basis for multilateral data-sharing and collaborati

The Data Space Industrie 4.0: A common set of technical, legal, and business foundations

A common base foundation must be established and agreed among all stakeholders. This enables multilateral collaboration and data-sharing based on trust, integrity, security, and individual sovereignty in the Data Space Industrie 4.0. A basic set of precompetitive foundations provide the “collaborative mesh” underneath the digital production ecosystem. Organizations can build on such a mesh to initiate and accelerate multilateral collaboration and data-sharing.

From today’s perspective, the architecture of the mesh is defined by examining at least three essential dimensions on the production landscape that must be considered comprehensively: business, legal, and technical foundations. Consequently, the Data Space Industrie 4.0 is described by a set of common basic foundations all stakeholders agree on. In this sense, the Data Space Industrie 4.0 does not include any individual business processes, specific technologies or use cases. The foundation of the data space rather prescribes frameworks, routines, standards, and guidelines that are agreed upon in a precompetitive context. It can be used to initiate efficient crosscompany collaboration and data-sharing.

Plattform Industrie 4.0: Implementing the Data Space Industrie 4.0 as a joint approach of public and private stakeholders

In recent years, the partners of the Plattform Industrie 4.0 have jointly developed concepts and endorsed solutions alongside the various dimensions describing the Data Space Industrie

4.0. With the Asset Administration Shell (AAS) a coherent information model has been developed and approved by industry to ensure cross-company interoperability. Further models describe industrial specifications for e.g. secure and trustworthy collaboration as well as the fundamental legal requirements associated with digital ecosystems.

Based on a well-established structure and world-wide network, the Plattform Industrie 4.0 together with its partners will now strive to describe a coherent architecture of precompetitive foundations defining the Data-Space Industrie 4.0:

Merging of individual perspectives into a coherent architecture for the Data Space Industrie 4.0 – In a multi-stakeholder approach recent concepts on technical, legal and business foundations are developed and agreed within the Plattform Industrie 4.0. The aim is to create a coherent set of specifications defining the Data Space Industrie 4.0 in accordance with industry needs.

Implementing prototypical industrial use cases for the Data Space Industrie 4.0 – With the implementation of lighthouse projects the business, ecological, and societal benefit of the collaborative sharing of data will be demonstrated. This happens through representative use cases with a strong footprint, e.g. in the automotive or process industry. We will address questions of architecture, governance, and business models to assess value creation. This will be the launchpad to start the journey on establishing pragmatic and ready-to-use frameworks that reduce scepticism and reluctance towards change. Consequently, results from these pilots will be communicated and transferred into the broader industrial landscape to enable a broad market uptake of the digital ecosystem.

Building the Data Space Industrie 4.0 on Gaia-X – The Data Space Industrie 4.0 and related business models fundamentally rely on a cloud/edgecloud infrastructure. This allows for a sovereign and transparent hosting and processing of data and services. With the European initiative Gaia-X numerous stakeholders are setting the foundations for a federated cloud infrastructure, empowering the creation of dataspaces in various sectors. The architecture for the Data Space Industrie 4.0 will be developed in close collaboration with the European Gaia-X foundation and national Gaia-hubs. In this way industrial requirements can be organically integrated within the Gaia architecture and vice versa.

Fostering a world-wide lead-market for digital industrial solutions in Europe – We will advocate the broad roll-out of common foundations for the Data Space Industrie 4.0 throughout the industry – in close alignment with European partners from industry, government, academia and the European Commission. This initiative will boost a collaborative culture that fosters digital entrepreneurship, promotes agility, and reduces reluctance against progressive change. The implementation of the Data Space Industrie 4.0 will accelerate the digital transition of the industry into the digital age.

The creation of the Data Space Industrie 4.0 will be successful as a joint effort that is orchestrated on European and international level. Here, Plattform Industrie 4.0 will develop governance structures that moderate and perpetuate multi-stakeholder dialogues among diverse international organizations. National and international organizations are invited to engage in this process and shape the tomorrow's digital ecosystems.

With ten years of experience Plattform Industrie 4.0 now pioneers the next phase of the digital transformation: Industrial concepts and solutions from different perspectives and disciplines developed so far, will be brought together into a coherent foundation for the data space, facilitating the >> 2030 vision of Industrie 4.0: An interoperable, sovereign and sustainable frame- work – enabling global, digital ecosystems!

Mobility

Expectations towards a Gaia-X roadmap for the European Mobility Dataspace

This position paper is based on a work initiated in October 2020 by the Mobility Dataspace within Gaia-X. The intent of this document is to share widely our dataspace's vision, mission and expectations, then to provide a first overview as to a desired 2021-2025 roadmap.

Mission and Goals of the Mobility Dataspace

The use of transport and mobility services needs and generates an increasingly important volume of information and would greatly benefit from a frictionless flow of both **open and private data**, currently siloed for various reasons, mostly in relation to legal, regulation, property, security and technical aspects. The members of the Mobility Dataspace, as providers of mobility services, aim to **foster a system of data exchange to deliver more value to the users, improve mobility services, as well as to increase business productivity and innovation in the sector**. However, by design, the dataspace architecture allows all the participants to it to **remain fully in control** about what they share, to who they share and at what price they share, contrary to a former vision of closed data ecosystems, like data lakes.

By focusing on **enhancing user experience in mobility, while creating new business opportunities for the ecosystem's members**, our working group aims to:

- offer end-to-end seamless and safer mobility for users,
- increase productivity for the various stakeholders of the mobility and digital sectors, whether public organizations, large private companies, or SMEs,
- be fully aligned with EU and National government priorities: Privacy Protection, Digitization, Environment, Industrial development, Territorial cohesion, etc.

We therefore actively support the emergence of federated service and data platforms which can provide for an **easy and trusted exchange of data**.

Challenges to be addressed

This mobility initiative seeks to address several challenges. Firstly, it should meet users' expectations to **benefit from seamless mobility, notably between different geographical areas (such as countries, regions or cities areas) and different mobility modes**. Therefore, the highest-level objective is to improve the mobility user experience and to make Europe the leader in this aspect, even considering the complexity of its political structure.

This initiative should ensure **data security for the mobility user, like for any other EU citizen**, and above all, assure that data shared is trustful, and cannot be used for another purpose than the one agreed by the user. Notably, it should also affirm that players are **compliant with the GDPR regulation**, this includes where the data is stored (EU territory or compliant with European Commission Adequacy list for instance) and what kind of data would be transferred abroad.

Participants to the Mobility Dataspace would share their data through **an infrastructure that they can trust**, and that they know will **respect "by design" European values for data**, independently of where and how they share their data. As several participants need to transfer

customer information outside European territory, data exchange must convey sufficient “GDPR qualification” to allow full compliance, traceability and audit at any time.

This initiative will also allow for the **creation, collection and sharing of open data** to foster **innovation** where possible in this field.

Another challenge regards the wide spread of storage options for data in the extended mobility ecosystem. Players in the mobility ecosystem are rapidly turning to cloud infrastructure systems to solve their data management requirements. These cloud infrastructure systems may be very different in functionality and operation. Therefore, the Mobility Dataspace should **ensure interoperability** between the members’ cloud infrastructure systems.

Data to be exchanged between users and Dataspace members will be both **open data and private data**, each to be managed and shared according to radically different principles.

Such an infrastructure for the Mobility Dataspace should be effective enough to be connected to without major investment, allowing SMEs to participate and to gain from it (typically the small hotel chains or the local travel content providers proposing activities at destination).

The mobility user is at the center of a huge ecosystem, composed of private companies and public organizations where multimodal travels are structured by nodes. For instance, a **multimodal node can be an airport** interconnecting airlines, railways, public transports, taxis and vehicles with drivers. Therefore, this ecosystem brings together a variety of business sectors:

- **Airlines, railways and travel sellers** for instance which are usually linked to long trips;
- **Customs and other governmental agencies** take a crucial part of this ecosystem to secure national and European borders, especially in airports;
- **Travel by car such as taxis, vehicles with drivers, or carpool** and more generally the whole variety of travel modes;
- **Buses, metros, and suburban trains**
- **Hospitality businesses** like hotels
- **Tourism actors** such as tourism boards or associations;
- And eventually the financial layers which support these activities, such as **insurance and payment services**.

Other sectors are also linked to this ecosystem:

- **Automotive** is directly linked to travel by car, especially with the recent emergence of autonomous cars and the associated infrastructures;
- The pandemic shed lights on the link between international travels and travelers’ **health information**, provided by a trusted third party in most of cases;
- The **environmental impact of mobility** is essential and urges to monitor data linked to **energy** consumption through data transparency.

Mobility Dataspace Objectives

We have outlined below three maturity-based stages that can drive future cooperation between us.

Short term (2021):

- Structure the working group through an **appropriate governance** and the creation of a **legal body**
- **Define and refine two main use cases** for products and services around data, which will be validated by the executives and business of the members of this endeavor. The use cases could be chosen between:
 - **Seamless travel & Digital Identity**
 - Health pass
 - Multimodal travel
 - Loyalty
 - Local content distribution
- **Design and implement a demonstrator** based on one use case, more focused on travel since it corresponds to the business needs of the founding partners
- **Enlarge the working group** to partners which are willing to contribute improving the mobility experience in a safe and seamless way
- **Secure access to funds and establish basic business cases**

Mid-term (2022-2023):

- **Provide minimum viable products** for a wider range of businesses related to mobility, with multimodal travels at its center
- These minimum viable products will be focused on the business needs of major players, but the possibility for **SME** to benefit from them will be tested
- **Define the business models** of these products, and for the most advanced the business plans
- **Define a series of use cases**, enabling a variety of partners to join the dataspace

Long term (2023-2025):

- **Provide competitive positions** for the private companies and significant added value for public organizations
- **Provide these services and products worldwide**, to non-European users and mobility companies

Public Sector

Digital sovereignty means the capabilities and opportunities of individuals and institutions to perform their roles in the digital world independently, self-determinedly and securely.

Digital sovereignty means the state's ability to communicate and act digitally. The public administration of the federal, state and municipalities in Germany has set itself the goal of continuously strengthening the digital sovereignty of the administration in its various roles as user, provider and customer of digital technologies by expanding open interfaces, standards and open-source solutions, thereby reducing dependencies.

Gaia-X pays towards this goal and promises to build a secure, digitally sovereign, networked European data infrastructure. There are similarities of interest and considerable overlaps with the existing activities of the public sector in Germany, for which independence, self-determination, data security and data protection of the public administration are priority goals. In addition, expertise on key technologies and their evaluation competence, secure and trustworthy infrastructures, and framework conditions open to innovation within Europe are further goals.

The digitization of public administration in Germany needs to be accelerated even more overall. The potential of using data has already been recognized and awareness of the "power of data" has increased. Data is the basis of decisions, and its systematic use can help to provide public services faster and better and to make the public infrastructure fit for the future. In addition, data and the information derived from it can reveal room for maneuver and enable value creation as well as business models through data-driven innovations.

The declared goal of Gaia-X's Public Sector domain is to drive exchange between public sector players and public and private IT service providers as well as platform operators and to make joint efforts to build a digitally sovereign, resilient, and cross-sector data infrastructure and to improve data use overall. Given Germany's federal structure, this also means informing the administrations at all federal levels about Gaia-X even more than before and convincing them to participate.

The public sector must protect its IT systems, data centers, communications networks, hardware and software from failure or attack and find a strategic way of dealing with lock-in and lock-out effects. Public sector data processing is therefore subject not only to high data protection requirements, but also to particularly high security requirements. In data processing, an appropriate level of security in terms of the protection goals of confidentiality, integrity and availability must be ensured and regularly checked. IT systems must be continuously protected; information security is therefore a dynamic process.

The provision of public sector data is subject to a number of restrictions that must be observed. Personal data in particular is subject to special requirements. It must be that the protection of the rights of the data subject is guaranteed. Special requirements also arise when processing classified documents or data, e.g. VS/NfD. Anonymization and pseudonymization concepts, as well as procedures for regular evaluation and review, must be developed to enable the transfer of personal data, because many data have a personal reference. The necessary legal bases for this exist, but there is often a lack of legal certainty and clarity in the question of which data may be used, passed on and published.

The Public Sector domain has the claim to formulate the special requirements of the Public Sector for the emerging Gaia-X architecture concept and to work towards ensuring that these are taken into account. Identity checks, rights and role concepts, and a directory of available data (data atlas or data catalog) and digital services for further processing of the data are indispensable for the Public Sector when passing on data.

Another obstacle is that data is often not available in sufficient quality or in machine-readable form. In addition, data is usually published in such different data structures and formats that data from different sources must first be merged at great expense. Measures must therefore be taken to open up the administration's data silos, prepare the data in a legally compliant and demand-oriented manner, and make it generally accessible for digital use in high quality with open interfaces. Making the state a pioneer in the provision and use of data is the declared goal of the German government's data strategy, which still needs to be translated into state and municipal strategies. An important building block for this is, for example, increasing and perpetuating digital and data skills among administrative employees and promoting a culture of data sharing, including open data. This goal is explicitly supported by the Public Sector domain.

Data generated by the public sector should serve the common good. Within administration, data will play an increasingly important role because it enables evidence-based governance of policy and administration and user-friendly online services. Therefore, data should be made usable not only for public administration, but also for companies, startups, for science and research, and for civil society. Conversely, data from the private sector can also be used by public administration for its own purposes in the sense of datadriven government. The use cases within the domain exemplify which data is in demand and which applications can be developed from it.

In order to activate both the demand and the supply of public sector data, suitable data infrastructures ("data rooms") as well as concepts for data portability and interoperability must be established and tested. Currently, public cloud infrastructures and services are still in the process of being built, and some effort is needed to achieve the desired goals of interoperability and portability. It is necessary to dovetail the activities of the public administration in Germany in setting up a federal administration cloud with the activities of Gaia-X and to describe the special requirements of the public sector for Gaia-X's "Dataspace Public Sector". In this context, the publication of data as "open data" must always be taken into consideration.

At the European level, it is necessary to cross-link the players in public administration in order to identify which specific activities exist in the member states and how these can be coordinated if necessary. Activities of the Open Data Committee or the European Data Strategy 2020 and the revised Public Sector Information Directive should be taken into account.

In the medium-term, possible technical implementation concepts for public administration data rooms should be discussed. The existing administrative infrastructure is extremely heterogeneous. In Germany, the federal system means that the federal government, the states and the municipalities have built up their own infrastructures and that a large number of private and public IT service providers offer services within the public administration ecosystem. This diversity means that there is no comprehensive overview of the available data

and services. The Public Sector domain is working to ensure that consolidation potential is tapped in the medium-term and that such a catalog or directory is created, and to this end it will cooperate with other bodies pursuing the same goals. In this context, the offerings of private-sector operators, which are used by numerous actors in public administration, must also be taken into account in the creation of Public Sector data rooms and catalogs.

The concrete design of data rooms for the public sector involves legal, organizational and technical aspects. On the one hand, it is necessary that the existing federal and diverse system is orchestrated and joint agreements are made. With the IT-Planning-Council, there is a body that can be considered for this and must be involved. For technical implementation, platform and multi-cloud strategies must be developed and state-of-the-art approaches must be tested in practice, incorporating existing solutions and standards. Other aspects include the development of data governance models and the design of data trustee models. In addition, machine learning and predictive analytics applications and the resulting automation and simplification are becoming increasingly important in the public sector, and with them the availability of scalable platform and cloud services. Open and high-quality administrative data is required for application development and data preparation, including for deep learning methods. Appropriate capacities are already being built up for this in various places, e.g., at the platform and IT service providers, and initial architecture models are being developed. These need to be discussed and tested further and aligned with the activities of Gaia-X.

The European data platform Gaia-X, in combination with the "Public Sector Data Space" will become an enabler and marketplace for a new form of data economy for public administration in Germany and Europe. This offers the opportunity to make high-quality data available for further processing and provision of innovative applications.

German Gaia-X Hub, 25.03.2021

Members of the Domain Public Sector

Geoinformation

Mission and Goals of the data space

The development of the world economy and its industries - especially those in Europe, depend highly on the availability and accessibility of data. As in early studies on public sector information in Europe stated, more than 80% of data is somehow related to space. Spatial relationships of nearly every data type will lead to extensive leverage effects for economies and markets in different perspectives of sustainable development. Thus, it makes sense to develop and maintain a general basis, better: a network of geo platforms, geo services, and methodologies, to serve different industry demands on a technology open, common and harmonized geo data and geo information basis.

Europe is aware of the high impact of spatial data, information, and relationships: this is proven by the revision of the Open Data Directive of the European Commission, as most of the High Value data Categories (geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, mobility) are clearly related to space.

During the past decades, the gathering, maintenance, and provision of geo information was achieved by several different spatial information systems, like cadaster, utility networks, roads, topography, environment, space, meteorology, water, shipping, defense, and so forth. Those systems serve lots of administrative processes in rural and urban planning, smart city development, property registration and ownerships, environmental monitoring, tourism, land surveying, and mapping.

Beneath the public Spatial Data Infrastructure (SDI) approaches, like the Infrastructure for Spatial Information in Europe (INSPIRE), the Data Information Access Services (DIAS nodes) of the Copernicus programme, and several open data portals on each administrative level (Europe, Federal, State, Local), geo information systems being operated within private sector industries, like real estate, insurance, architecture, engineering, construction, transportation, logistics, mobility, smart cities, energy, and utilities, too.

To build and support geo related infrastructures, a worldwide geo spatial industry has emerged over the past 50 years. It is based on sectors like positioning, geo information systems (GIS), spatial analytics, earth observation sensors, and 3D. Due to a study on Copernicus by the European Commission from 2018, the overall market potential for the EMEA region was expected by 5.016 billion USD in 2020 with an average growth rate of 9.33%. Having this in mind, Gaia-X geoinformation domain's mission is to follow a technology open geo based digital tin approach. Intention is to unlock the power of spatial relationships through geo spatial enabling technologies by integration in value adding business processes. The characteristics of Gaia-X, known as trust, governance, data sovereignty, data portability, and openness will lead to additional value within geo data dissemination, provision, integration, and usage.

From a visionary point of view, a "horizontal", and value driven network of interconnected, private, and public geo platforms serve the demands of other "vertical" Gaia-X domains based on Gaia-X principles of value, trust, interoperability, transparency, and openness. The overall benefit and leverage effect emerge through a "geoinformation spillover" into other Gaia-X domains. This "cross domain approach" fosters geo information sharing, creates business

value, enforce crowd sourcing, collaboration, innovation, and makes the application of spatial technologies more economical and efficient.

Challenges addressed

As stated above, the European geo spatial market is a developed market driven by user demands, private cloud infrastructures, public spatial data infrastructures, public open geo data, free and open earth observation data, commercial products, open-source components, and the societal demands to create a sustainable future on the basis sharing and collaboration principles. This leads to more understanding of the natural and built environment to precede smart actions for an increased sustainability.

Following challenges are foreseen:

- to reach out Gaia-X principles into a widely developed geo spatial market
- to address the geospatial market and identify the business value of Gaia-X principles
- to align successful business models of the private industry and public sector developments with Gaia-X architecture, values, and governance principles
- to adapt commercial software products to Gaia-X and vice versa to make specific public information systems (e. g. cadaster, high resolution images) accessible with Gaia-X principles
- to implement use cases without having Gaia-X core services available
- to agree with the domain industries and communities on common governance principles, and implement those
- to consider and develop current relevant standards and initiatives, such as OGC, ISO, GDI-DE and INSPIRE within a modern and future oriented Gaia-X based standardization process
- to develop interoperable data models, which allows the spatial data providers on local, regional, and national level to be integrated into the specific Gaia-X services via standardized spatial data services
- to have billing/pricing principles available to sell commercial data and services into/out of Gaia-X
- to have early available a set of generic Gaia-X services to foster the development of applications and solutions (e.g. administration, industries, Infrastructure as a Service approach)
- to enforce the availability of information as a service, and actionable data/Information/services generation (value adding services to generate ready-to-use products and solutions)
- to implement a clear und transparent lifecycle management (research & development & economy/productization)
- to make high valuable datasets widely accessible
- to have a testbed/sandbox environment available early to build PoC or show case studies for evidence-based standardization/processing or using federated catalogs such as tools, data, or services

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

Partners of the ecosystem

- GIS Software vendors
- Building Information Modeling (BIM) Software vendors
- Private and public geo data providers
- Platform and service providers, PaaS, SaaS providers
- Governmental agencies, public sector; including cadastral and land management information/systems
- Private Sector/Private companies, professional service companies, system integrators
- All levels of Public administration (European, National, State, Local level)
- Citizens (to benefit from open geo data and end-to-end services)
- Researchers, Space administration, SDI initiatives
- Other domains and data spaces
- NGO (disaster and risk management, World Bank/Development Organizations, UN) using data with global dissemination
- International institutions (Development Banks, EEA, WHO, WFP)

Use-cases (scenarios) within the data space and the current state of development

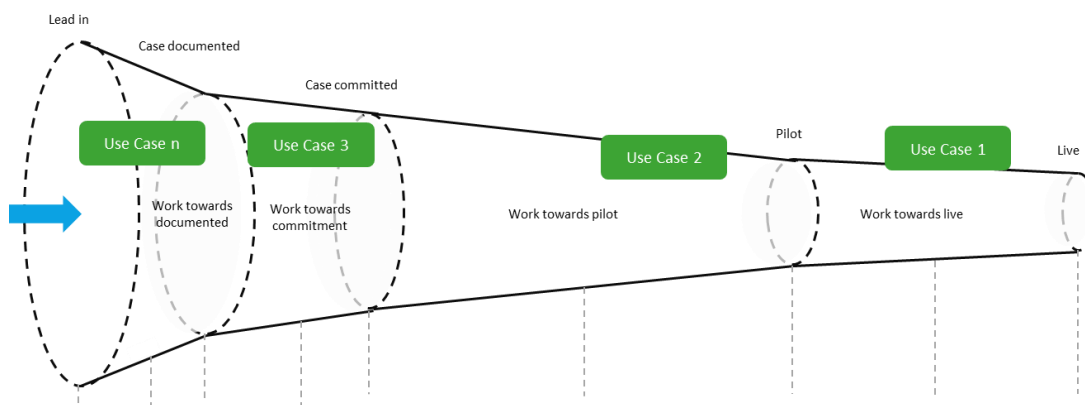


Figure 12

Use cases play an important role in the domain geoinformation. Since the onboarding process of domain members is ongoing, new use cases will appear continuously over time. However, not all of them will succeed as this depends on commitment and voluntary engagement at the beginning. Since business models must be developed, and the reach out to customers/users, too, some funding has to pave the way in the second half of the shown “Use Case Funnel”. Successfully implemented cases will have a “going live” to show and point out the domain geoinformation itself and cross domain benefits. As geoinformation is not abstract data and

can be depicted very vivid and clear on digital and interactive maps, it is an excellent basis for public Gaia-X value communication.

Use Case 1 – Space4Cities

Solution

Space4Cities provides solutions for the joint exploitation of earth observation, machine learning and modern information and communication technologies to provide decision makers from science, economy and policy with precise, up-to-date and ready-to-use information on the highly dynamic built environment – from local to global scale.

Problem solved

Cities are one of the most rapidly evolving systems on Earth. Hence, effective management and development of the built environment requires detailed, timely and holistic knowledge about the status and changes of the urban system. At the same time the digital transformation taking place in all areas of life has led to a massive increase of digital data – in particular, related to the places where and the ways how we live. It is here, where the enabling technologies developed in Space4Cities allow users to effectively access, process and analyse the petabytes of available big city data in order to gain innovative knowledge and insights that help making the human habitat more sustainable, livable and resilient. At the same time mechanisms for the integration of citizen sciences, crowdsourcing and open knowledge foster the inclusive, creative, and transparent character of the applications and services of Space4Cities.

Partners/Ecosystem

The Space4Cities activity includes and serves a broad spectrum of partners and users from science, urban and environmental planning, economy, and policy. A particular focus lies on the active involvement of international institutions (e.g., United Nations, World Bank, Asian Development Bank) in order to substantially support the Sustainable Development Goals, contribute to the monitoring of global change, and improve disaster and risk management.

Main technology/Gaia-X components

The main technological components of Space4Cities and related Gaia-X instances include instruments for big data collection, structuring and processing, earth observation, and smart data analytics. The related technical solutions cover the automated information extraction from large-volume and heterogeneous data sets (big earth data) by means of machine and deep learning, empowered by distributed computing infrastructures for high performance processing. These approaches are supplemented by modern information and communication technologies, which allow the autonomous control of data access, management, and processing (automation). Key components of the Space4Cities framework have successfully been applied in the global urban monitoring initiatives of the Global Urban Footprint (GUF) and the World Settlement Footprint (WSF), whose information products are already used by over 600 renowned institutions worldwide.

Concrete benefits

The value of the Space4Cities solutions lies in the synergistic evaluation of databases of earth observation, open data, and public administration to break down data silos and offer - GDPR-compliant -tailored information products for city planning and development and for new digital products and business models (e.g. in the area of sharing economy or local public transport). Ultimately, Space4Cities is supposed to help improving the quality of life for citizens and save costs by making processes and decision more effective and sustainable. Therefore, Space4Cities finally supports cities in stepping toward a digital metropolis that is fit for the future.

Use Case 2 – Smart Infrastructure Management

Solution

Gaia-X enables the early engagement of organization as well as the fusion of various data sources in a trustful data sharing ecosystem towards one information base in order to plan, execute and run complex Architectural, Engineering or Construction (AEC) projects in an interdisciplinary, collaborative, and timely manner. Additionally, Gaia-X enables the provision of a crowd participation service for the affected citizens to ensure early community involvement.

Problem solved

Planning and approval procedures have reached a very high degree of complexity in the last two centuries, making it very difficult for decision-makers in planning offices, authorities, associations, and courts to master these procedures. Moreover, construction planning needs to consider a large number of interests and stakeholders, including for example property owners, transport ministries, land surveying offices, environmental protection agencies and the general public.

Currently, the various stakeholders are engaged successively. This is partly due to the different restrictions in place regarding the purpose and use of data, as well as questions of competence and partly due to isolated approaches that make it impossible to merge relevant data. At the moment, there is no uniform platform for collaboration and the secure provision of data with all the necessary functionalities and rights to access and use data. As a result, processes are mostly limited to simple data provision, leaving the great potential of real-time cooperation untapped, including advantages in terms of flexible and targeted data use. In addition, objections from communities and associations often go unheeded in the planning phase, leading to later changes and additional project costs for the planners.

Partners/Ecosystem

The use case combines several stakeholders such as road construction authorities, planning companies, responsible developers for instance bridge structures, Ministry of Transport, Environmental Protection organizations, surveying offices, affected municipalities and citizens, providers of critical infrastructure (e.g. energy suppliers), SMEs and the general public.

Main technology/Gaia-X components

This use case impressively demonstrates that an ecosystem of central services based on the Gaia-X Cloud can be established to enable secure and user-friendly operation for relevant use cases. The use case shows that the secure provision of data always includes the necessary functionality to process this data in a relevant business process. A common ecosystem of users and providers is created. Therefore, component such as Catalogue, Identity Access Management, and control of digital identities are crucial.

Concrete benefits

The ‘Smart Infrastructure Management’ use case lays the foundation for better interdisciplinary and user-friendly cooperation in planning beyond the boundaries of organizations and authorities – while ensuring their self-determination and sovereignty. According to modularity and interoperability, Gaia-X ensures the continuous availability of various data sources as well as integration of additional existing solutions or data bases. Based on the digital data infrastructure Gaia-X, which ensures data exchange across national boundaries, this use case accelerates the procedures and quality of project planning – and therefore contributes to the European added value.

Use Case 3D Planning Platform

Solution

The use case describes a solution for using a digital twin of a city as the backbone for planning projects and processes. Projects can be visualized, observed, and checked throughout their entire life cycle. Planned developments can be examined within a 3D model under the existing real conditions in the respective city. It is also possible to streamline the process of creating and sharing 3D zoning and land use plans. The use case enables citizens, real estate developers and other community stakeholders to participate in the planning and development process.

Problem solved

The planning, information and participation process is digitized and made more efficient thanks to a common 3D planning platform, the Digital Twin. Existing 3D context data from the surveying offices form the basis of the planning platform and can be used for this and many other applications, resulting in benefits for city and planning offices, regional planning authorities and the population. Cities and states get easy access to the relevant data and software to guarantee a sustainable and citizen focused development.

The planning and development processes are open for various stakeholders (government, public, research, AEC, real estate). In this way, the community can be asked for feedback at an early stage. Stakeholders can use the Smart Urban Planning platform to access a web-based view of all plans and projects if they are publicly approved.

Partners/Ecosystem

The use case allows various stakeholders (government, public, research, AEC, real estate) to participate in planning and development processes via a digital twin of a city.

Main technology/Gaia-X components

Gaia-X provides the necessary infrastructure for the integration of the Digital Twin and its data space, ensures secure data exchange, and allows data to be made available without media disruption. The solution is offered as SaaS and considers the pursuit of modularity and interoperability. Specific software (e.g. ArcGIS Enterprise) is run on Gaia-X for access to the 3D models or the digital twin services. Gaia-X is also used for the hosting of the 3D services.

Concrete benefits

Gaia-X provides a fast and secure environment to host 3D services and to run user specific (planning) software. Gaia-X offers stakeholders like cities or states a user-friendly and easy access to the relevant data and software. According to the concerns about political data protection regulations and preferences Gaia-X provides the planning processes based on legal foundations such as land use and development plans. The calculation and comparison of future scenarios and their impact on capacity indicators (jobs, energy consumption, etc.) is a time-consuming task and requires considerable expertise. The Digital Twin serves not only for planning purposes but for many processes in smart cities (e.g., AI, IoT, green infrastructure).

Use Case Automatic 3D Spatial Content Generation

Solution

The use case describes the automation of processes to create and update 3D surfaces and features in high detail and accuracy with methods of artificial intelligence.

Problem solved

The concepts and technologies of Open Data Science are utilized to generate or update 3D surfaces and features from geospatial incoming raw data automatically. Additionally, market available software products are used to prepare the training data and to handle the incoming geospatial raw data. It exists a trained algorithm of artificial intelligence to do the job based on various incoming data sources, e.g., aerial imagery or LIDAR point clouds.

Partners/Ecosystem

The use case is aimed at companies, SME, research institutes. Potential end users are citizens, Industry 4.0, smart cities, agencies, security organizations - for planning and operational purposes.

Main technology/Gaia-X components

Along the chain of finishing the training of algorithm of artificial intelligence, a lot of cloud resources are essential. Therefore, Gaia-X will put the safe usage of artificial intelligence over the edge and provides a certified environment of trust. The intellectual property of the labeled training data and the final algorithm must be protected. Gaia-X will enable the protection. In addition to authenticity and trust, components such identity access management are crucial.

Concrete benefits

Gaia-X enables reducing costs of creating and updating 3D content and improves the frequency of such data updates. In addition, it is possible to save worthwhile human resources to manage the information instead of having to produce it manually.

Use Case Earth Observation Federate Data Access

Solution

Earth Observation (EO) data federated access system is the solution for any stakeholder who wants to contribute or to use EO satellite data in complementary and direct way, online and secure from one comprehensive source. Proposed cloud-based system will be able to efficiently deliver data to Destination Earth tasks. Our system could also be a perfect solution for not only satellite but aerial, health image diagnostics, IoT or any other big data challenge and partners.

Problem Solved

Problem definition:

- Instant, online, and immediate accessibility of huge amount of EO data.
- Availability and easy use of enormous data sets – Copernicus and other EO data may be augmented with some in-situ data. Data sets of the size 50PB or more – a challenge, too.
- Online, immediate access to long time and space series of data for any cloud or HPC infrastructure in Europe
- Data easily discovered – for interactive and m2m applications
- Data tagged for potential AI applications

Rapidly growing amount of collected Earth Observation data should be provided to the users in predictable and reliable manner. The vast amount of public data is generated in an open access Copernicus - European Earth Observation programme. Nature of the data sources ecosystem consist of growing time series of freely available satellite products (e.g. Sentinel), increasing number of data-based services (e.g. Copernicus Services) and of the commercially available VHR data. Successful use of this wealth of information will depend on the possibility to access them in a fast and reliable way. It is as well to be expected that the number of the sources (data repositories) to be consulted in each project will grow with time even if some data sources will federate. Similarly, the size of individual repositories will increase. Already full Sentinel Data repository takes well over 30 PBytes of data with daily growth exceeding 25 TBytes per day and new missions are planned. It is thus necessary to design robust, open source solution standards allowing data providers (data repositories) and data users to 'meet' and to process the data in cloud environment. It is a challenge itself to manage 50 PBytes or even more of heterogenous data (EO, in-situ and statistic) nevertheless this use case/solution can solve this enormous challenge.

EO data federated access solution includes and serves a broad spectrum of partners and users from commercial, scientific, and administrative worlds.

We propose to build cloud-based federated data access service using API and interoperability standards to provide access to a very large amount of EO data from various sources while keeping data traceability and accounting but protecting user rights. In particular, developing and testing definition semantic metadata standards for AI data labelling are needed in order

to include AI-ready data stores. Project will include specialized tools for data providers to make their collections discoverable.

Partners/Ecosystem

Partner ecosystem would consist of data and information providers, cloud operators, data users and information end users including VHR data providers, specialized associates for data processing algorithms development, science and research institutions, federation partners. Since Earth Observation and Services usually address specific challenges (geohazard, water or coast monitoring) the project will involve already existing ecosystems focused on their specific area of interest.

Main technology/Gaia-X components

The project would build on the already existing European solutions such as:

- Open source based public clouds
- Open source based Big Data storage
- Existing Open data access standards
- Big Earth Observation data repositories (e.g. CREODIAS with over 21 PBytes of the data, WEKEO, other DIAS-es), other data and information sources
- PaaS solutions – data discovery, on-demand processing
- AI tagging – open source plus necessary developments

Challenge – scalability of storage, efficiency of data access (more than 2 PB daily delivered), federation of existing data sources and development of PaaS elements (in terms of new functions and increased efficiency). Offered data will be consolidated, indexed, tagged, and prepared for potential pre-processing, making the service adjustable to any need. The main Gaia-X components to be used are the Federated service catalog, Identity management and Standards of interoperability. The solution while secure and robust should be flexible and open for evolution.

Concrete benefits

Federated, stable, comprehensive, efficient, and easy access to Earth Observation data for instant use and/or further processing for every customer, maximizing the use of available data on European platforms are the main benefits. Fully controlled European solution with European competencies and infrastructure, located and managed only in Europe and by Europeans harmonizes and constitutes with the main assumptions of Gaia-X.

Added value of the proposed system comes from the limiting carbon footprint by de-duplication of data and keeping it instead of repetitive generation on-demand. In this way maximized use of existing resources will be achieved (no need to copy or reprocess data to make them usable). Elasticity and multitasking - possibility to extend the model to other fields (like telemedicine data) is another example of potential positive impact of our System on environment as well as a community and business.

Thanks to this project Earth Observation data will become a part of an open, transparent digital ecosystem, where data and services can be made available, collated, and shared in an environment of trust.

Maturity indication of the data space, health status

How is the demand side represented?

- by other domains data spaces
- through domain members from public and private institutions
- by customers of private companies
- by representatives from technical and scientific organizations
- European citizens, participation processes, transparency of political decisions if spatial relevant
- through shared knowledge generation as basis for innovation, business
- by decision makers, politicians, and delegates from crisis management centers

How is the supply side represented?

- Software vendors (e.g., GIS, BIM)
- Platform operators and providers
- Public geo data agencies
- Private geo data providers
- Developers
- Startups
- Crowd sourcing communities (business perspective)

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

- Yes
- Open business models; usage tracking; buy versus use on demand
- Enhancement of business models

Is the story of the data space well documented?

Data space members didn't agree on a common story yet. However, the fundamentals of a storyline are outlined in section 1. As every information is somehow related to space, the geo data space is a common data space that enables business value in other data spaces through spatial referencing by a network of interoperable platforms, data sharing, and collaboration. As most of today's challenges towards a sustainable future needs several data from different industries (big data paradigm), the geo data space provides a common and harmonized spatial reference. On one hand side application neutral data and services is provided (maps, terrain, land use, 3D city models, satellite images etc.), and on the other hand, several spatial methods achieve spatial intelligence to lots of domain specific data. In other words, the geo data space

is beneath Gaia-X itself another common nominator of Gaia-X domains to foster efficiency and innovation.

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

Gaia-X provides federated cloud architecture components on an open basis, and common governance principles. Following the rules, Gaia-X geospatial platforms are interconnected and serve as a “platform of platforms” (geo based digital twin approach) other Gaia-X domains with geo spatial data, information, relationships, and location intelligence. In terms of a business model the geo data space creates value through spatial relations and location intelligence to other domains without having extensive geo knowledge by themselves. Interconnected identity management, open API and interfaces, and a system of systems lead to better insights into data, and smart intelligent decisions.

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

It is worth to mention, that the use of data sets, information layers and intelligent geo spatial functionality depends on the availability of geospatial technology and geo data. The Gaia-X federation services will lead to certified interfaces and APIs to feed data and services into a decentralized, federated and independent catalog that is maintained by many involved actors of Gaia-X. Together with a common and cross domain identity management, and standards of interoperability, this achieves the basis for data sharing, collaboration, billing, and pricing.

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

As geoinformation is a fully developed industry, the adoption to Gaia-X depends on additional value created by Gaia-X. This business value might result from Gaia-X characteristics as far as others did not cover them with their own business models. To scale up, value, benefits and incentives might be necessary to involve software vendors, data, and IT infrastructure providers to adopt to Gaia-X and roll out products with Gaia-X compliant interfaces and APIs.

How can the commitment of the parties involved be proven?

The commitment of involved parties depends highly on a successful governance of Gaia-X. If Gaia-X services and features are required through procurement processes within public sector, public administration and public owned companies, the commitment could be checked by existing Gaia-X compliant IT products. If there is no further “real” market pull, Gaia-X would just be nice to have, and not a must.

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

To awake the spirit of Gaia-X by resources for further implementation, innovation and business value must go hand in hand to tackle the markets. Resources of companies will be available if differences to existing solutions and offerings would bring additional business value. This also depends on further funding and tenders.

Evolution of the data space

Roadmap of the evolution

- Onboarding of platform providers and software vendors (GIS, BIM etc.)
- Interconnect with Gaia-X federated services (catalog, identity)
- Agreement to common governance rules
- Implementation of use cases
- Marketing and business development

Quick wins (for 2021)

- Find geoinformation services by a federated Gaia-X catalog service
- PoC to access an identity controlled geoinformation service by “n” other domain/data space
- New innovative capabilities to produce, update and share high quality 3D x/y/z and spatiotemporal x/y/t and x/y/z/t geoinformation in a controllable manner
- Interdisciplinary fusion of relevant geoinformation towards a more realistic digital twin as a basis for BIM or other trustful interconnected processes

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

- Running a data hub on basis of federated platforms and identity access management
- Improvement of usability for open data proposals
- Deeper integration of geoinformation and appropriate functionality to manage them into the valued adding process chains of the ecosystem
- Accessibility of non-profit and profit geoinformation for training of AI algorithm in a trust-fair environment
- Availability of a deep, high quality geoinformation database to discover and research phenomena of climate change even better
- Improvement of environmental and biodiversity protection by a better and interdisciplinary accessible data base of geoinformation of high scale
- Connecting with relevant European Spatial Data Infrastructure activities, such as INSPIRE, EarthServer, and DIAS nodes (on European and national level),

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

- Gaia-X acting as a living marketplace of data services
- Access to Gaia-X based storage and computing capabilities
- Deployment of geoinformation out of Gaia-X connected platforms as services

- Effective integration of geoinformation into all value-adding processes to achieve a better exploitation of the value of geoinformation by trust-share and digital business models

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

- Align public geo spatial data infrastructure with Gaia-X
- Provide geoinformation out of Gaia-X
- Implement technology open value chains for successful market impact
- Provide tools and methods to combine, analyse, visualize geo-based information
- Follow on open data initiatives
- Simplify existing initiative in a more user focused manner (usability)
- Implement government and platform interconnections to exploit the value of existing geoinformation as well as to improve a high-quality basis of accurate and current geoinformation
- Enable trustful-sharing and real-time collaboration for and with geoinformation
- Enrich 4D digital twin initiatives with geoinformation
- Foster relevant standardization process towards open standards in the domain of geoinformation
- Foster business opportunities on all levels, including Public-Private Partnerships

Smart Living

Mission and Goals of the data space

In the future Smart Living technologies and applications will become an inherent part of our lives. The "Smart Living" ecosystem is developing rapidly (Figure 13). In Europe 250 million residential buildings have a massive potential for Smart Home/Smart Building technologies and cross-domain applications/services. Consequently, the global Smart Living market is expected to raise above 22% (Figure 14). This development's significant driving factor is artificial intelligence (AI) and its potential for highly individualized, context-aware, and cross-domain applications and more intelligent systems and devices.

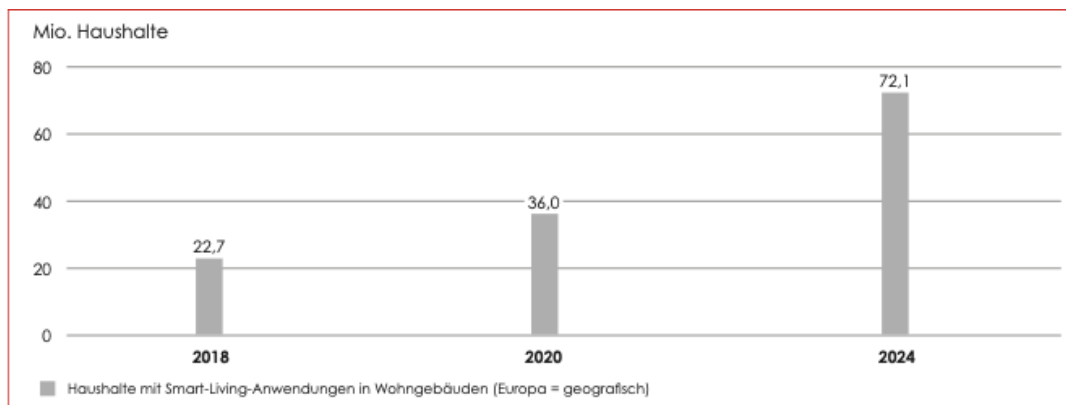


Figure 13: Number of Smart Living applications in residential houses in Europe (Source: SmartLiving2Market 2020, page 15)

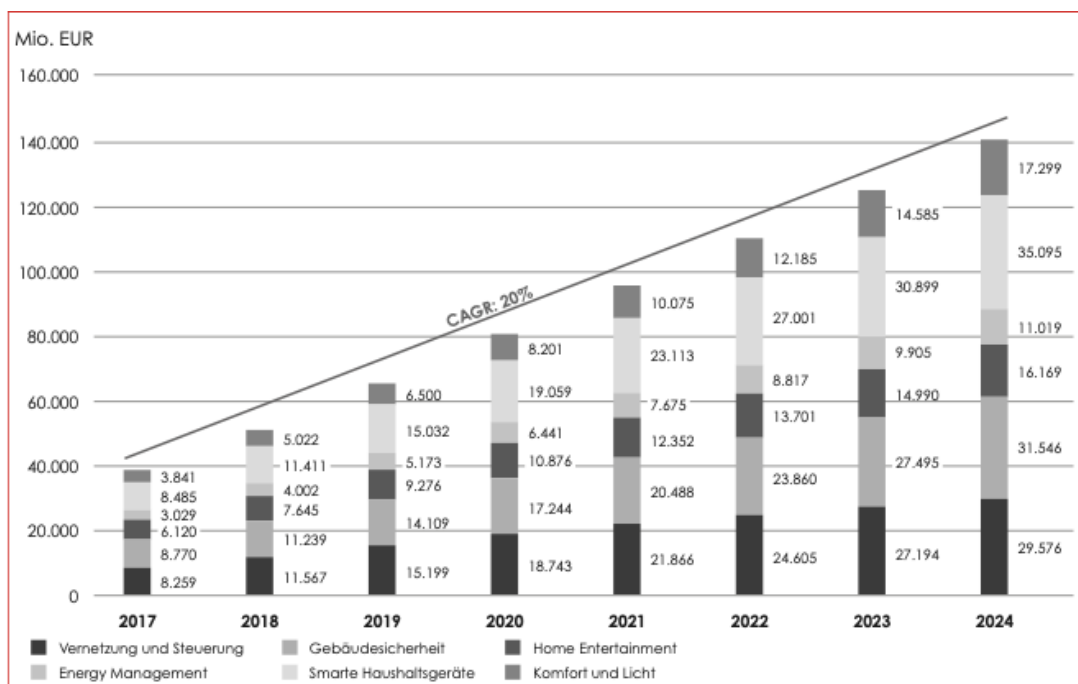


Figure 14: Prediction of global market development of Smart Living and Smart Living applications (Source: SmartLiving2Market 2020, page 13)

Given the fact of intensive connections to adjacent domains, such as smart energy networks, smart city, assistance and AAL, etc. (Figure 15) the overall concept of Smart Living is “Building as an intelligent Service”. A smart home, a smart building or a group of smart buildings should be able to interact more or less intelligently with smart energy networks, smart city structures, smart mobility offerings and smart assistance services. Information gained from smart buildings are useful and even essential for all these domains. The mission and goal of a Gaia-X Smart Living Shared Data Space is to enable such cross-domain services and applications.

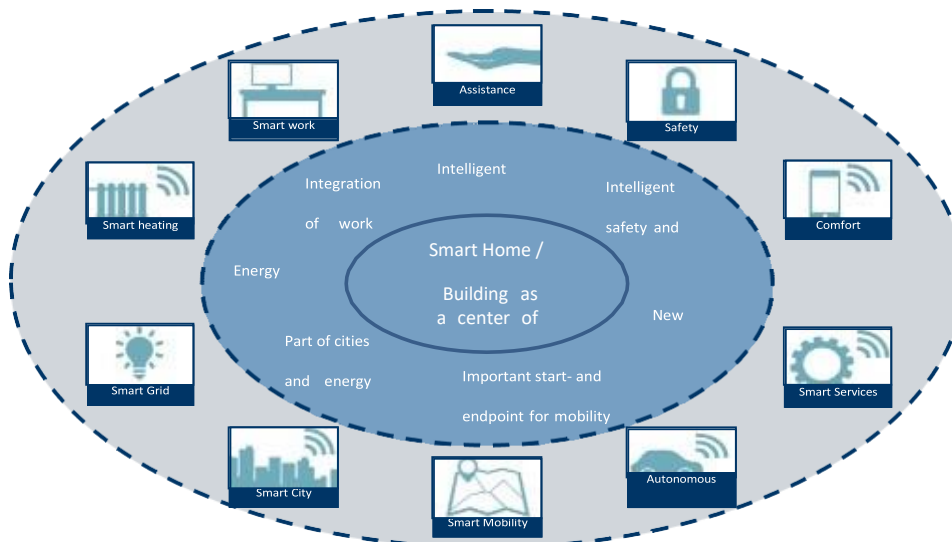


Figure 15: Smart Living and adjacent, connected domains (ForeSight Community)

Service development should be simplified drastically to meet future (and urgent) requirements, such as seamless integration of (smart) buildings in smart energy networks for an improved energy efficiency or to make smart buildings ready for urgently required assistance at home.

The potentially massive Smart Homes and Buildings scalability and federation of distributed clouds and edge devices are considered a vital issue for further development. Reasonable standards to describe clouds, resources, and data provided by specific clouds, useful semantic search functionalities, federated and maybe domain-specific resource catalogues, and an integrated cross-resources access management for improved data sovereignty are key factors for the future Smart Living applications. Consequently, a Smart Living Shared Data Space is a next logical step towards a virtualization of data locations, data access as well as computation, AI, and storage resources. Service developers need to concentrate on development of intelligent services and not on handling numerous clouds, searching highly distributed data, implementing numerous interfaces to different clouds and services and access mechanisms.

To realize these goals, Gaia-X offers the Smart Living domain and the housing industry an improved, easy, and secure access to a multifunctional cloud environment in the GDPR space. Gaia-X will provide required technologies to connect regional and functionally specialized data centres and thus supports the task-specific scaling of smart-living applications that are dedicated to edge computing due to high latency requirements. Gaia-X facilitates suitable

standardization requirements for linking the growing volumes of data and data sources and thus promoting the emergence of further AI-applications, mainly through cooperation between the digital economy, housing industry, and electrical industry.

Challenges addressed

Number of stakeholders: Before a building gets smart, several stakeholders must work together. Real estate industry plans (Building Information Modelling - BIM), builds, operates, maintains and markets a building, manufactures (electrical industry) develop, configure and market innovative equipment and components and tech industry develops and connects smart services using those installations (Figure 16). Today, Smart Living still comprises completely separate trades that will/must in the future interact in cross- domain applications. These applications require intelligent, situation-adaptive buildings that fit seamlessly into comprehensive structures.

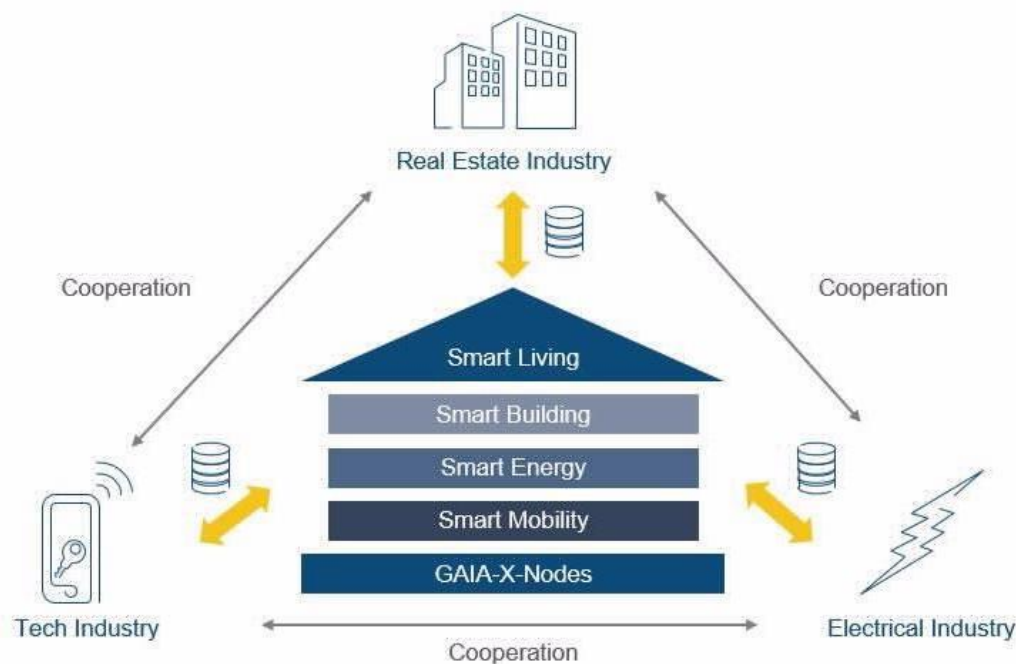


Figure 16: Stakeholders of Smart Living

Heterogenous ecosystem: A prerequisite for such, in many cases AI-based services, is consistent data acquisition, processing and networking. The Smart Living ecosystem consists of numerous (69,000 housing companies with 53,2 billion turnovers in Europe) mainly small and medium companies. Within this heterogenous structure service developers are confronted with many different, in most cases incompatible smart home system families and devices, highly individual installations, more and more widely distributed cloud representations of components and devices with incompatible interfaces as well as with many different stakeholders owning data and providing access to required data. Development of useful and intelligent Smart Living applications requires a coordination of numerous domains,

integration of separated, incompatible devices, implementation of methods and interfaces for individual data handling, an adoption on different markets, system families and different standards, and of different scales (Figure 17).

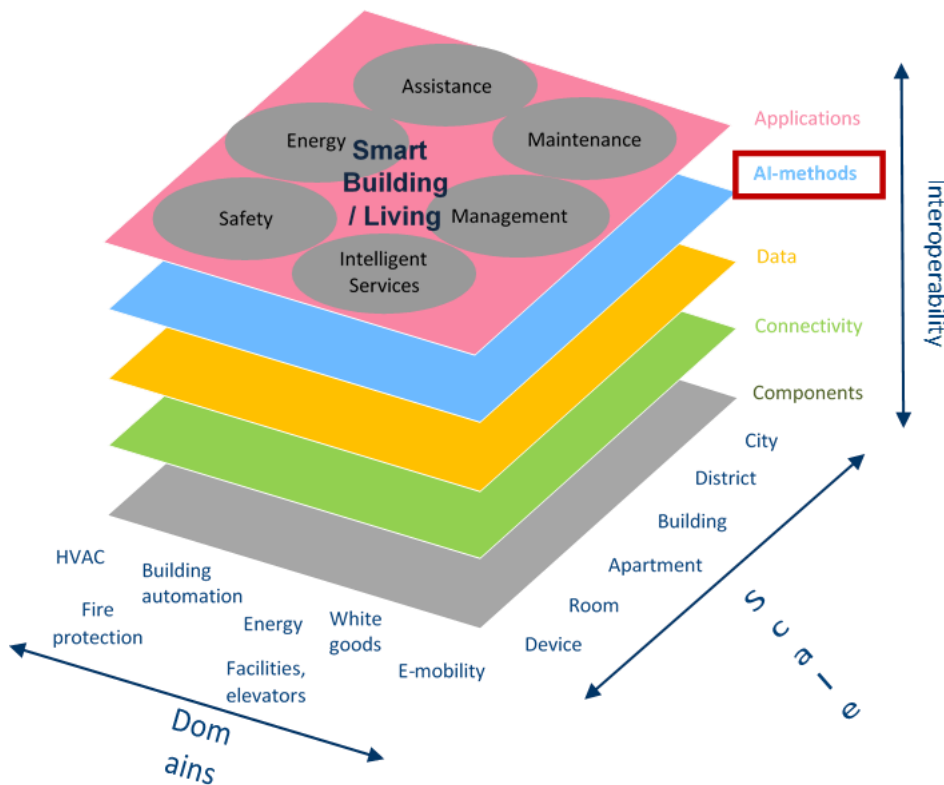


Figure 17: Smart Living data model (ForeSight Community)

Scalability and semantics: Due to heterogeneous structure a consistent semantic description of data and resources as well as more or less publicly available, federated resource catalogues with intelligent search functionalities. There is a need for a secure, scalable, and high-performance cloud environment that also includes local edge devices.

Data protection: Smart Living data is often (strictly) personal and sometimes health-related, i.e. highly sensitive, especially because data of third parties like guests is considered as well. This includes, for example, the recognition of activities, patterns of movements, usage habits, presence, and emergencies. Other scenarios, such as predictive maintenance for building technology or smart energy management, are less sensitive or not sensitive at all. Gaia-X must offer configuration options for these scenarios.

Transparency: Companies in the housing industry want to offer intelligent services and automation solutions and in return store and process customer data flexibly, transparently, and locally in Europe (DSGVO room). Transparency and easy to use settings where data is located and what can be done with it are required. An increasingly important factor is to clarify of what AI is allowed to do with data – and what is off limits!

Integration of data and functionalities: Smart Living applications have to integrate data and functionalities on several levels (Figure 17). Data from smart homes must be merged with data from the building and its technical installations for numerous applications. Automated buildings should be monitored as efficiently as possible for maintenance requirements and technical problems. The combination of occupant-specific smart home installations, intelligent household appliances, etc. in conjunction with their specific clouds and other manufacturer-specific clouds for building technologies with data from the tenant management demands a sensibly structured, secure, and reliably available data management.

Performance: Smart Living applications in a broader sense have very different requirements on performance (some applications need reactions in hard real-time, some others are fine with more extended periods), amount of data, scalability and level of data protection and IT-security. Besides high availability requirements up to 100% up-time for critical smart-living services such as emergency detection and fire protection are required in building services engineering.

Interoperability: As described above, Smart Buildings and Smart Homes consist of different system families and smart devices. So called vendor clouds provide users (and developers) complex digital twins of components and devices. Most vendors offer cloud-based, proprietary JSON-interfaces to interact with data and functionalities. For service developers the challenge is that each cloud and digital twin has specific interfaces and for the vendors the challenge is the high effort to transfer a given, complex data space to a different cloud provider, so that a virtualization layer based on Gaia-X might improve interoperability on different levels.

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

The use of open standards in the implementation of Smart Living Data Spaces will be a key success factor on the way to a Gaia-X compliant architecture and a wide acceptance. In this way, the ecosystem achieves global interoperability, scalability, and resilience while avoiding vendor lock-ins. These characteristics, in turn, drive innovation and enable users to freely choose services.

Cloud infrastructures: For cloud infrastructures and the service implementations based on them, this applies. The choice of the right provider, the implementation of standard solutions and a community-led strategy are success factors here. Already today, many data centres use a mix of different cloud services from different providers. This makes multi-cloud management necessary, which pursues two main goals: Users from the business departments should be able to obtain the cloud services they need as far as possible independently in self-service, while from the IT point of view, automated billing takes place, among other things. Industrial companies are developing their own cloud and IoT enablement platforms based on the major hyperscalers or are increasingly becoming providers themselves. This is accompanied by several cloud management tasks that contribute to a functioning overall system: Infrastructure Management, Application Management, Service Management, Managed Services, Full Managed Services. Interoperable tools for common tasks including deployment, autonomic scaling, and monitoring are extremely helpful. The focus here is on integration, portability, interoperability, and innovation.

Various tools have already been established in the implementation community, some of which have also been used in the use cases described above. In the foreseeable future, containerization will be the tool of choice for scalable and resource-efficient provision of cloud services. The major cloud providers rely on their own platform solutions for setting up and managing the services. However, there are different cloud computing platforms based on open standards for the deployment of public and private clouds, such as OpenStack. For service management, OpenShift is probably one of the most powerful tools on the market right now, offering many of the functions described above integrated. For orchestration, Kubernetes is often used, which as an open-source-platform achieves excellent results in the management of workloads and services in the containers. The overall goal must be to simplify any transfer of existing cloud and data spaces and established cloud-based applications. Finally, event streaming platforms will play a major role in the architecture of data rooms. Here, a growing community has already successfully deployed Kafka Server. The open-source platform is already used by a variety of companies for high-performance data pipelines, streaming analytics, data integration, and mission-critical applications. In the project ForeSight we are establishing Kafka to evaluate its capabilities for the Smart Living domain.

Standards and virtualization: As described above, a Smart Living Shared Data Spaces should be a virtualization layer on top of existing cloud and edge environments. In order to gain a wide acceptance, it should be based on international standards – in the best case W3C and well-accepted industrial standards. A data space should offer data providers and service developers:

- Standards for semantic description of cloud resources, cloud functionalities (storage, high security, AI, etc.)
- Standards for semantic description of resources hosted in a cloud or on an edge device
- Standards for semantic descriptions of data, services, and things
- A set of basic ontologies for typical smart living and connected domains e.g. iotschema.org, SAREF, SAREF4Ener, GTFS, etc.
- A comprehensive set of tools for the required mapping between different data models, and ontologies
- A comprehensive set of tools to perform automatic semantic uplifting of data and interfaces that are not yet semantically described. Those tools should be predefined for a standard set of data models of the Smart Living domain
- Catalogues for those data, services and things with a comprehensive search functionality and a fine-grained access control
- Concepts to setup and connect federated catalogues as well as to enable cross catalogue semantic searches and reasoning
- Concepts to describe data origin, data quality, quality of service, as well as limitations for an AI based data processing
- Easy and cross cloud deployment of containerized services and environments
- Easy to use, predefined AI environments for machine learning and rule-based AI
- Tools for data analytics beyond AI

Integration of IoT: Smart Living is an umbrella of many sub-domains in the Internet of Things (IoT) and IoT is a collection of advanced technical tools that only fully develop their potential in such complex and heterogeneous scenarios. Consequently, IoT platforms should be made Gaia-X compliant - and Gaia-X should support this. Figure 18 shows in principle how an open IoT platform can be used even more flexibly as a federated IoT platform, e.g., by splitting it among different Gaia-X-compliant clouds or by IoT-as-a-Service offerings described via federated catalogues.

Some IoT platforms already provide semantic descriptions, in this example accessible via a graph API. Other approaches include triple stores and SPARQL instead. These are massive steps to support (symbolic) AI and simplify service developments in the context of high data sets/device volumes. The integration of such IoT platforms into Gaia-X should provide a common set of mappings between different semantics and typical data models for further simplification and thus scalability opportunities.

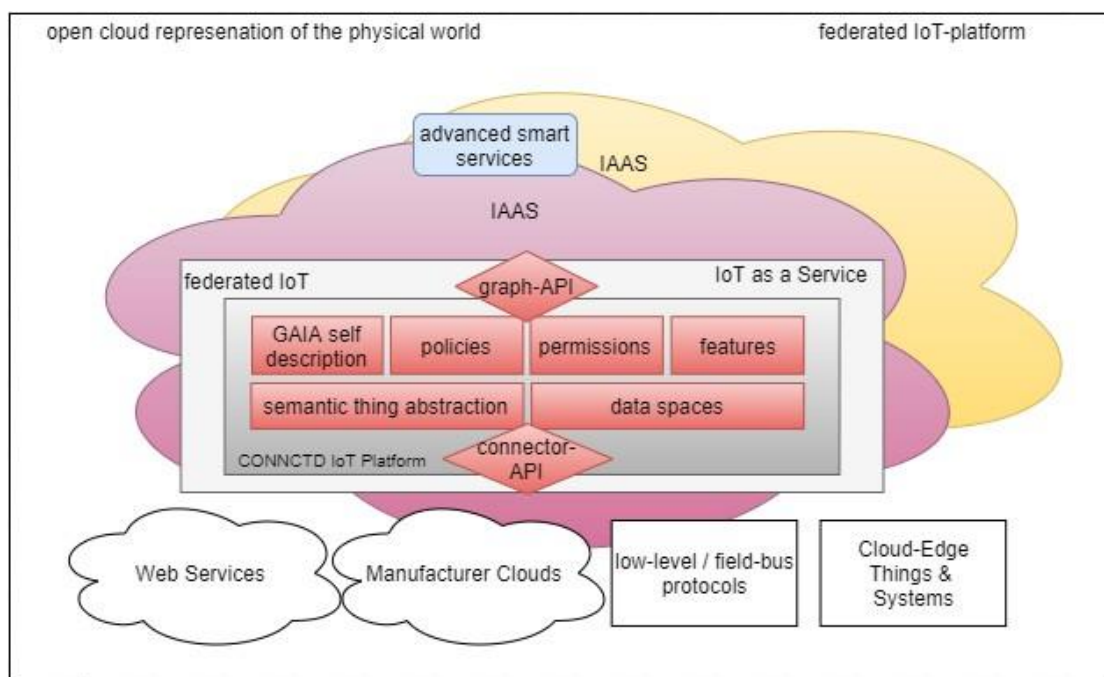


Figure 18: The concept for a federated IoT platform (IoT connctd GmbH 2021)

ForeSight Smart Living Shared AI Data Space, digital twin(s): Figure 19 shows the existing prototype of a (partially) Gaia-X compliant shared data space. The so-called ForeSight AI platform has been developed in the Federal Ministry for Economic Affairs and Energy (BMWi) funded project ForeSight and is a Smart Living AI Data Space. It comprises a federated catalogue (registry), semantic things descriptions, digital twins of connected appartements/buildings, a SPARQL end point, a reasoner, a multi-agent system for data collection and dynamic removal and injection of new things and services in the registry, a concept for virtual sensors, a tool set and predefined AI cloud environment for machine learning as well as a set of essential AI services for Smart Living applications. All services are semantically described and stored in the registry. For data resources that are not yet semantically described a set of mapping tools/services has been developed to uplift those data and interfaces on a semantic level. ForeSight uses Web of Things to semantically describe

“things” on an abstract but machine-readable level. The base IoT ontology of services inside that data space is iotschema.org. This data space can be seen as a first prototype for a Smart Living Shared AI Data Space.

Identity and access management (IAM): All services, as well as all digital twins, are protected with IAM- system “Keycloak” as suggested by the Gaia-X consortium. Keycloak offers various realms, roles, and groups, which in our case provide different scopes of rights, for example, to give tenants, guests, patients, doctors, and relatives appropriate access options to the data they need. A system such as Keycloak allows extremely fine-grained adjustment of access rights, which protects individual data endpoints based on roles and at the same time on attributes, taking into account other factors such as the time of access and highly secure identification and authentication to address the privacy and security requirements in Europe.

GDPR: Gaia-X-compatible Smart Living systems need to be GDPR compliant and should lead by example when it comes to data acquisition and data management strategies:

- Acquire a minimum of necessary data to fulfil the purpose of the systems service, e. g. make distinctions of necessary location levels (miles for weather related purposes, meters for positioning-tasks)
- Enable a user to set such data access rules easily, i.e. by offering a similar approach like currently in cell phones, which will be extended to address all relevant Gaia-X-specific challenges.
- Forbid unnoticed data combination and offer built-in segmentation strategies to achieve this goal.

Report a maximum of real data access related to user’s data, e. g. offer a dashboard for each user.

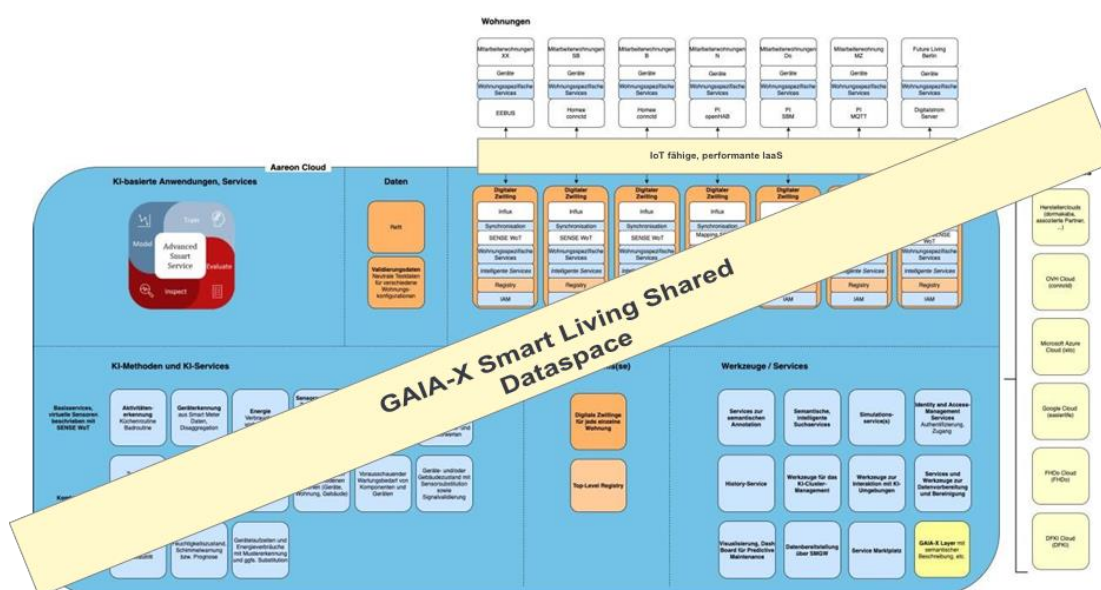


Figure 19: ForeSight Shared Data Space (ForeSight Community)

AI: From a user’s perspective, there are several points to consider when AI technologies are used. A classical approach that is usually employed nowadays is to transfer data to a cloud

system with virtually infinite computing and storage resources. However, this approach has a couple of disadvantages. First, with growing numbers of sensors and increasing sensor capabilities/resolutions, the utilized bandwidth is as well. Second, many people would like to avoid transferring highly sensitive data that is inevitably created in their own homes to cloud systems owned by companies or governments they do not trust, and which are not under their control for privacy and data protection reasons. Another important aspect in the energy domain is that a system should react instantaneously to specific events to provide a high level of safety. This means that AI systems should be able to provide decisions in real time, i.e., in deterministically predictable amount of time.

Edge, Edge AI: One possible approach to increase the confidence of users, minimize amount of data to transfer and storage and provide real-time capable AI functionality is to process the data locally, i.e., where it is generated. In this case, AI methods run on computing devices that are located in or close to the apartment, home or building of the corresponding owner or tenant and transfer only specific data to the external cloud system, e.g., data that describes the occurrence of important events as predicted by the AI methods or significant changes of the state of the environment. This concept is called Edge Computing or Edge AI for the specific case of AI computations.

For Edge AI it is usually preferable to use power efficient and small embedded computing devices. To provide the computing performance required for AI methods, application specific hardware accelerators that are targeted towards AI computations can be used. This has several reasons. First, for many applications it is desired that the computing hardware is “invisible”, i.e., deeply embedded into the application and not recognizable as such by a user. Second, from an energy efficiency perspective, the AI methods should need as little energy as necessary, since the number of computing devices used for the Edge-AI concept might be very high and the devices are continuously running.

Partners of the ecosystem

As described above our ecosystem consists of many different stakeholders or sub ecosystems that are in many cases not or not well connected. Stakeholders are amongst others:

- Residential companies as owner and/or managers of buildings
- Residents
- Manufactures/vendors of smart home components, devices, systems
- Manufacturers/vendors of intelligent wide goods and kitchenware
- Manufactures of smart elevators, heating, cooling, photovoltaic, battery equipment
- Manufacturers and/or operators of internal energy management systems
- Manufacturers and/or operators of internal access control systems
- Manufacturers and/or operators of internal fire protection systems
- Operators of charging stations for e-cars
- Operator of surrounding energy network
- Services providers
- Maintenance providers

- Care provider
- Financial industry
- Municipalities
- Waste management companies
- Utilities (gas, water)

Consequently, the existing ForeSight community comprises currently 70 partners.

Use-cases (scenarios) within the data space and the current state of development

Smart Living addresses two societal highly relevant topics, namely energy efficiency and the ageing society in many European countries. Additionally, to these comprising and extremely challenging use cases the Gaia-X Smart Living working group currently develops a set of minor use cases tackling single aspects of a Shared Data Space.

An important attraction and extension of the existing Smart Living Shared Data Space is AI. Nearly all use cases that partners describe are somehow related to AI in general. Because all hyperscalers offer at least more or less ready-to-use machine learning environments, partners expect such functionality plus specific AI-based essential services that simplify application development remarkably. Thus, ForeSight aims to develop a valuable set of such basic services for different purposes and to develop an advanced virtualization layer for a dedicated AI-cloud environment.

Studies have shown that in the Smart Living domain the “one and only” Shared Data Space makes no sense. Use cases, services and applications are too heterogeneous to bring them all together in one Data Space. In fact, applications/use cases require specific Shared Data Spaces that provide precisely the data and resources required. Those project-specific Smart Living Data Spaces need to have reasonable concepts for semantic descriptions of and comprise cross-Data Space search functionalities for resource and data catalogues.

Use Case Smart Living Assistance/AAL

Solution

The elderly, sick, or care-dependent people are primarily faced with one question: Do I stay home, or do I move to a nursing home? One objective of “Smart Living” is to help the person concerned to maintain their independence and quality of life, despite health limitations. The ability for them to stay in their own home is an essential factor here. In addition, younger people may profit from the monitoring of health and fitness data to prevent e.g. civil illness.

To make this possible, an optimal living environment needs to be created with the affected person, their relatives, the care, medical and emergency services, insurance companies, local authorities, and, if necessary, the housing company. Solutions that assist within everyday life (Ambient Assisted Living (AAL) solutions) can be a great help here. In addition, medical treatments and diagnoses can also be supported with digital technologies. AI based Monitoring and hazard recognition of vital data based on embedded AI systems for example can support

online care and event detection by resident doctors e.g. in case of stroke patients. AAL solutions also comprise alarms to monitor and/or signal an emergency, smoke, air quality, indoor climate, water consumption/leakage, and stove use. They also encompass door openers and sensors, sensors to detect absence and monitor getting up as well as activity and inactivity, internet-based communication services and voice assistants. A Gaia-X Shared Data Space is a perfect basis for a confidential data protection compliant and scalable cross trade solution.

Problem solved

Through the use of AI, it is now possible to develop processes that incorporate existing sensor and actuator technology and offer better care and decision-making support by linking information from all data sources together. For example, information on electric power consumption, i.e. from smart meter data, could also be used to obtain important indicators of activity/inactivity or for monitoring residents' behaviour.

Besides, there are all kinds of services supporting everyday life up to the affected person's professional care and support. The array of processes concerned must be coordinated between all service providers and systems involved, and, in particular, information from AAL systems is made accessible to all the persons concerned so it can serve as an essential additional source of data.

In the case of pandemics and the associated restrictions on social contact, networked assistance solutions that assist with everyday life are essential for enabling the person affected to maintain their independence and quality of life in their interactions with other people.

Main technology/Gaia-X components

- Protected edge/cloud environment for IoT sensors on humans and smart home IoT devices
- Time-series data bases (e.g. Influx DB)
- Federated catalogue/registry
- Semantic description of all data and resources of things as well as thing locations
- Predefined semantic searches (SPARQL based)
- IDS as a semantic description of related clouds
- iotschema.org as IoT ontology and SENSE WoT for semantic description
- Keycloak as IAM system for all services
- OpenShift for containerization and deployment

Concrete benefits

Solutions that provide assistance with everyday life are used in various domains, each with different requirements depending on the particular protection classes concerned (for example, equipment technology, energy, health, and "human" data) Gaia-X-Governance enables all of these different requirements to be taken into account.

It is vital to ensure all participant's data sovereignty, provided that Gaia-X methods for data exchange in the AAL area are applied and supported. AAL can thus use all available data sources without violating data protection rights.

Within the Gaia-X framework, it would thus be possible to enable transparent and secure data exchange between the various actors (affected persons, relatives, medical, personal care services, emergency call services, health insurance companies, and corresponding AAL platforms).

Use Case Smart Living Energy Efficiency of Smart Buildings

Solution

Buildings are important entities in the context of a more flexible balancing of (smart) energy networks. The ratio of the total energy consumption is currently around 40%. If we want to realize the macroeconomic and overarching energy efficiency goals for individual buildings, neighbourhood solutions, quarters and smart cities, we need to create overarching, interoperable networks that cross local boundaries (Figure 20). Smart Buildings/Smart Homes must become an intelligent counterpart to smart and intelligent energy networks and should be enabled to internally manage their energy consumption and production. But only a cloud system with governance rules that provide clear guidelines for data protection will be able to manage the balancing act between data protection, individual user habits, and energy supply security in the long term. If such a solution were available, it would also be relevant for our secure critical energy infrastructure, which is under special regulatory protection. Here, clear distributed approaches under national co-regulation can guarantee availability. Gaia-X offers the opportunity to provide such a cloud infrastructure and protected access for all actors (residents, owners, energy service providers, and contractors) to the usage data needed for energy forecasting, e.g., energy consumption and behaviour-specific information such as arrival and departure times and any absences.

Gaia-X enables a uniform and holistic pool of information to be established on all relevant devices installed in a building. This is vital in order to develop scalable, cross-cutting energy products.

An overarching IAM-system provide secure access (confidential chain of action) to the devices from private and semi-public areas, provided that the customer or the owner has provided their consent. This access system can either be used to benefit the grid (clear guidelines, timetables) or consist of rate-based incentive schemes. After that, the focus will be on the mass market.

By pairing the iMSys infrastructure with the Gaia-X cloud infrastructure, further energy efficiency gains can be made.



Figure 20: Smart Building as entity in an energy network

Problem solved

Today, around 40% of final energy is consumed in our homes. Germany has around 21,7 million buildings that consume 870 TWh. As part of the energy efficiency strategy for buildings, Germany's building stock is to be made virtually climate-neutral by 2050. Meeting this goal will only be possible by using smart solutions that increase energy efficiency and cloud-based coupling between the electricity, heat, electric mobility, and buildings sectors. This will involve data-based solutions to network the relevant energy consumers, producers, and storage systems across different buildings. In addition to the devices that are already widely connected today - such as heating and cooling installations and all types of IoT and household appliances - an important role will also be played by power-to-heat (PtH), power-to-gas (PtG), power-to-mobility (PtM) as well as electricity and heat storage devices.

Buildings as well as associated charging infrastructures for e-cars are important factors for the balancing of any energy network. With more and more decentralized regenerative generation of electrical energy and reinforced by the development of private and public charging infrastructure for electromobility, the task arises in the distribution network to carry out suitable intelligent control. At grid nodes (transformer stations), for example, a grid-serving feed-in/load profile of the instantaneous electrical power can be aimed for. This can be done in relation to this point via a suitable intelligent central control system. The sum of the powers must be balanced at any time. In the future grid, it consists of the supply from the upper grid segment, the PV feed-in, the supplied/discharged power of the controllable electrical/thermal storage units, the load from charging stations, heat pumps and households, and the load from the companies in the grid segment. Controllable components can be made available through

appropriate switching on and off. The goal can be in this case to carry out a grid-serving balancing aggregation of the supplied and dissipated powers and, if possible, to follow a predefined overall load curve through demand side management. For this purpose, the instantaneous powers of the billing-relevant measuring points in the network segment must be recorded in fine granular form and the loads and storage facilities must be controlled appropriately according to the specifications of the central control system.

In the information technology model, the switching and measuring points within the distributed energy network are data points in the IoT that permanently supply or are supplied with data in encrypted form at specific measuring intervals in accordance with IT security requirements and in accordance with IT service quality (QoS) requirements in terms of sampling rate and packet delay. Endpoints/edge controllers (smart device controllers, SDC) are smart meters and switching devices at smart meter gateways (SMGW) or energy management controllers (EMC) that communicate via IP interfaces with cloud systems and a Gaia-X service environment.

By semantic description of the data points according to the WoT specification supplemented by the geo- position, a digital twin of the grid segment can be built up, which can describe the state of the distribution grid segment to be spatially observed in each measurement interval. This in turn can be used for intelligent dynamic adjustment of the network quality. By intelligent use of resources with a location-based service for intelligent demand side management and prediction of feed-in and consumption in the cloud, the peak power in the grid segment can be reduced. For example, a service in the cloud has the task of minimizing energy needed in the short term to regulate production and consumption by balancing feed-in, consumption, and storage. This saves both CO₂ emissions and power-related costs.

Another service in the cloud is, for example, the operational management of the assets of the network in terms of predictive maintenance. A reduction of possibly required network expansion costs due to the higher energy consumption of electromobility and the conversion of heating systems to heat pump systems and theoretically possibly also regenerative feed-in can be achieved.

The aforementioned services allow integration into energy marketplace systems that bring network aspects of local energy distribution networks into connection with energy trading transactions. This should also be seen in relation to the BDEW traffic light concept considered in Germany.

Main technology/Gaia-X components

- Protected edge/cloud environment for IoT sensors on humans and smart home IoT devices
- Time-series data bases (e.g. Influx DB)
- Federated catalogue/registry
- Semantic description of all data and resources of things as well as thing locations
- Predefined semantic searches (SPARQL based)
- IDS as a semantic description of related clouds
- iotschema.org as IoT ontology and SENSE WoT for semantic description

- Keycloak as IAM system for all services

Concrete benefits

In order to balance generation, storage, and consumption, various processes need to be established around the building, from power management at the grid connection (smart meter gateways), to smart energy management, through to energy trading across all buildings. Smart solutions are required that enable the systems to be designed around the individual needs of the residents.

Maturity indication of the data space

With the ForeSight Smart Living Shared Data Space, a demonstrator has been built for “Digitalgipfel 2020”. The demonstrator (Figure 9) is wholly based on W3C standards and known technical concepts and artifacts out of the Gaia-X technical working groups. Therefore, we would sort it in in TRL 5. We did not tackle and implement software artefacts for the topic trust. Our federated catalogue is based on RDF-triple-stores and SPARQL. Mapping services have been implemented to translate between different ontologies and data models.

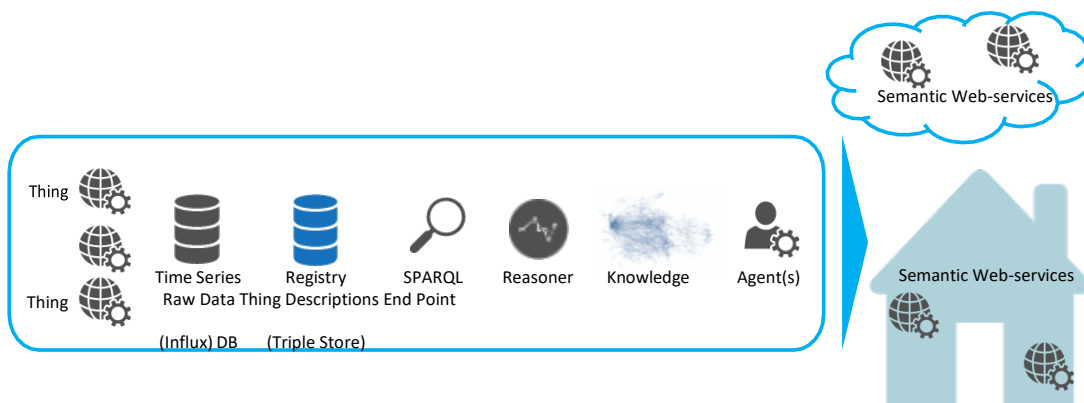


Figure 21: Elements of Smart Living Shared Data Space

One focus of the demonstrator is the concept of a registry/catalogue. This is because, the smart living data ecosystem is characterized by a high degree of heterogeneity and complexity. Use cases in the domain often require that previously isolated products, services, and data sources that are distributed among different actors are integrated into overarching systems. For example, the development of a smart building access management service (‘intelligent doorkeeper’) requires the integration of data from building ERP systems of the housing industry, biometric data of residents and visitors, context data, and the integration of services such as an AI-based facial recognition and a door lock API. In turn, the processing of this sensitive data requires a high degree of transparency and traceability, which is primarily in the hands of the service engineer (i.e., the developer) of the overall service, so that he or she can consider data flows and data storage early in the development process. To achieve this, it is essential to provide the service engineer with tools to evaluate relevant information about products, services, or data in order to help him make the best decisions and thus certify to users the privacy-compliant characteristics of the overall system. For this purpose, the service

registry was developed as part of the existing demonstrator. It instantiates the basic idea of a Gaia-X federated catalogue for the Smart Living domain and extends its scope to the service level. The registry allows service engineers to register their services in a central system, store meta-data that facilitates the use of the services by other actors of the ecosystem and search it to make use of existing services. The registry thus represents the central starting point for service development within the Smart Living data ecosystem. On the one hand, it lists services currently registered in the domain (Figure 22), on the other hand, it provides detailed information on individual services (Figure 23). The detailed information includes, for example, information on the provider, the version, dependencies, the costs, or the security classification. To ensure compatibility the self-description of each service is based on the W3C Web of Things (WoT) (<https://www.w3.org/TR/wot-architecture/>) description and extended by domain specific aspects like the Smart Readiness Indicator for Buildings (<https://smartreadinessindicator.eu>) of the European Commission. In addition, the registry also includes Gaia-X specific information, such as the ID and physical location of the executing Gaia-X node. Since the ecosystem's services often process sensitive data, information about the protection classification of the data being processed is crucial for the engineer to ensure appropriate processing that meets the requirements of the data provider. Of particular importance is the data visualization component of the service registry. Its goal is to increase transparency within the data ecosystem by showing data flows between individual services. In this way, data producers who share data with the ecosystem can get an overview of how this data is further processed. On the other hand, the user of a service (e.g., an end user or an intermediary, like the above-mentioned service engineer) can track which services a particular service uses and thus ensure compliant use of his own offer.

Our prototype of the service registry demonstrates the concept for the domain Smart Living following central Gaia-X values: Transparency and interoperability. This is realized by extending domain-specific standards (W3C WoT), allowing access via a web-GUI and a REST API and providing information on dataflows.

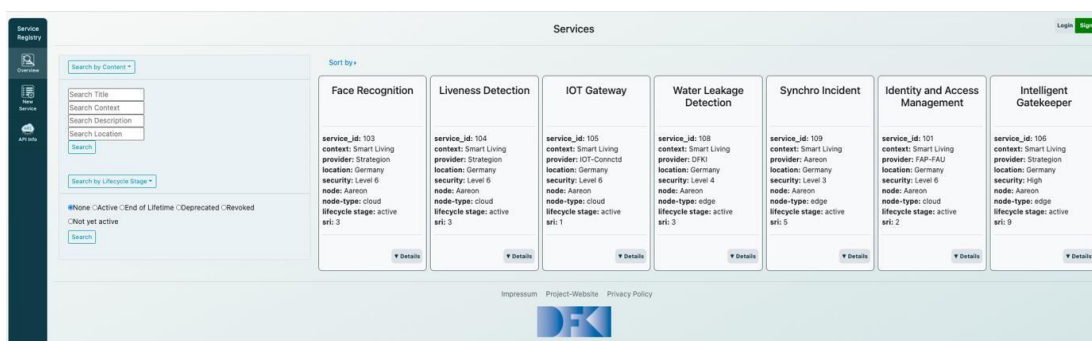


Figure 22: Screenshot of the landing page of the service registry showing all available services

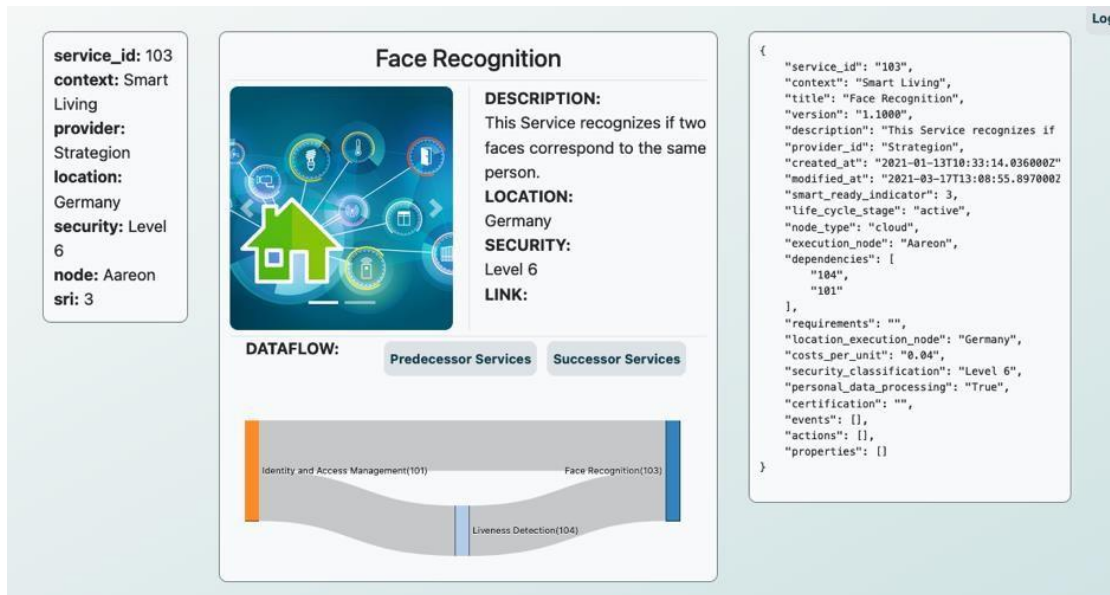


Figure 23: Detailed information regarding a particular service, including the transparent visualization of data flows and the service self-description

Main technology/Gaia-X components

Figure 24 provides an overview on technical components of the Smart Living Shared Data Space that have been partially used for the implementation of the demonstrator described above.

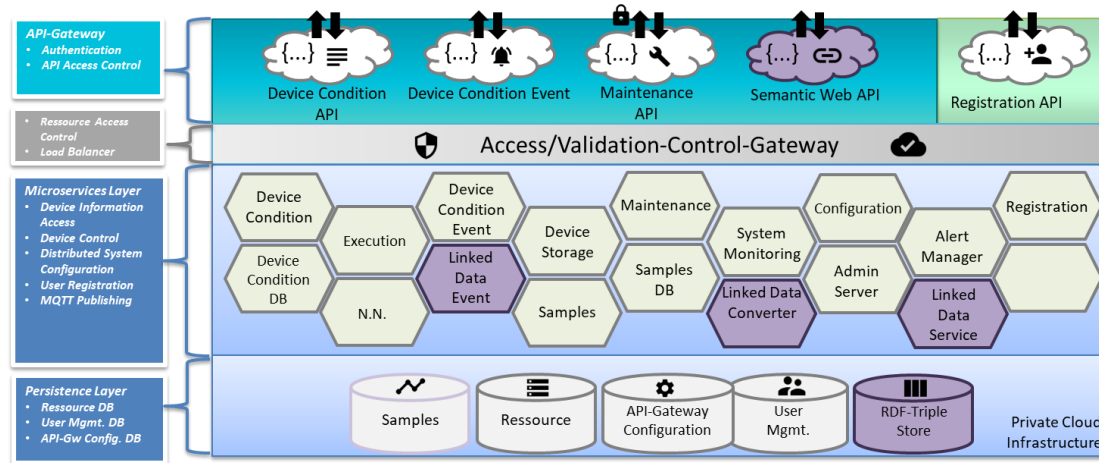


Figure 24: Technical components of the Smart Living Shared Data Space

Technical overview Key components are:

- Protected edge/cloud environment for IoT sensors and actuators
- Time series data bases (e.g. Influx DB)
- Federated catalogue/ service registry
- Semantic description of all data and resources of things as well as thing locations
- Predefined semantic searches (SPARQL based)
- IDS as semantic description of related clouds

- iotschema.org as IoT ontology and SENSE WoT for semantic description
- Keycloak as IAM system for all services
- OpenShift for containerization and deployment

How is the demand side represented?

Partners of the ForeSight community.

How is the supply side represented?

Partners of the ForeSight community.

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

Not yet.

Is the story of the data space well documented?

Scientific publications regarding the demonstrator and its components are currently written, a promotion video by the Digitalgipfel 2020 can be accessed here:

<https://foresight-plattform.de/newsroom/digitalgipfel-2020/>

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

An essential part of Smart Living ecosystems/projects is service engineering. The concept of basic services, predefined AI environments, etc. are made to simplify service engineering and application development. The community comprises numerous service developers of the domains housing industry, IoT, assistance, maintenance, and energy management. The partners take over at least parts of ForeSight Smart Living AI Shared Data Space and are currently integrating those concepts in their own business mechanics. The process is still ongoing because ForeSight is running for another 1 ½ years.

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

Catalogues/registries, cloud descriptions, semantic searches, mapping, semantic databases for IoT Things, AI tools and layer, deployment of services, transfer of services between clouds, integration of edge devices and maybe the concept of digital twins.

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

One smart building might not be a considerable challenge in terms of resources required. However, if we take energy efficiency and societal changes seriously, smart Living should be more or less completely implemented for 21,7 million buildings in Germany and 33 million in France. Around 100 million smart homes are expected in the year 2025 (Figure 25 and Figure 26)

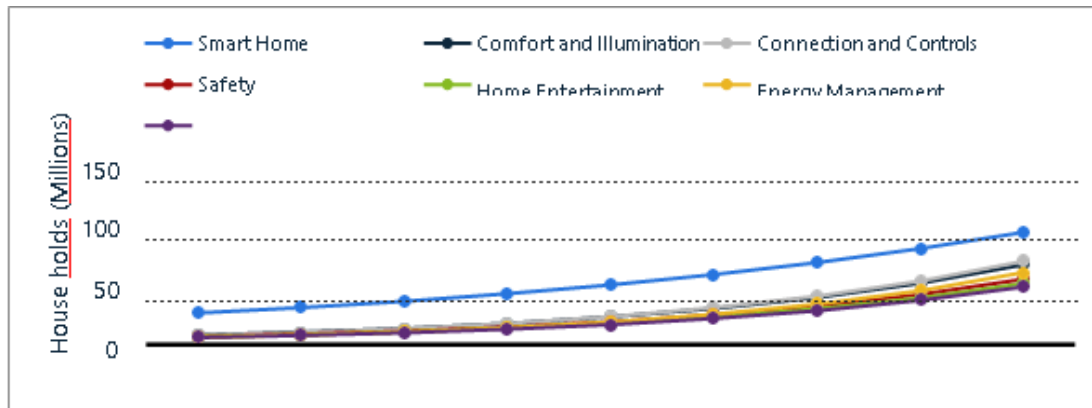


Figure 25: Forecast of the number of smart homes in Europe

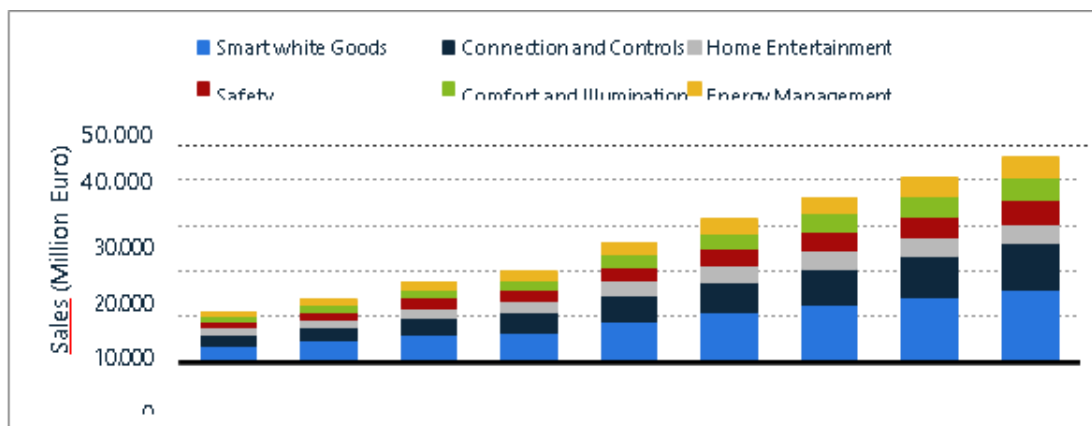


Figure 26: Forecast of sales

How can the commitment of the parties involved be proven?

All partners involved are already partners of the ForeSight community. All partners are actively participating on development of numerous use cases in the field of assistance, energy efficiency, safety, and predictive maintenance. Nearly all partners are now members of Gaia-X working group Smart Living.

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

Yes, due to funding of the BMWi.

Evolution of the data space

Roadmap of the evolution

The roadmap (Figure 27) follows the work plan of ForeSight and possibly a domain-specific Gaia-X project currently in the acquisition.

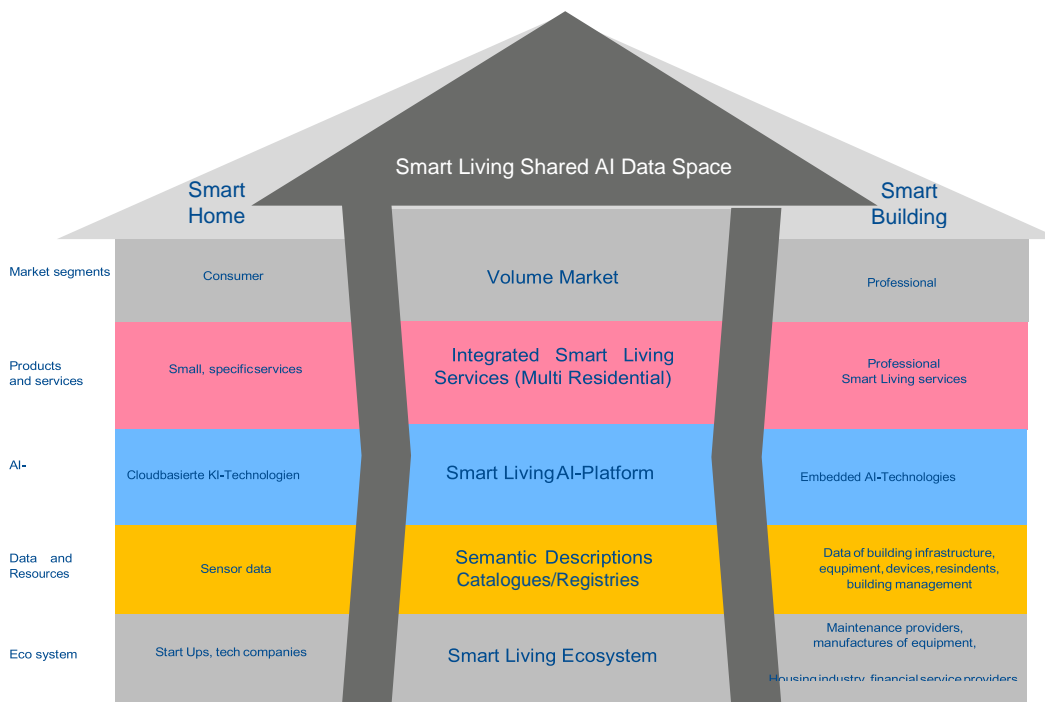


Figure 27: Roadmap towards Smart Living Shared AI Data Space

Quick-wins (for 2021)

The technology demonstrator is available and ready to be presented. All basic components of our Data Space are ready to run and are used in different use cases and application developments along with the ForeSight workplan. The Data Space is expected to grow during runtime of ForeSight. The concept of associated partners within ForeSight (currently 70 partners) will bring up new service ideas and maybe prototypic implementations for our domain.

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

Following the ongoing development within the context of ForeSight we expect a first integrated prototype of our Smart Living Shared (AI) Data Space at the end of 2022 to contain all basic AI-services, digital twins, test data and connections to iot and manufactures clouds, AI-cloud and a data cloud. We expect useful evaluation of the Kafka streaming platform in the context of Smart Living also.

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

The next step is to realize a prototype of a cross-domain data space based on Gaia-X.

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

The Smart Living domain will definitely take advantage of a future Gaia-X infrastructure. Due to the heterogeneous character of our domain many different and up to now separated edge and cloud solutions must work together seamlessly for innovative interoperable (AI-based) applications and services. Existing cloud spaces and interfaces are using W3C standards. We recommend that Gaia-X respects and strengthens these standards and possibly participates actively at W3C. We also recommend developing Gaia-X in a way that enables service developers to more or less easily setup cross domain, maybe AI-based applications that are using many different data sources. We support the driving vision of more data sovereignty.