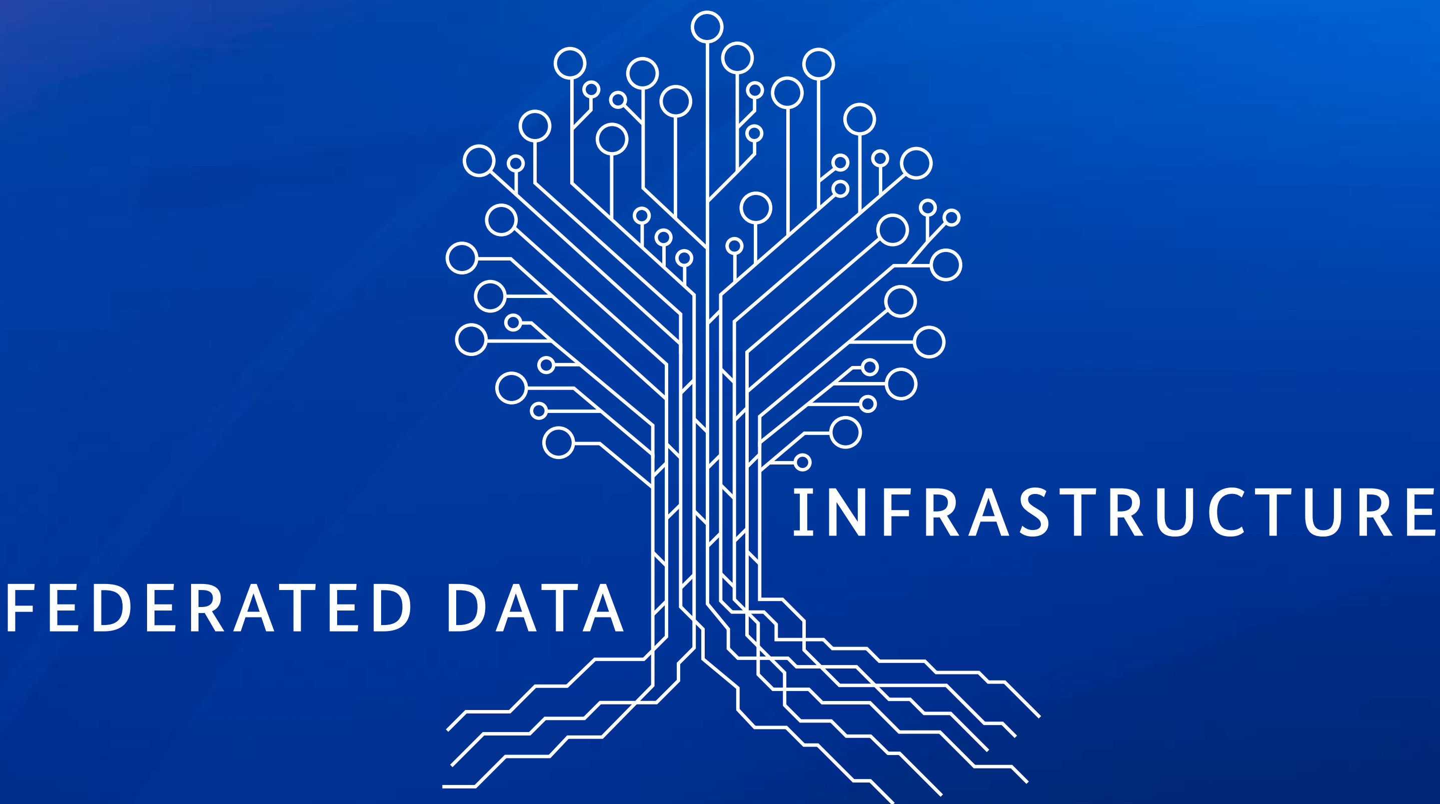
Federal Ministry
for Economic Affairs
and Energy

Federal Ministry
of Education
and Research

Project GAIA-X

A Federated Data Infrastructure as the Cradle of a Vibrant European Ecosystem



Imprint

Publisher
Federal Ministry for Economic Affairs and Energy (BMWi)
Public Relations
D-11019 Berlin, Germany
www.bmwi.de

Status
October 2019

Design
PRpetuum GmbH, 80801 Munich

Illustrations
gettyimages
Alistair Berg / p. 23
ANDRZEJ WOJCICKI/SCIENCE PHOTO LIBRARY / p. 10
Busakorn Pongparnit / p. 16
Caiaimage / Agnieszka Olek / p. 29
Caiaimage / Robert Daly / p. 7
Erik Isakson / p. 43
Monty Rakusen / p. 19, p. 21
MR.Cole_Photographer / p. 4
powerofforever / p. 33
Qi Yang / p. 45
svetikd / p. 31
VICTOR HABBICK VISIONS / p. 39
Westend61 / p. 25
Wutthichai Luemuang / EyeEm / p. 37
Yuichiro Chino / p. 35, p. 41

This publication as well as further publications can be obtained from:
Federal Ministry for Economic Affairs and Energy
Public Relations
Email: publikationen@bundesregierung.de
www.bmwi.de

Central procurement service:
Tel.: +49 30 182722721
Fax: +49 30 18102722721

Contents

1. Preamble

We, representatives of the German Federal Government, business and science communities, are striving to set up a high-performance, competitive, secure and trustworthy data infrastructure for Europe. To this end, we have drawn up the foundations for a federated, open data infrastructure based on European values, giving it the provisional project name 'GAIA-X'. The project serves the following objectives:

1. Stating the specific technical and economic conceptual work involved in building such an infrastructure; and

2. On this basis, creating a common ecosystem of users and providers from public administration, the health sector, enterprises and scientific institutions; and

3. Establishing supporting framework conditions and structures.

We see 'Project GAIA-X' as the cradle of an open, digital ecosystem where data can be made available, securely collated and shared while enjoying the trust of its users. Our goal is to join forces with other European countries to create the next generation of a federated data infrastructure for Europe, its states, companies and citizens; a data infrastructure which satisfies our highest aspirations in terms of digital sovereignty while promoting innovations.

Taking European values as our starting point, we are guided by the following principles:

1. European data protection

2. Openness and transparency

3. Authenticity and trust

4. Digital sovereignty and self-determination

5. Free market access and European value creation

6. Modularity and interoperability

7. User-friendliness

Business, science and politics have firmly resolved to work together to set up the preconditions for an innovative data economy that is focused on the future, in both Germany and the rest of Europe. Business and society rightly expect the digital infrastructure to provide a high level of security and availability.

We understand **data infrastructure** as a federated technical infrastructure, consisting of components and services that make it possible to access data and to store, exchange and use it according to predefined rules. We understand a **digital ecosystem** as the network of developers, providers and users of digital products and services, connected with transparency, wide-based access and a vibrant process of interchange. Such a system thus serves as a crucial foundation for European growth, digital innovations and new business models.

We are banking on Europe's enduring strengths. Among others, these include the diversity of offerings, together with strong decentralised structures suitable to small and medium-sized firms. In this way, we link up the many investments in digital technologies made throughout Europe, enabling them to have an even greater effect.

Project GAIA-X makes provision for the federalised structuring of infrastructure services, especially cloud instances and edge instances, to transform them into a homogeneous, user-friendly system. The federated form of data infrastructure that results from this strengthens both the digital sovereignty of sources of demand for cloud services and also the scalability and the competitive position of European cloud-service providers.

The openness to national and European initiatives with a similar objective gives the project a decisive impetus, directed at a joint European solution. Building upon existing solutions and their further development, we want to launch competitive offerings from Europe out onto the world market. Involvement is also open to market participants outside Europe who share our goals of data sovereignty and data availability.

The federated data infrastructure is tailored to the needs of both the providers and users: it enhances transparency and visibility on the provider side; advances innovations in the data economy; clearly commits to interoperability of offerings and links up companies of all sizes – from major industrial corporations to small and medium-sized enterprises (SMEs), through to start-ups.

In order to implement the federated data infrastructure, we consider it necessary to establish a central organisation at the European level. This organisation would lay the economic, organisational and technical foundations of a federated data infrastructure. Its task will be to develop a reference architecture, to define standards and to determine criteria for certifications and product quality seals. It should be a neutral mediator and the hub of the European ecosystem.

Through this concept of a federated data infrastructure, we enable Europe to develop its potential in its data economy with a dynamic ecosystem. We aim to establish an ecosystem that distributes sovereignty and benefits among business, science, the state and society in equal measure.

2. Current Situation and Motivation

2.1 Trend

The advance of digitalisation generates constantly increasing quantities of data; the use of this data promises substantial potential to society and the economy. For instance, as a society we benefit from improvements in healthcare, in the targeted distribution of scarce goods, and through greater resource efficiency. The economic advantages include productivity growth, process optimisation, or innovations in the form of new products and services. Linking up and analysing various data sources open up additional value creation opportunities, notably thanks to the methods and processes of artificial intelligence (AI).

The central gathering and analysis of much of this data in the cloud indicates a higher level of value creation in the digitalisation process ('As A Service' model), especially for consumers. This development explains why the rapidly scaling cloud offerings have emerged from the market of large web providers. The existing cloud offerings are dominated by non-European providers with significant market power and rapidly upscaling cloud infrastructures. European alternatives do not offer comparable market capitalisation, scalability or breadth of applications; they are active in specialist niches at best.

We understand **Cloud Computing** as the offer, use and charge for IT services dynamically adapted to demand and supplied through a network (consistent with the definition by the Federal Office for Information Security, BSI[1]). The breadth of the services offered in the cloud-computing context encompasses the whole spectrum of information technology – among other things, infrastructure (e.g. processing capacity, storage space), platforms and software. Thanks to the use of cloud services which are only loaded onto the customers' end-devices for further processing where necessary, the customer enjoys reduced costs for investment and for providing, operating and charging for these services. In this regard, cloud services are offered and used solely through technical interfaces and protocols.

A paradigm shift in cloud use is now evident: in the Internet of Things (IoT), huge data quantities are generated on a decentralised basis (e.g. on a sensor or a wearable device). This trend is also boosted by the increasing performance capability of mobile end devices. Due to real-time requirements, or due to intellectual property protection or data protection, this data often also requires decentralised processing. These are scenarios where the so-called Edge and Cloud solutions come into play.

1 Federal Office for Information Security 'Basics of Cloud Computing' *(Bundesamt für Sicherheit in der Informationstechnik (BSI)/Cloud Computing Grundlagen: https://www.bsi.bund.de/DE/Themen/DigitaleGesellschaft/CloudComputing/Grundlagen/Grundlagen_node.html)* (called off: 14.10.2019)

We understand **Edge** as a decentralised data architecture principle. Edge Computing processes data not solely in a cloud but also where it is generated, worldwide, i.e. close to the production processes – also using cloud technologies. This is highly relevant for real-time applications where a few milliseconds of reaction time (latencies) are decisive, thereby making cloud processing impossible due to its time constraints. This approach guarantees a subsequent further processing in the cloud, which is expected to be possible.

At the same time, companies or organisations are part of complex value creation networks. The digital transformation confronts them with a double task of integration: They must orchestrate the interplay of edge-based and cloud-based solutions and resolve this challenge across the boundaries of individual companies.

This transition becomes especially evident in the realm of Industrie 4.0: in production, the sensors of countless machines generate vast quantities of data which are administered by various manufacturers. Due to the proximity to production required by the technology, this data is processed in Edge computing centres. For analysis, the various machines' bodies of data must be both interconnected and linked up with business data. This process generates efficiency gains, production improvements and additional value creation. If we want to make greater use of these advantages, we need to use distributed bulk raw data to generate usable, data-based and domain-specific data pools. These pools serve as the basis for innovative business models across individual companies' boundaries. They thereby secure our future economic viability and value creation in industries in which Germany, and Europe more broadly, are presently market leaders worldwide.

So far, the response to these challenges has been individual project solutions or bilateral solutions, at best. The federated data infrastructure built upon in the project links up existing and new upcoming edge and cloud solutions respectively. Combined Edge and Cloud computing connects the advantages of decentralised and centralised data architectures; it does so – depending on the application case and necessary processing times – by distributing and coordinating the data processing, storage and analysis between local Edge services and central Cloud services. A federated data infrastructure ensures broad data availability, while taking into account the prevailing requirements.

2.2 Our goals

2.2.1 We are striving for data sovereignty

Europe faces the following challenge: firstly, maintaining its liberal and socially cohesive economic and societal model in the face of increasing dependence on critical digital technologies (e.g. for obtaining, exchanging, and analysing data) and oligopoly tendencies in the platform economy; secondly, positioning that model in a context of international competition. International tensions, trade conflicts and the digital divide intensify this problem. This discussion is also reflected at the European level. We must safeguard our strategic capacity for action in order to be able to operate digitally in the long term on a free and self-determined basis. For this, we must also maintain digital sovereignty in the realm of data.

We understand **digital sovereignty**,[2] consistent with the definition of terms used by the Digital Summit Focus Group 'digital sovereignty in a Connected Economy', as the 'possibility of independent self-determination by the state and by organisations' with regard to the 'use and structuring of digital systems themselves, the data produced and stored in them, and the processes depicted as a result.' Our project primarily addresses the aspect of data sovereignty included in this definition of the term: i.e. 'complete control over stored and processed data and also the independent decision on who is permitted to have access to it.'[3]

We strive towards a data infrastructure that does justice to the freedom-based values and self-determination of all European citizens and companies, thereby guaranteeing their data sovereignty. In doing this, we invoke Europe's strengths, emphasising diversity, open ecosystems and plurality. Our goal is fair and equal conditions of competition and the endorsement of free competition by all market players, committed to non-discrimination with open systems as the foundation of a joint partnership with international service providers.

2 **Note for the English version:** The German term "Digitale Souveränität", which is used in the German original version of the concept paper, does not have a direct equivalent in the English language. Both "digital sovereignty" and "digital autonomy" are frequently used, with slight variations in meaning. When we use the term "digital sovereignty" in this translation, is has the meaning defined in more detail in this paragraph. The same applies mutatis mutandis to related translated terms, e.g. "Datensouveränität"/"data sovereignty".

3 Definitions of terms from 'Digital Sovereignty and Artificial Intelligence – Preconditions, Responsibilities and Recommendations for Action', Focus Group 'Digital Sovereignty in a Connected Economy', 2018; 'Digital Sovereignty in the Context of Platform-Based Ecosystems', Focus Group 'Digital Sovereignty in a Networked Economy', 2019; and also from 'Role Model 2030 for Industrie 4.0 – Structuring Digital Ecosystems Globally', Plattform Industrie 4.0, 2019.

2.2.2 We want to reduce dependencies

In principle, a digital infrastructure encompasses three architecture levels:

1. data-transfer networks and hardware (the network level),

2. A data storage level, including operating systems and databases for data storage (the data level),

3. A data-processing and use level, including application systems, functions and services (the service level).

By "data infrastructure" we refer to the second and third architecture levels. Advantages of scale are decisive for the market success of data infrastructure services.

The European economy's strength is primarily founded upon highly-specialised domain knowledge of industry and a high degree of integration competence in complex value creation networks. These capabilities will also be used to build up digital 'business to business' (B2B) platforms that make new business models possible. These platforms' success primarily depends on access to data or respectively on companies' willingness to share it in a way that enables monitoring and inspires trust. This is precisely our approach with Project GAIA-X. In many industry domains, data are highly sensitive or needs to be protected. The current market structure risks dependency on international providers. Technical, economic and contractual obstacles in the path of a data migration to another infrastructure provider (so-called 'lock-in effects') constrain companies' freedom in their operations – both in running their firms and in the event of political conflicts. In order that platforms and whole industries can continue to secure and expand their value creation, data infrastructure that strengthens the users' digital sovereignty is necessary.

Lock-in effects emerge between customers and providers of cloud services if the switchover to an alternative provider of solutions or services is made more difficult, or indeed impossible, by switchover costs and barriers. The barriers to a switchover can be of a technical-functional kind (dependence on the specific features of certain providers); they can arise from contractual agreements (e.g. license models and penalty costs), but also result from a high, customer-specific degree of personalisation, from familiarisation effects, or from the sheer data volume that is to be migrated.

2.2.3 We want to make cloud services attractive on a broad basis

In Europe, SMEs, in particular, often remain critical of cloud services. This is due to a lack of trust in the existing product offerings, misgivings about high investment costs, and a lack of specialist personnel in companies, as well as a fear of dependency. Those opting to deploy cloud services often only use a small sub-area of the whole spectrum of services offered.

This may lead to competitive disadvantages, especially for SMEs: efficiency gains are not taken advantage of and innovative ideas are less likely to be translated into new business models.

This is why companies increasingly bank on the simultaneous use of several cloud providers (so-called 'multi-cloud strategies'). Yet this also brings greater complexity and the challenge of achieving a uniform approach to link-ups, semantics and data-processing. A European federated infrastructure that creates trustworthy cloud offerings according to clear rules, e.g. based on existing solutions or also on international providers, and extended to include edge components, addresses these reservations, notably on the part of SMEs.

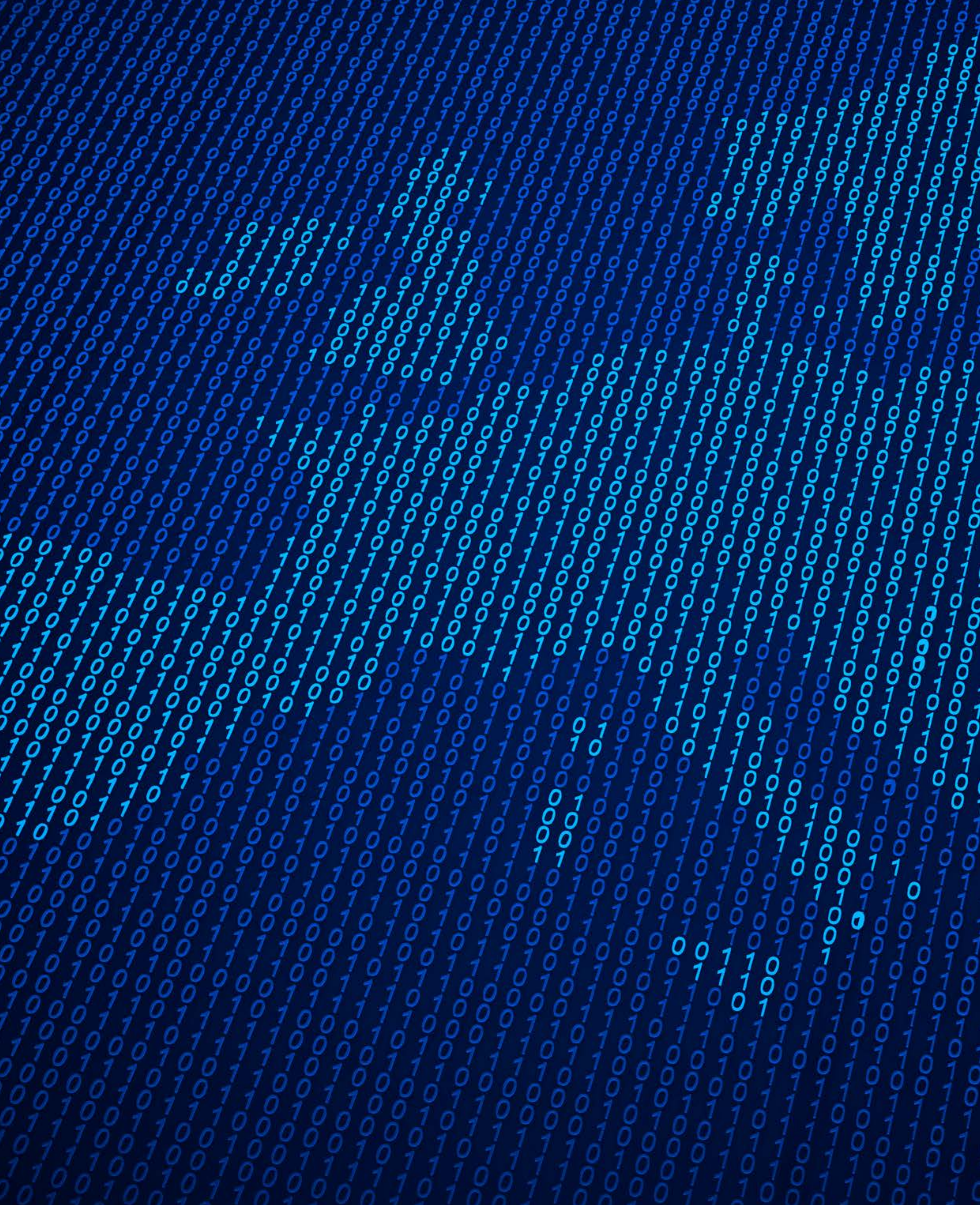
The core of such an infrastructure is to create the necessary trust, most notably for SMEs that usually only have limited quantities of data. In the future, the federated data infrastructure will enable SMEs to use their data jointly and better, thereby breaking through one of the key barriers to new business models.

2.2.4 We are creating an ecosystem for innovation

Europe is making extensive investments in digital technologies and innovative business models. We must ensure that those who drive innovations forward are also those who benefit in economic terms. Our Project GAIA-X lays the foundations for an open, digital ecosystem to help European companies and business models to scale up competitively worldwide. This is how we secure European value creation and employment.

Value creation and employment emerge through innovation. And innovations emerge due to ideas, the linking-up of information and data and also through cooperation among intelligent people. Intelligent people want challenges, attractive conditions for bringing their ideas to fruition and the necessary resources for implementation. In this context, the scaling-up – and thereby the long-term success – of many ideas and business models also depend both on access to risk capital and the availability of large quantities of data. Thus, what we need is a simple, reliable opportunity to voluntarily exchange and to jointly use Europe's data treasure trove. Our goal is to unleash this creativity. This requires start-ups, science and companies of all sizes to be able to cooperate even more closely and easily with one another. We must raise the visibility of European offers, network the providers and users and create an attractive environment for investments, smart people and powerful ideas.

3. Approach to a Solution: GAIA-X



Driven by the developments described (Industrie 4.0, edge computing, cloud computing, etc.) multi-cloud strategies and thereby interoperability are of increasing significance. Users expect flexibility, functionality, usability, global availability, services across company boundaries, specialisation of services, or distributed data processing and data holding. Simple migration to other cloud or edge providers must be possible. Users also demand interoperability, i.e. finding services and using them efficiently via new intermediaries for data and services. This also includes transparency about services offered. They would like to be able to distribute the data processing flexibly across many providers, availing themselves of robust processes as they do so.

Interoperability is therefore one of the central requirements of a federated data infrastructure. It is taken into consideration on three infrastructure levels: technical and semantic interconnectivity at network, data and service level. Within Europe there are already network infrastructures and concepts of interconnectivity for cloud infrastructures that can be integrated through our approach, and indeed that make a distributed mode of working possible in the first place. Accordingly, for instance, three of the five largest interconnection hubs (IXPs) worldwide are in Europe.

A federated data infrastructure must enable both intra and inter-domain-specific exchange as an added value for the user; it must also enable data and services to be interlocked, across the boundaries of individual providers and customers. In particular, collaboration between edge instances and cloud instances must be eased. This lowers the hurdles to entry into edge and cloud computing, for SMEs as well as others. Common standards help in breaking up domain-specific data silos, which cannot be linked-up and assessed due to a lack of data interfaces. Tailored solutions emerge for each application case. New value creation networks and processing networks become possible, and new, extensively distributed processing models and business models are promoted, such as intermediary activities for data or services. Our project can support these requirements in the sense that basic services for interoperability are offered; users and providers respectively use these for domain-specific interoperability services. This ranges from matching systems and vocabulary systems right through to contractual and commercial basic services for invoicing, certification and contracts.

3.1 Target

We are pursuing the goal of creating a secure and federated data infrastructure in Germany and in Europe more broadly. ‘Project GAIA-X’ strengthens data sovereignty for business, science, the state and society, in the context of storing, exchanging and using data and services. The federated data infrastructure makes data and services available for artificial intelligence applications; it protects rights, interests and intellectual property – regarding the data and the expertise associated with it. The infrastructure is provider-independent; it takes into account the interests of data producers, providers and users. The ecosystem extends domain knowledge among European players, reinforces the exchange of inputs both within domains and between them, and makes innovative business models possible in the digital single market.

The federated data infrastructure strengthens fair competition in the value creation networks of the global data economy. It offers the market players efficient access to all relevant digital or cloud-based applications while simultaneously ensuring the highest possible degree of self-determination and data sovereignty.

3.2 Solution

'Project GAIA-X' connects centralised and decentralised infrastructures (in particular cloud services and edge services) into a homogeneous, user-friendly system. The distributed ecosystem that emerges from this strengthens both the digital sovereignty of sources of demand for cloud services and also the scalability and competitive position of European cloud providers. In particular, SMEs gain from the market transparency, broad-based access to alternative solutions and the resulting possibilities for action. This ecosystem also takes into account various preferences regarding security aspects, latency times and the breadth of application; it supplies tailor-made solutions and enables the use of various cloud providers.

The advantages of pooling together, which we attain with the federated data infrastructures, merit particular emphasis. The multiple use of cloud and edge instances leads to efficiency gains due to an increase in performance range. Redundant nodes increase security. If a node becomes non-operational (for instance caused by a power outage or a natural disaster), another node comes into action. Moreover, calculation-intensive applications can be processed 'in a distributed way', on cloud instances arranged for that purpose; this leads to a more than proportional boost in performance. Users can always use the most modern cloud technology, because individual new investments and replacement investments are continually introduced in the alliance between the many nodes. Based on the federated data infrastructure, an ecosystem can also be established for software development projects that benefit business, science, the state, and society. The federated data infrastructure's declared commitment to neutrality stabilises this ecosystem: each cloud-service provider, whether already on the market or new, can become a node in the network (GAIA-X node) by using GAIA-X technology and its reference architecture.

The essential characteristics are the following:

- Use of secure, open technologies provided by the federated data infrastructure; interfaces for simple and secure data-exchange; possibilities to use third parties' applications and functions; and adherence to standards that make data integration easy. In this context, the available technologies are used; meanwhile, technologies and services that are lacking are developed and made accessible by the ecosystem's participants, also on the basis of open-source technologies.

- Each node of the federated infrastructure forms an autonomous unit, follows the reference architecture, and is unambiguously identifiable and reachable. A self-description, linked-up with this, provides transparency regarding the individual nodes' specific aspects and capabilities. Notably, it includes statements on the location of data storage and processing; on the technologies used; on calculation and storage performance; and on the functionality made available. In addition, the following characteristics are stated: real-time capability, data sovereignty based on certified degrees of protection; consumption models (e.g. spot market); the price model and environmental sustainability (e.g. energy efficiency and consumption). The self-description requirement plays an important part in strengthening data sovereignty in fields of application that are dependent on having data storage within the country, for instance, or at least within the validity area of the EU's General Data Protection Regulation (GDPR). As part of this, the nodes can be set up both as a public and as a private cloud and also as edge nodes.

- A software repository makes components available that, depending on the categorisation, must or can be used by all providers, especially identification services and authorisation services, interface components, or certificates. In technical terms, these components could be made available on a centralised basis, via a 'peer-to-peer' approach, or distributed across several nodes.

- The individual services within a node and for the exchange between different nodes are implemented as functions ('Function as a Service'). This makes the following possible: a high level of interoperability between different nodes; exchangeability among different service providers (to avoid a lock-in effect); and efficient use and charging of services. The interfaces, services and products needed for this are to be harmonised through standards; they must also be able to be identified simply and used easily in a central directory for all participants. Through this we develop an infrastructure that is the basis for a vigorous and growing digital data ecosystem.

- Services from the GAIA-X ecosystem can be used directly through the provider or via digital platforms. For this purpose, an overview showing all available nodes and service capabilities is being made available; on its basis, the users themselves or the digital platform, as the intermediary, select the respective suitable service. Backed by this central directory service, we help to find suitable service providers and to identify relevant data pools quickly and securely. Alongside the technical agility, and in order for the contractual framework conditions to be in place, flexible framework contracts are introduced (possibly in the form of Smart Contracts); they fundamentally depict contractual rulings; when the respective service is called up, their role is in specifically concluding a contract between user and provider.

A precondition for taking part in the ecosystem is the respective service providers' binding commitment to common rules. Proof of adherence to these rules can be provided through a certification of the providers, nodes and also services, especially with regard to the necessary prerequisites in technical-organisational terms. In particular, this includes IT security, service levels, the degree of data sovereignty attained, and contractual framework conditions. It is a matter of principle that the certification should be proved by means of a transparent check undertaken by an independent, trustworthy third party. The process for this builds upon auditing and certification procedures already established, as well as others currently emerging (e.g. minimum standards for the use of external cloud services of the Federal Office for Information Security (BSI), C5, ISO 27001 and Trusted Cloud). Innovations must be promoted in order to move towards certifications that can be made on a technically automated basis. Existing reference architectures are to be included, e.g. those of the 'International Data Spaces (IDS) initiative'.

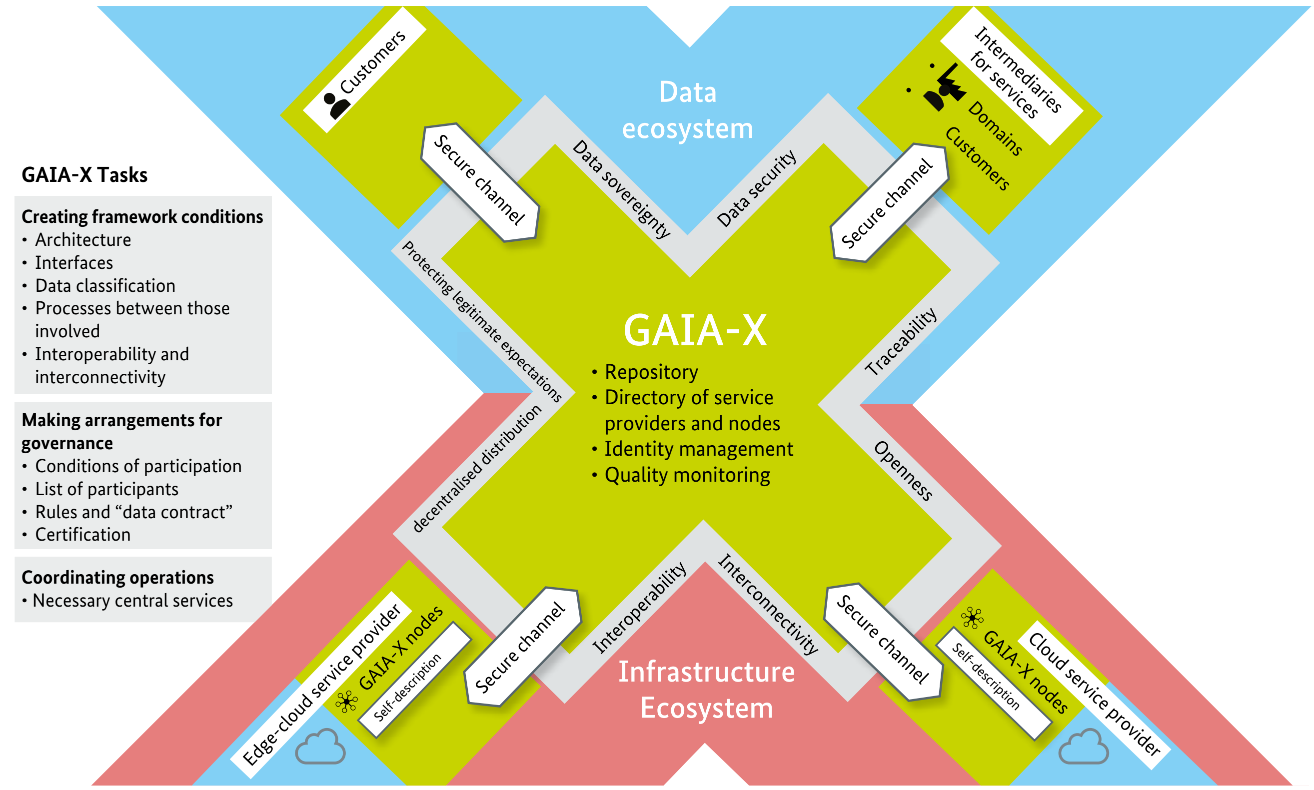
The **International Data Spaces Association** offers a reference architecture sustained by over 100 partners; this makes an ecosystem possible for the exchange of data, maintaining data sovereignty, and with clearly defined rights of use; it thereby describes a component of the federated, open infrastructure which Project GAIA-X is striving to attain. The reference architecture of the International Data Spaces Association defines a technical infrastructure and a semantic body of rules for data exchange and data use in ecosystems. In technical terms, one software component, the IDS Connector, either allows data link-ups and data analyses between various members of an ecosystem, or blocks them respectively. In this way, existing and new cloud services can be embedded in an interoperable digital business, while data sovereignty is safeguarded.

From the outset, implementation of the project is European and lends itself to international link-ups: existing offerings and initiatives with similar objectives will be integrated into the concept's further development, based on cooperation, to safeguard rapid availability. For instance, the IDS initiative's complementary activities to pursue data sovereignty are used, as is the Trusted Cloud's initial work on certification. Likewise, European initiatives and activities, first and foremost the European Commission's projects, will be included in the remainder of the process.

3.3 'Project GAIA-X' from the user's perspective

The federated data infrastructure is the basis for an ecosystem that effectively integrates the specific strengths of various participants and promotes cooperation. The users obtain access

Figure 1: Overall picture of data infrastructure and ecosystem



Source: BMWi

to a relevant portfolio of products and services: the federated data infrastructure

- integrates existing digital and cloud-based state-of-the-art products and services; it also offers the possibility that further specific areas of need can be addressed by modular offerings, e.g. from specialised or small-scale providers.

- offers full transparency by providing authentication on verified data protection and regulatory criteria of the products and services offered. Via self-description, it provides transparency regarding the level of confidentiality of all the ecosystem's participants. This is reflected in the guarantee on data use monitoring (data sovereignty).

- simplifies the management of IT interfaces and integration, especially regarding multi-cloud strategies and data-pooling; this is done by a high level of interoperability of compatible products, as well as by the provision of authentications that span across individual security domains. Domain-specific data silos that cannot yet be linked-up and assessed due to a lack of data interfaces, could now be broken up. As a consequence, lock-in effects can be avoided. This makes it possible, or easier, to have perfect-fit solutions for the relevant application. Moreover, an important contribution can be made to acceptance of AI use in instances involving particularly sensitive data.

- makes it possible to store data where users consider it useful to do so, considering the respective data classification. This means that the user can retain command over particularly sensitive data, while simultaneously sharing other data with partners for joint use.

- creates the preconditions for optimising the users' data strategies. Their nodes, i.e. decentralised and/or centralised cloud infrastructures, can be linked-up with one another. This link-up generates options regarding how data and algorithms can be used securely. So, for instance, various cooperation partners along the value creation chain are also given the possibility of migrating the data to the applications. To protect intellectual property, users can thus retain their algorithms and data with themselves, for instance.

- makes an important contribution to the emergence of digital ecosystems in the various user domains, by enabling them to make the transfer from bilateral individual-project solutions to marketplace solutions. Standardised contracts and procedures reduce transaction costs, data markets can emerge and data availability is improved.

3.3.1 Examples of demand from the user's perspective

To illustrate the project's added value, we now present some use cases, taken from the following sectors: Industrie 4.0/SMEs, Smart Living, the finance sector, health, public administration, and science. This compilation serves as a set of examples, to the extent that the selection neither claims to be complete, nor sets priorities. The aim is for the examples to highlight areas of potential for federated data infrastructure, based on application patterns that, in principle, are also relevant to other sectors (e.g. mobility, energy). The project is always open to additional examples of application to be published by the Federal Ministry for Economic Affairs and Energy (BMWi) online.

3.3.1.1 Industrie 4.0/SMEs, Smart Living and the finance sector

The use cases are supported by the Federation of German Industries (BDI), the Mechanical Engineering Industry Association (VDMA), and the German Electrical and Electronics Manufacturers' Association (ZVEI); these examples come from the realm of Industrie 4.0[4]; they show that companies, especially SMEs, face very similar challenges regarding data integration and assessment, and also regarding solutions to those challenges; such solutions must have a uniform language, semantics as well as modular structures. In the classical business relationships of component producers, mechanical-engineering companies and system operators, questions of how best to manage and control the data transfer and thereby secure intellectual property are of high importance.

For the finance sector, the relevant protagonists' cooperation is significant, primarily from the regulatory viewpoint. Open digital platforms and the use of AI methods enable a more efficient interplay to emerge between, for instance, the stock exchanges, the supervisory authorities, and companies.

By contrast, in the Smart Living use case, also supported by the German Electrical and Electronic Manufacturers' Association (ZVEI), the focus is rather on user-friendliness and open access to various providers' cloud-based platforms. Up to now, numerous standards and the user's need to obtain multiple authorisations on different platforms have been factors obstructing broad-based use.

4 Beyond this, the Federation of German Industries (BDI) also supports the activities directed at development of use cases for the following application areas: Smart Living, the finance sector, and healthcare, in addition to public administration and science.

EXAMPLES FROM PRACTICE

A: On the path to Industrie 4.0 – how companies can achieve trustworthy cooperation

Example from practice; current challenges faced

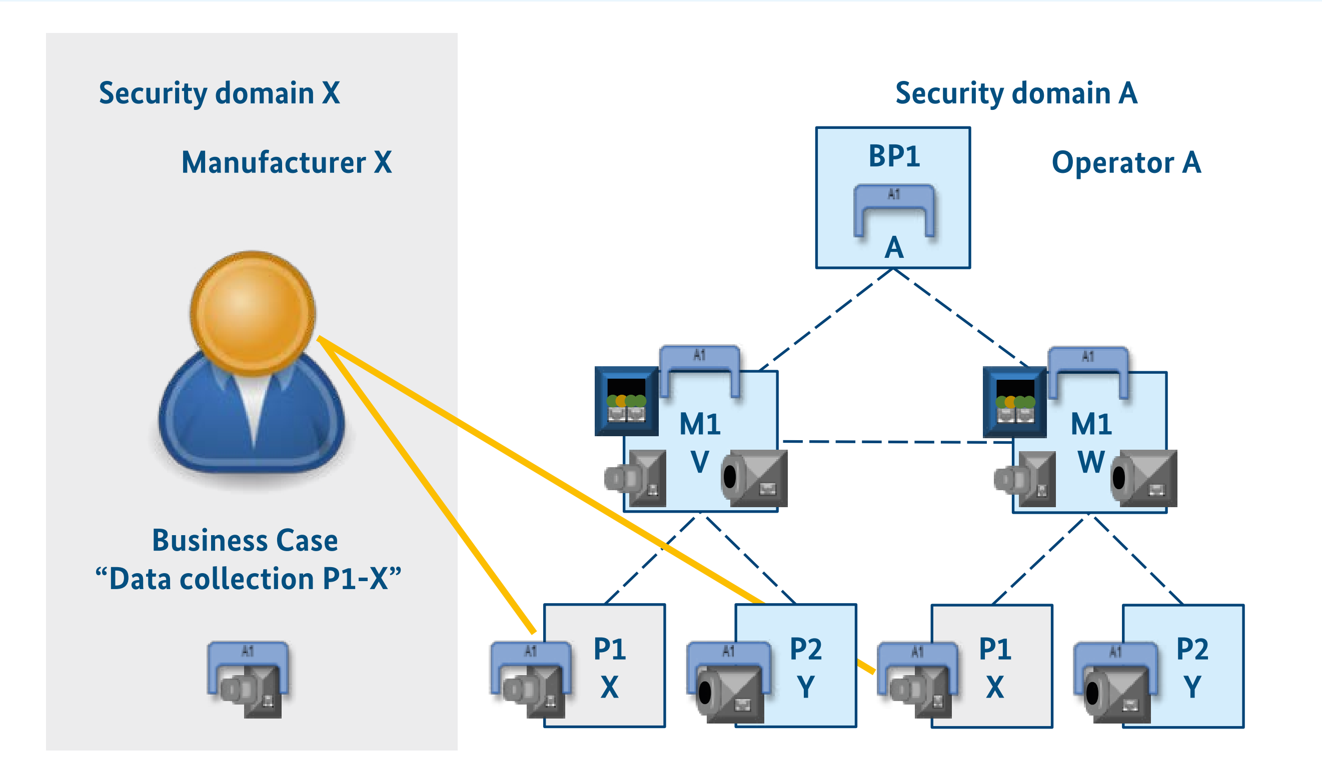
- The manufacturer of an automation component would like to gain access to operating data for the product, which is built into a machine which, in turn, is operated by a third-party company. Permanently recording the machine's latest current status enables the manufacturer, for instance, to optimise the security and efficiency both of its components and of the facility as a whole.

- This very simple and yet typical scenario of an industrial business relationship gives rise to some questions, for instance: To whom does the component manufacturer's data belong? Who may obtain access to it, and for what purpose? How can data be monetised? Are the data available in a standardised format?

- At present, these questions can be resolved bilaterally along a defined value creation chain. Yet the data relevant to the manufacturer is not available in an aggregated form across many operators; it is not possible to scale up across value creation networks.

- For the component manufacturer, the mechanical-engineering company and the operators to be able to collaborate on condition monitoring (i.e. the continuous summarising, analysis and presentation of operating data and current-status data, using sensors), they need a trustworthy infrastructure for data exchange and common rules for cross-company authentication and access management.

Figure 2: Use Case: Trustworthy Cooperation on the Path to Industrie 4.0



Source: Bosch, Plattform Industrie 4.0

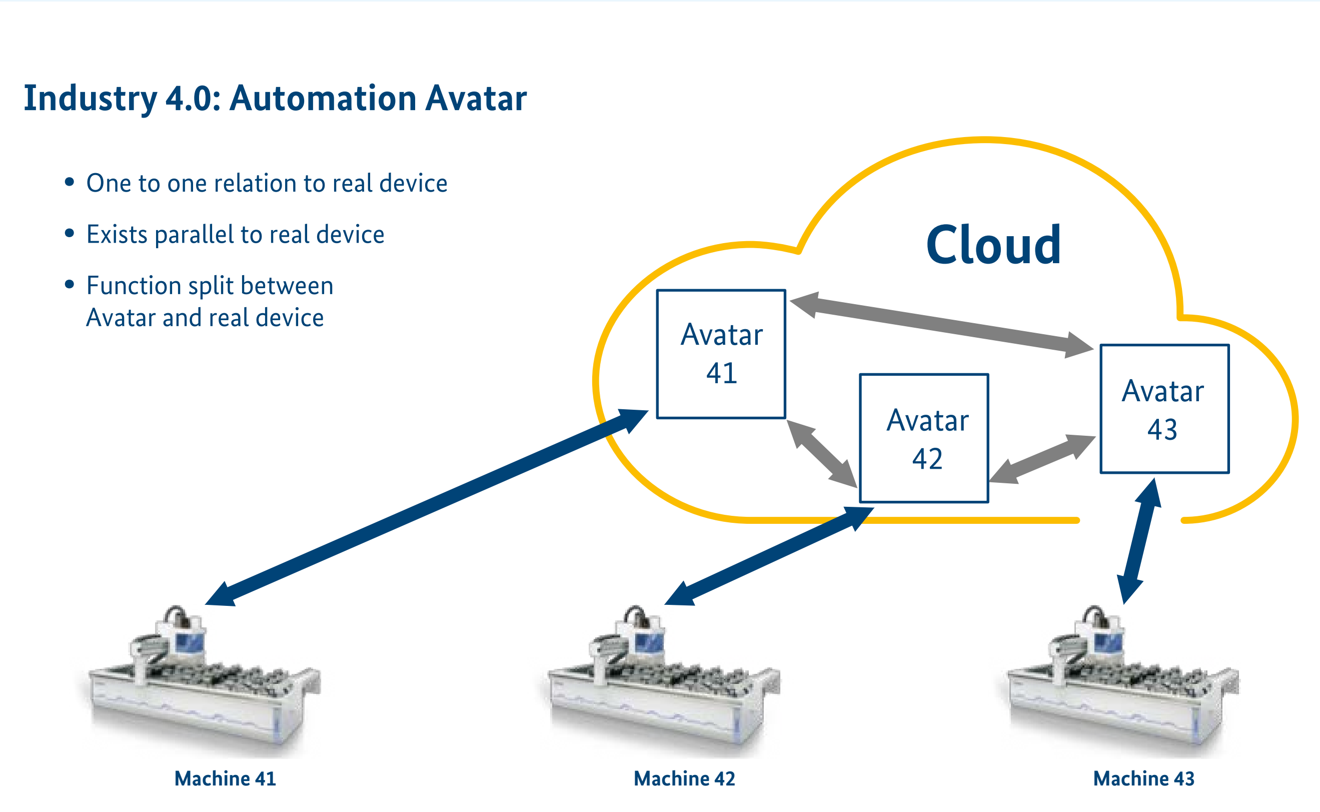
What added value does Project GAIA-X offer?

- GAIA-X nodes act as 'anchors of trust', as they enable the authentication of the partners in the value creation network and the regulation of access rights. This dispenses with laborious bilateral coordination processes. The different protagonists can communicate across the various security domains of the companies in the network, based on trust relations and a trustworthy infrastructure that enables a secure exchange of data. Each company decides for itself where its data is stored, as well as who may process it and for what purpose.
- The project can provide the foundation for a marketplace that monetises operational data in industrial value creation networks. Simultaneously, incentives can be generated for data exchange, spanning across the various parties involved.

Patron
Michael Jochem –
Bosch and
Plattform Industrie 4.0

B: Practical implementation of Industrie 4.0? A genuine feat of strength!

Figure 3: Use case: Industrie 4.0 – Automation Avatar



Source: Beckhoff Automation

Example from practice; current challenges faced

- Implementing Industrie 4.0 in practice means networking all components along the whole value creation chain, in order to thereby offer value-added services based on data: this presents major challenges, particularly for SMEs.

- Modern production facilities consist of many different machines and components; in part, these also use a variety of cloud systems. At present, the data and systems are linked-up and integrated by means of laborious project work. The lack of standardised workflows, of data availability from beginning to end and of framework contracts for data exchange are factors explaining the slow, faltering adoption of Industrie 4.0 solutions.

- Operators and manufacturers of machines and integrated facilities also present major demands with respect to data sovereignty: they want to decide for themselves where they store production data and applications that sustain know-how – for instance, in system control, 'on-edge' or private cloud instances, instead of in their customer's cloud.

- The breakthrough to across-the-board implementation of Industrie 4.0 can come only when an ecosystem is set up for turnkey use of value-added services in heterogeneous production environments.

What added value does 'Project GAIA-X' offer?

- The project creates an added value by placing itself above existing hyperscaler/cloud providers, as a type of multilateral administrative layer; this connects the production infrastructure and clouds with higher-level semantics and data exchange services; in this way it simplifies interface management.

- Specifically, the project can reduce the workload that emerges due to the need to find a bilateral individual-project solution with the machine supplier, for integrating each machine and coordinating the access on security domains. Thanks to the federated data infrastructure, interoperable standards can be used to integrate all machine suppliers efficiently, via central interfaces, taking into account the security requirements. Consequently, the workload that Industrie 4.0 projects involve can be substantially reduced.

- This means that an ecosystem can grow in a decentralised, heterogeneous and scalable form: on the one hand, it connects the various levels, consisting of computing and storing, between edge and cloud; on the other, it enables data to be stored and algorithms to be used in accordance with IP rights.

Patron
Gerd Hoppe – Beckhoff Automation
(Example of demand was coordinated with the Mechanical Engineering Industry Association *(Verband Deutscher Maschinen- und Anlagenbau e.V.)* and the German Electrical and Electronics Manufacturers' Association *(Zentralverband Elektrotechnik- und Elektroindustrie e.V.)*

C: Using synergies in supplier networks

Example from practice; current challenges faced

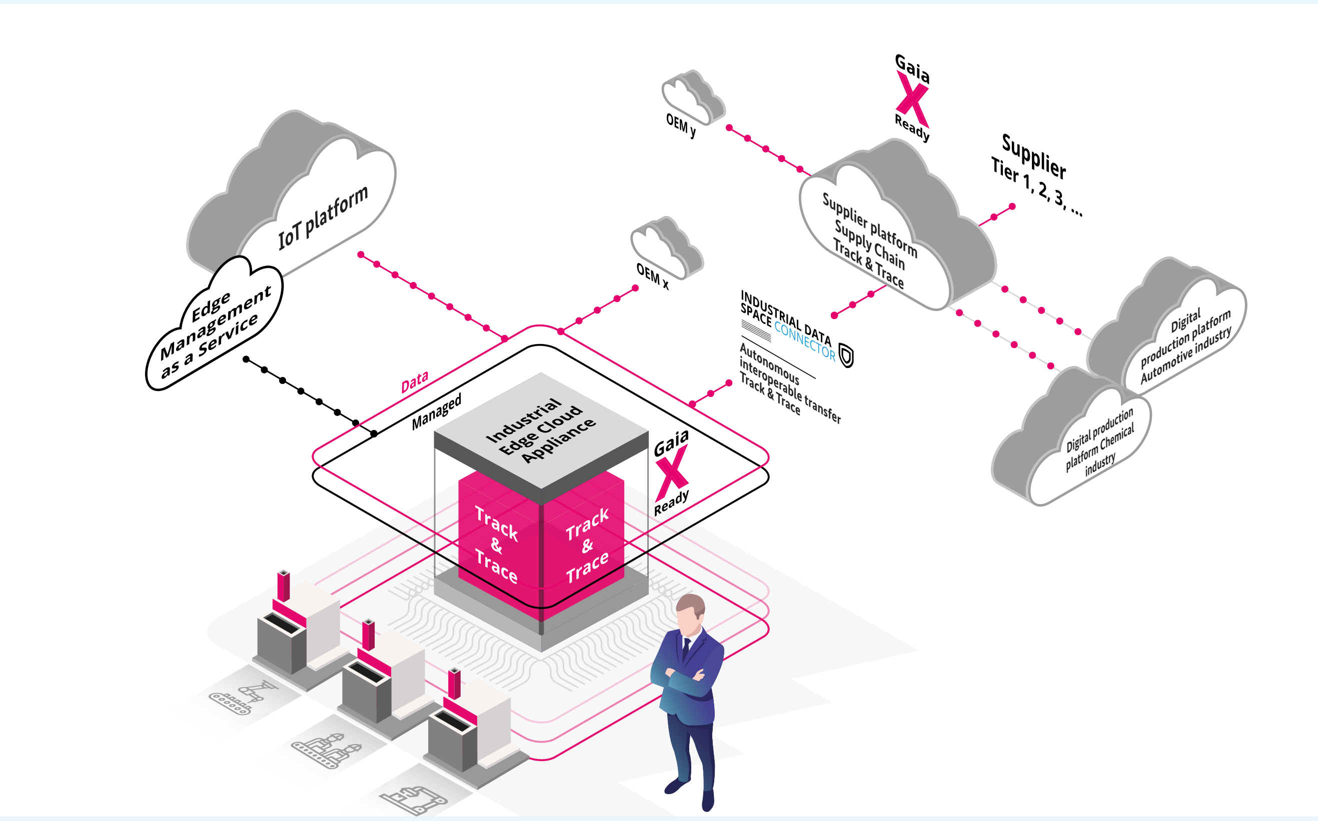
- Ever greater diversity of versions, technological discontinuities and the integration of ever new suppliers into the production networks: these aspects characterise the manufacturing industries' digital transition.

- Simultaneously, there are increasing requirements on transparency and the ability to control supply chains, reaching right into an individual company's own facility. Securing the supply of components is a big challenge, not only in times of international trade conflicts. In cases of product call-back, the decisive issue is to know which component was installed in which production batch. It is also a matter of improving the early recognition of production-series errors in the field (track and trace).

- The biggest challenge in this is to provide, and to link-up, data from non-homogeneous IT systems, obtained from different stakeholders, while safeguarding data sovereignty along the whole chain of production and supply. A complete picture of the data is essential in order to be able to unambiguously identify disruptions to operation.

- Up to now, the cross-company exchange and link-up of heterogeneous data in sector-specific solutions is primarily driven on the bilateral level. Common rules for participation in the ecosystem and collaboration could substantially reduce the workload and the obstacles to participation, especially for SMEs. In this way, new business models can emerge and synergies in the value creation network can be used even better.

Figure 4: Use Case: Synergies in Supplier Networks



Source: SupplyOn, German Edge Cloud, IoTOS, IDS Association

What added value does 'Project GAIA-X' offer?

- The project makes it easier to provide clear and readily comprehensible requirements for participation in the ecosystem, or cooperation in it; this reduces the bilateral workload of coordination between interested companies.

- Through one-off 'qualification' as a GAIA-X node, protagonists can make their own data available to a potentially larger circle of interested parties, in order to improve their own products or to develop digital services.

- The project makes it easier to selectively pass data on, thereby strengthening data sovereignty.

- Standardisation rules and uniform semantics mean that data sets can be linked-up with one another better. Typical application cases, such as traceability in pre-products, can be better put into practice - also as part of uniform identity management. At the same time, the implementation workload associated with track-and-trace solutions across company boundaries can be substantially reduced.

- The project can serve as an integrator. An overlapping, open ecosystem emerges, one that gives impetus to data exchange and stimulates new business models.

Patrons
Markus Quicken – SupplyOn
Sebastian Ritz – German Edge Cloud
Dieter Meuser – IoTOS
Lars Nagel – IDS Association

D: Secure and multi-functional cloud environment for the housing industry, to generate Smart Living solutions with demanding latency requirements

Example from practice; current challenges faced

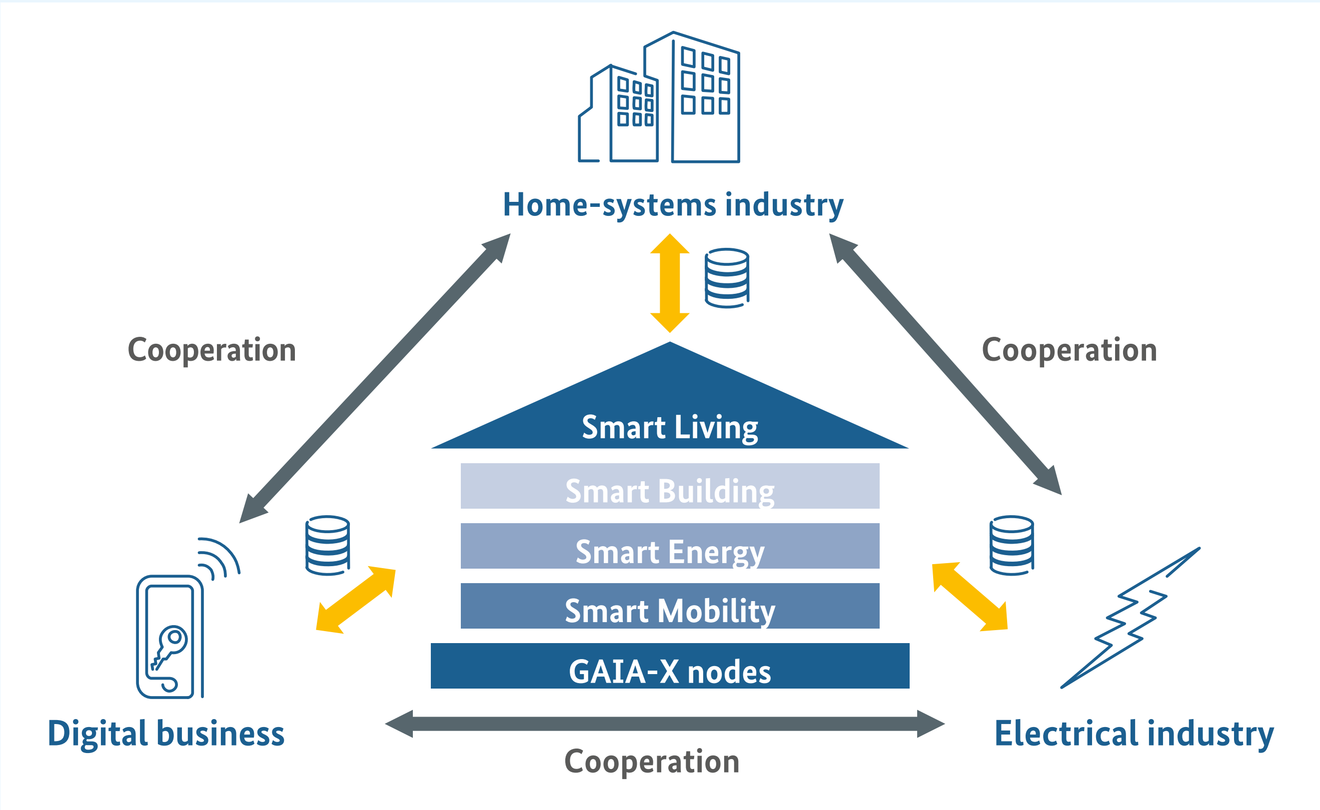
- The ecosystem centred around Smart Living is developing into an attractive market. 23 million rented homes in Germany alone could be equipped with automation solutions and intelligent devices.

- These devices need to be able to interact seamlessly. A precondition for AI-based business models is end-to-end data recording, processing and networking (in the figure: see 'GAIA-X nodes'). A suitable cloud environment is needed for this.

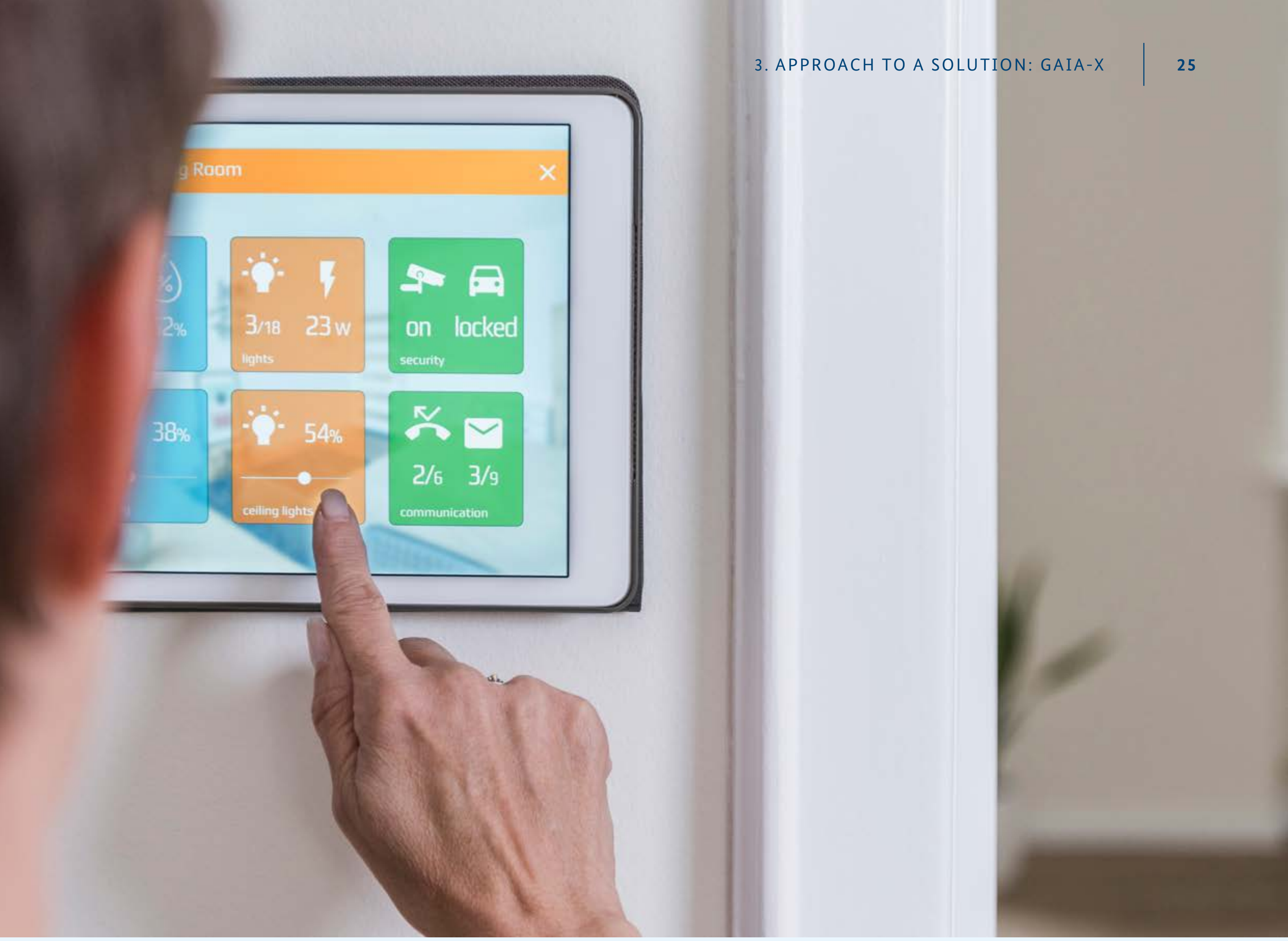
- Many companies in the housing industry want to store and process customers' data exclusively in cloud environments located in Europe (GDPR area).

- In addition, intelligent services are sustained by the link-up with data from other related systems, such as Smart Energy or Smart Mobility in AI-based applications. This data must be structured in a meaningful way, securely and reliably; this is not yet the case.

Figure 5: Use Case: Smart Living



Source: ZVEI, for the consortium "Smart Living"



- To scale-up Smart Living, a secure, scalable European cloud environment with strong performance is needed; it must also include local edge devices, to avoid excessively long latency.

What added value does 'Project GAIA-X' offer?

- The project offers the housing industry simple and secure access to a multi-functional cloud environment in the GDPR area (e.g. data storage, machine learning, semantic description, pattern recognition, predictive actions).
- The project sets up networks of regional and functionally specialised computer centres, also thereby supporting the scaling-up of Smart Living applications; the latter depend on edge computing because of the high level of latency-related requirements.
- The project makes it easier to introduce suitable standardisation rules for linking-up the growing data quantities; it thereby fosters the emergence of further AI applications, especially through the collaboration between the digital economy, the housing industry and the electrical industry.

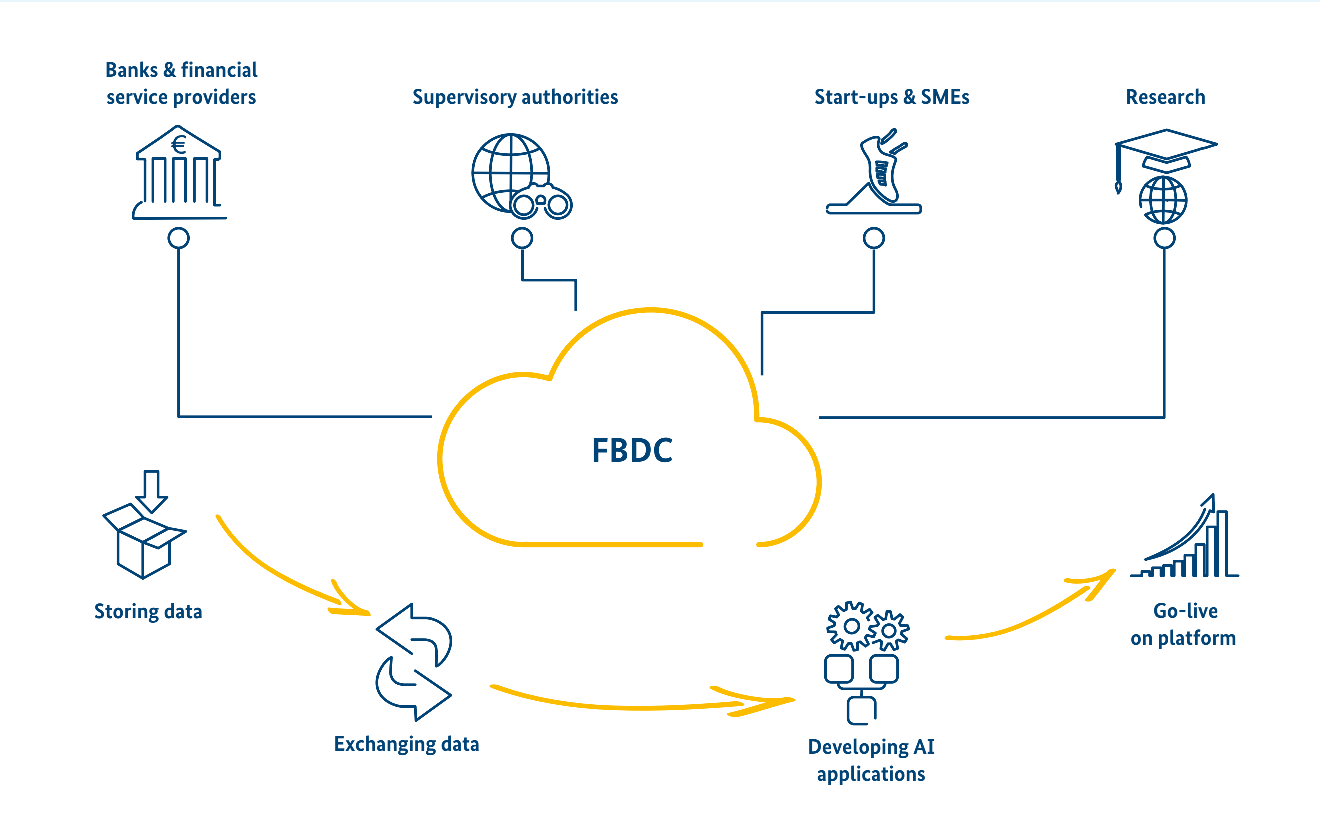
Patrons
Platform for context-sensitive, intelligent and predictive Smart Living services – ForeSight
Anke Hüneburg and Jochen Schäfer – German Electrical and Electronic Manufacturers' Association (ZVEI), for the 'Smart Living' consortium, funded in the context of the following innovation competition: 'Artificial intelligence as a driver for ecosystems relevant to the overall economy', commissioned by the Federal Ministry for Economic Affairs and Energy (BMWi)
Dr. Hilko Hoffmann – DFKI
Kerstin Bergmann – Bosch
Thomas Feld – Strategion

E: Financial Big Data Cloud – strengthening the German and European financial marketplace

Example from practice; current challenges faced

- Big Data and AI act as the finance sector's central drivers of innovation. They are not only important research areas; they are also already shaping the business models of banks and stock exchanges.
- Likewise, supervisory bodies for financial markets will use AI applications in the future, for instance to take more effective action against money-laundering and market manipulation.
- Substantial players in this sector are drawing together to form a Financial Big Data Cloud (FBDC), in order to build up a cloud-based data platform for the finance sector. The aim for this platform is to integrate the finance data of companies, public authorities and research institutions, which are not yet linked up, in a common data pool; the platform is also to be optimised to suit the development of AI applications and systems respectively.
- The IT infrastructure, serving as the foundation, needs a secure treasury of data – one that satisfies the finance sector's demanding statutory and regulatory requirements. The data is to be made available to the various user-groups in different phases, depending on its level of sensitivity.
- The platform should also make analysis tools, data-exchange tools and computing capacities available.

Figure 6: Use Case: Financial Big Data Cloud



Source: Ministry of Economics, Energy, Transport and Housing, State of Hessen, Deutsche Börse, Deutsche Bundesbank and TechQuartier

What added value does 'Project GAIA-X' offer?

- The project can serve the FBDC in the context of a multi-cloud approach, as an alternative infrastructure platform, one that offers improved possibilities for developing and using AI applications.
- The project raises the level of transparency in the cloud market, thereby making an important contribution to data sovereignty. users can obtain a targeted overview of the various offerings and the related conditions (e.g.: where is the data hosted?). This is especially important for stakeholders in the finance sector. They are subject to – at least in part – strict rulings on where the storage of data is allowed and how it may be used.
- The project thereby provides further participants with easier access to the FBDC; in that way, it can enlarge the target group and the reach of the developing ecosystem that is based around financial data. There is also particular potential in linking up the FBDC ecosystem with additional ecosystems specific to other business sectors, for example in industry.

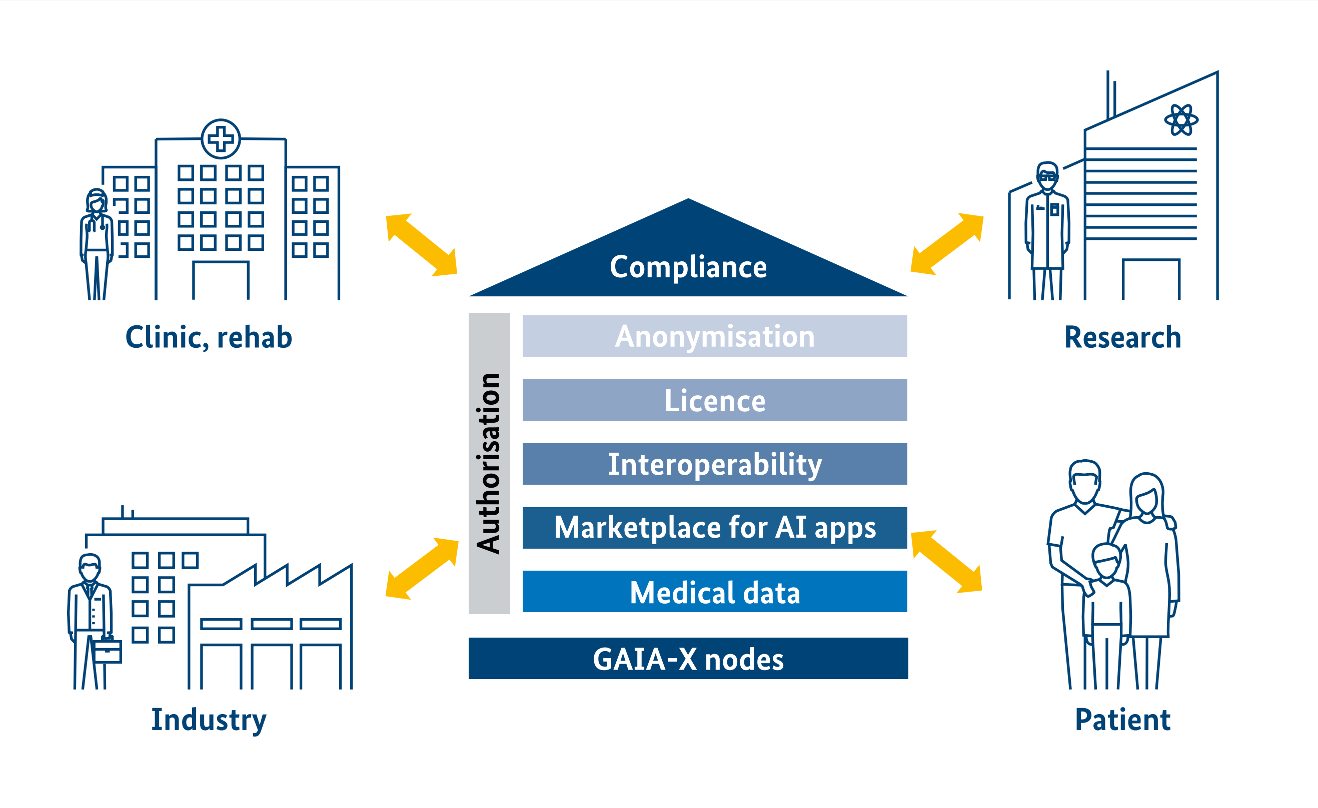
Patrons
Dr. Stephan Bredt – Ministry of Economics, Energy, Transport and Housing, State of Hessen
Konrad Sippel – Deutsche Börse
Prof. Stefan Bender – Deutsche Bundesbank
Dr. Kevin Bauer – TechQuartier

3.3.1.2 Healthcare

The health sector is regulated rigorously and operates on a highly decentralised basis. All use cases shown here focus on the use of AI in healthcare. Development of the appropriate solutions entails a major workload in integrating the requirements of the various people and organisations involved. Simultaneously, dealing with patient data imposes exacting data-protection requirements. The Federal Ministry of Health (BMG) is pressing forward on digital solutions and AI in healthcare. The basis of such solutions includes (among other things) the introduction of the telematics infrastructure, including the electronic patient file. The BMG supports the use cases in a fundamentally positive way; it is receptive to the use of the offer to be developed within the project framework. Of particular relevance in this context is the ability to connect to the telematics infrastructure. For this purpose, organisations and people involved can take advantage of the current experience of gematik GmbH.

F: Artificial intelligence for clinical studies

Figure 7: Use Case: Artificial Intelligence for Clinical Studies



Source: Raylytic, for the consortium "Artificial Intelligence for Clinical Studies"

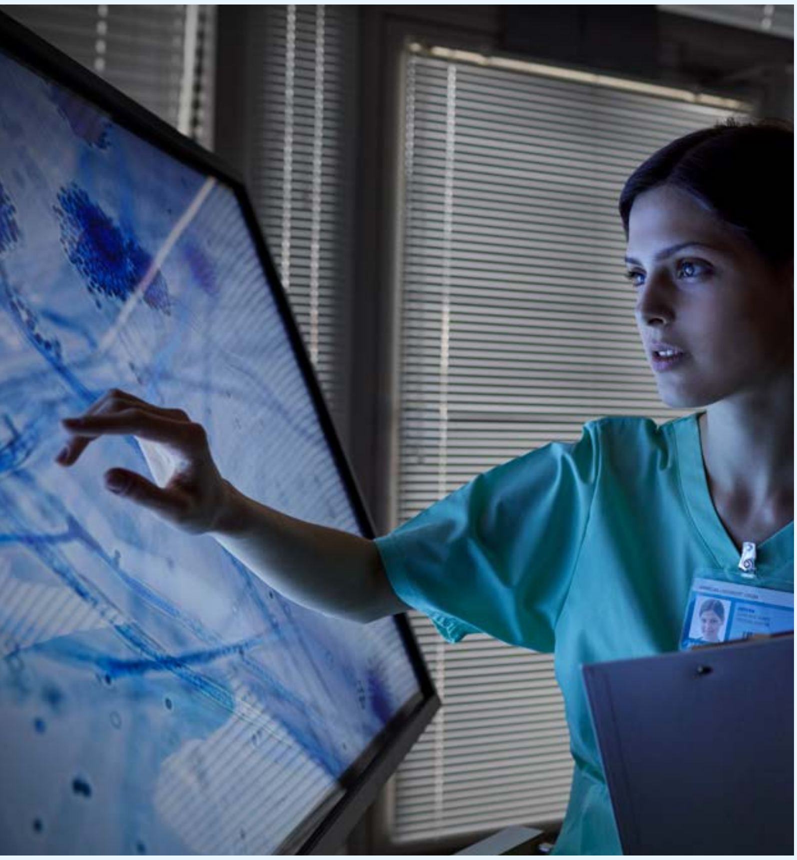
Example from practice; current challenges faced

- Starting in May 2020, EU law will require manufacturers of medical technology to make data available on product performance and safety throughout the whole product life cycle. This data is obtained in clinics using the products. In the task of automatically extracting and assessing the data, AI methods can render support.
- The continuous recording of data can benefit many health sector stakeholders: the product manufacturers, doctors, health insurance companies, state institutions for market monitoring, and also research establishments in the fields of medicine and health provision.
- At present, a large portion of clinical data cannot be used for statistical or scientific purposes, due to a variety of data formats and the unstructured form of the data. As such, it does not contribute to the creation of new medical knowledge.
- Regarding health data use, numerous regulations, norms and legal requirements in the national and international contexts must be met; these require demanding technical measures which obstruct ease of data use.
- There are misgivings in this sector about entrusting health data to cloud-service providers with closed-source systems.

What added value does 'Project GAIA-X' offer?

- The project can provide the basis for a reliable system for health data use while providing technical implementation of international provisions on legal security, data security and cyber security. The provision of storage space and computer capacity in a secure framework means that, even in areas of sensitive activity, clinics and SMEs can benefit from the economies of scale that apply to cloud-based services.

- The project can include functions for standardised anonymisation and pseudonymisation, as well as for data classification; these are needed for use by various user groups to be compliant with the law, e.g. for training AI models (so-called 'secondary use').



- The project could make standardised interfaces available; using these, various user-groups and healthcare service providers can access a cloud environment that merits trust.

- Networked solutions from a variety of project partners can emerge from bilateral individual projects; these solutions enable use of the data that renders use of AI easier, thereby fostering synergies.

- Using modular, distributed solutions it is possible to separate data processing and data hosting. Anonymised data can be brought together in a federated structure, where the data is necessary for an analysis. In this way, highly sensitive data can remain in its location, for instance in the hospital, whilst other data can be exchanged for purposes of processing and analysis.

Patron
Frank Trautwein – Raylytic, for the consortium 'Artificial intelligence for clinical studies' (KIKS), funded in the context of the innovation competition 'Artificial intelligence as a driver for ecosystems relevant to the overall economy', commissioned by the Federal Ministry for Economic Affairs and Energy (BMWi)

G: AI-based e-triage in the emergency room

Example from practice; current challenges faced

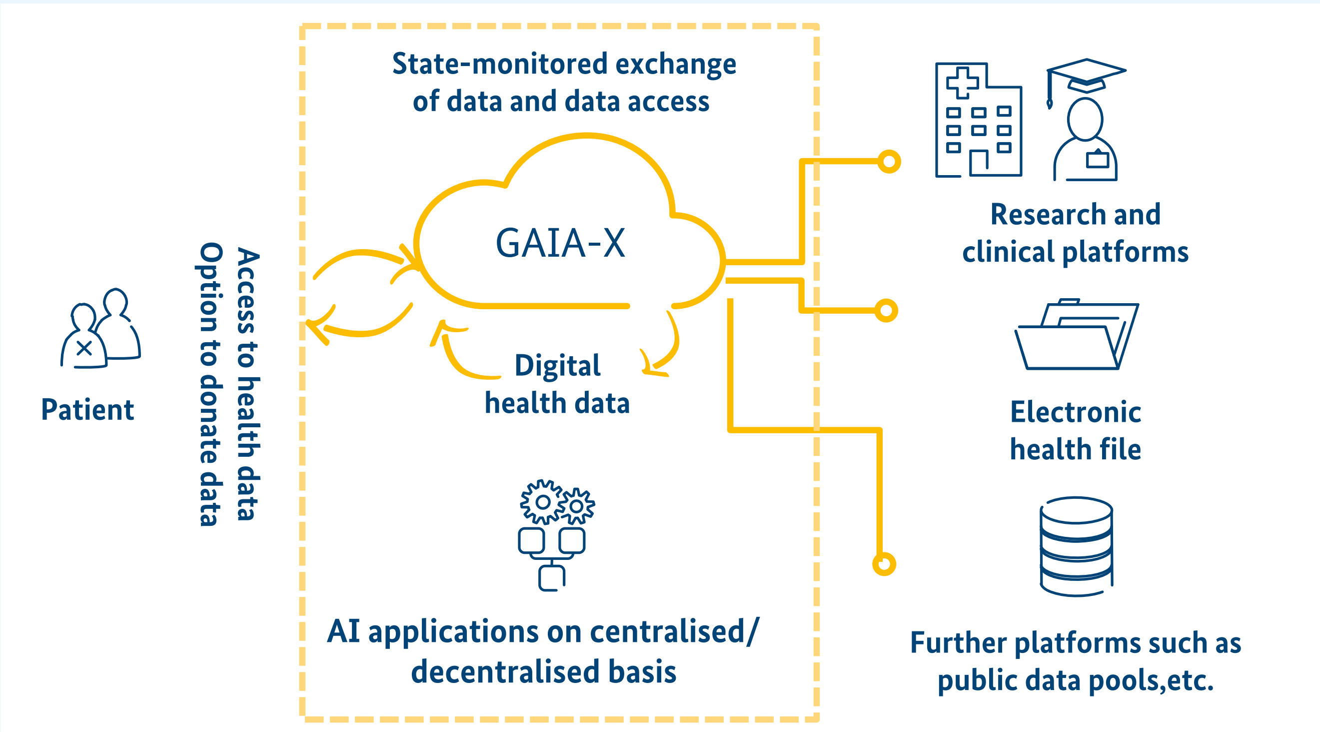
- German emergency room facilities are burdened by many non-urgent cases. From over 10 million emergency cases annually, 3.5 million are not emergency cases, which leaves less time and capacity for genuine emergencies. Therefore, risk is high that doctors take wrong decisions, e.g., that important treatments are initiated too late.

- AI-based assistance systems could tackle this problem: using e-triage, they can improve the direction of patients, starting from arrival at the hospital through to treatment, thereby supporting medical staff at an early stage of treatment. Among other tasks, the assistance system assesses data that patients themselves already submitted in the waiting area – supplemented by further findings from preliminary examinations.

- The precondition for developing these AI applications and others in healthcare is access to high-quality, real world data. These data should be provided via a digital infrastructure that guarantees patient protection and satisfies the most exacting data-protection and security requirements.

- With the electronic patient record and the telematics infrastructure, an appropriate platform is already in the pipeline for German healthcare. University clinic research data is also being rendered usable via the medical information science initiative. The intended federated data infrastructure would link data from medical research and care within a secure and anonymous environment, enabling medical progress and innovative applications such as e-triage.

Figure 8: AI-based Triage in the Emergency Room



Source: Charité – University Medicine Berlin, Learning Systems Platform

What added value does 'Project GAIA-X' offer?

- For e-triage, the various healthcare stakeholders (patients, doctors' practices, hospitals, health insurance companies, etc.) can access anonymised real data (among other sources, from the personal patient record, case files, wearables, research data from clinical studies or from various registers) via a trusted organisational unit, based on the federated data infrastructure. In addition, they can use this data in a highly secure data area (data on patients must be stored securely in Germany).

- By securely linking various data sources, health research generates very specific added value for society. This establishes an important basis for innovation and enables new business models to emerge in a highly regulated and sensitive environment, e.g. special e-triage apps.

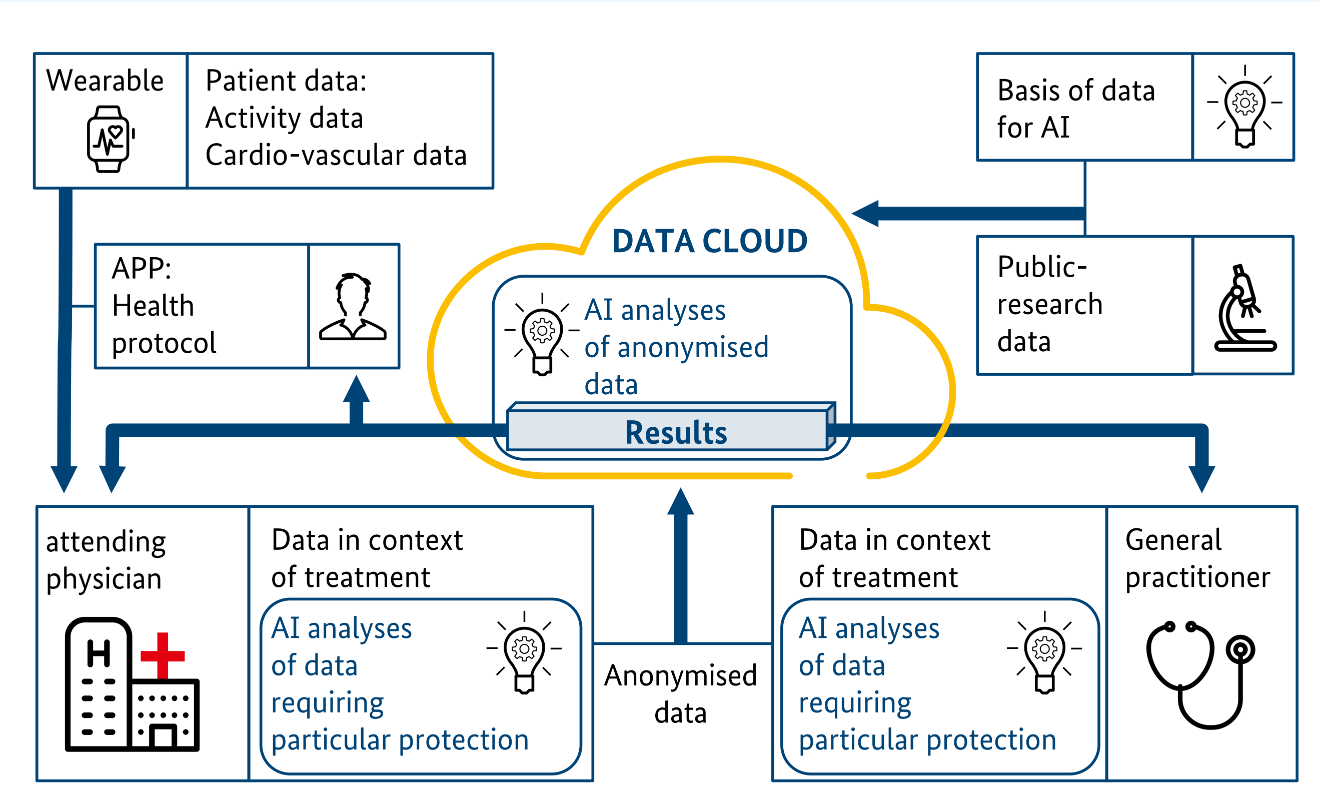
- The project supports the use of a database organised on a decentralised basis, thereby contributing to data sovereignty. By applying distributed-learning methods – so-called 'Federated Learning' – sensitive health data can be processed on the spot, in compliance with the law, e.g. directly in the hospital. A secure, federated and multi-functional cloud environment also enables the use of latest methods in AI analysis.

- Likewise, the project acts as the technical basis for an individual data donation; this means that, on a self-determined basis, citizens can make their own health data available. Furthermore, they can decide who is allowed to use the data and who is not.

Patrons
Prof. Dr. Klemens Budde – Charité – University Medicine Berlin and Plattform Lernende Systeme (Germany's Platform for Artificial Intelligence)
Dr. Thomas Schmidt – acatech – National Academy of Science and Engineering and Plattform Lernende System (Germany's Platform for Artificial Intelligence)

H: Better preventive healthcare with 'Smart Wearables' – how we can learn from data

Figure 9: Use Case: Smart Wearables



Source: Berlin Institute of Health, Charité – University Medicine Berlin

Example from practice; current challenges faced

- The recording of health data is undergoing a fundamental paradigm shift: as well as the customary data-gathering by doctors and clinics, typically prompted by specific events, data can increasingly be gathered and assessed continually via the patients' 'Smart Wearables'. Based on these large and revealing quantities of data, AI algorithms can be developed and refined extremely well to recognise critical situations at an earlier stage than before, for instance with regard to heart attack patients.

- In some instances, this requires hundreds of thousands of learning data sets of authenticated quality. Germany has not yet succeeded in gathering a sufficient quantity of data from various sources (e.g. hospitals, doctors' practices, wearables) to attain a critical mass for the applications described.

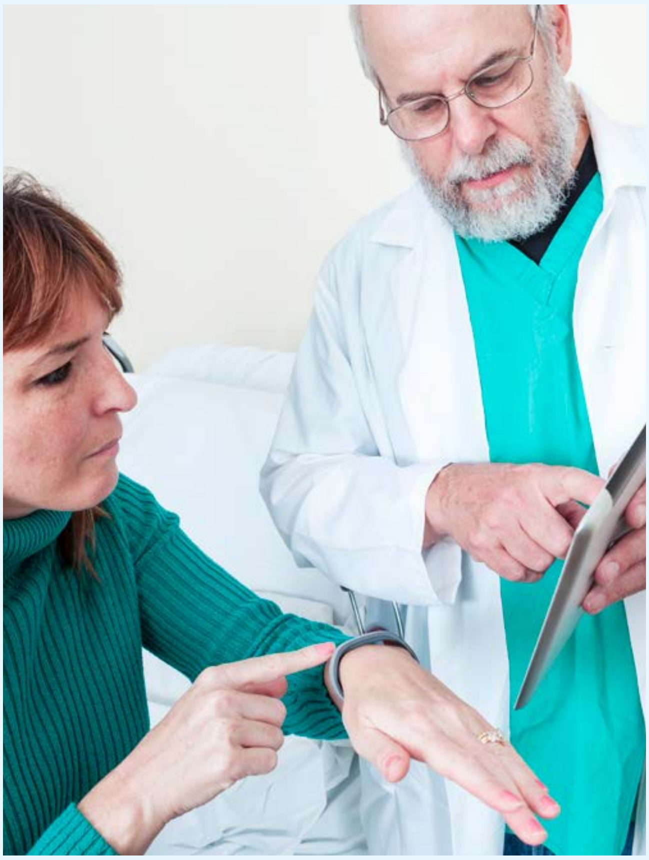
- It is not before the data areas are federated that the integrative analysis becomes possible. The clinical doctor and general practitioner can make use of the assessments; the patient can gain access to their own health data and analysis results.

- For particularly sensitive data, data storage and data processing are required to take place in Germany or within another EU country – in some cases, it must be done by German service-providers.

What added value does 'Project GAIA-X' offer?

- The project sets up the basis for developing and using standardised interfaces and suitable semantics for being able to link-up and assess the data of as many involved parties as possible, in a targeted way.

- Rigorous and binding data classifications and also centralised checks on the ecosystem participants make it easier to use the health-system data in compliance with the GDPR. The standards of federated data infrastructure can serve as anchors of trust.



- The federated data infrastructure also supports the health sector's particular demand for linked-up edge/cloud solutions: for instance, for data protection reasons, sensitive data supplied by a smart wearable must be able to be aggregated and assessed in the clinic (Edge). By contrast, the assessment results must be able to be brought together in the cloud, without needing to share particular data.

- The integration of cloud-based storage resources in Germany or in another EU country, as GAIA-X nodes, allows the storage of sensitive data.

- The openness and resulting flexibility of the project make it possible to connect existing (data) platforms to further research and health domains as well as international initiatives.

Patrons
Christian Lawerenz and Prof. Dr. Roland Eils – Berlin Institute of Health und Charité – University Medicine Berlin

3.3.1.3 Public administration and science

Similarly, public administration and science are important future application areas for cloud solutions and cloud-related services.

Particular requirements from the perspective of German federal administration: a current study commissioned by the Federal Ministry of the Interior, Building and Community (BMI) shows that, at present, public administration is highly dependent on a small number of foreign providers of software. Future solutions must both fulfil rigorous information-security requirements and provide legal certainty within the framework of the GDPR, in order to retain the digital sovereignty of the public administration. In addition, the Online Access Act (OZG) requires the Federal Government, the States *(Länder)* and municipalities to offer their administrative services online by the end of 2022, and to form a linked-up alliance of their administrative portals. A prerequisite for this digital transition is a high-performance cloud infrastructure, guaranteeing a high degree of digital sovereignty.

The Federal Ministry of the Interior, Building and Community is receptive to the use of the offering to be developed within this project's framework. On the basis of Project GAIA-X, the Ministry is currently reviewing scenarios that could be implemented in the future, under the following framework IT security conditions.

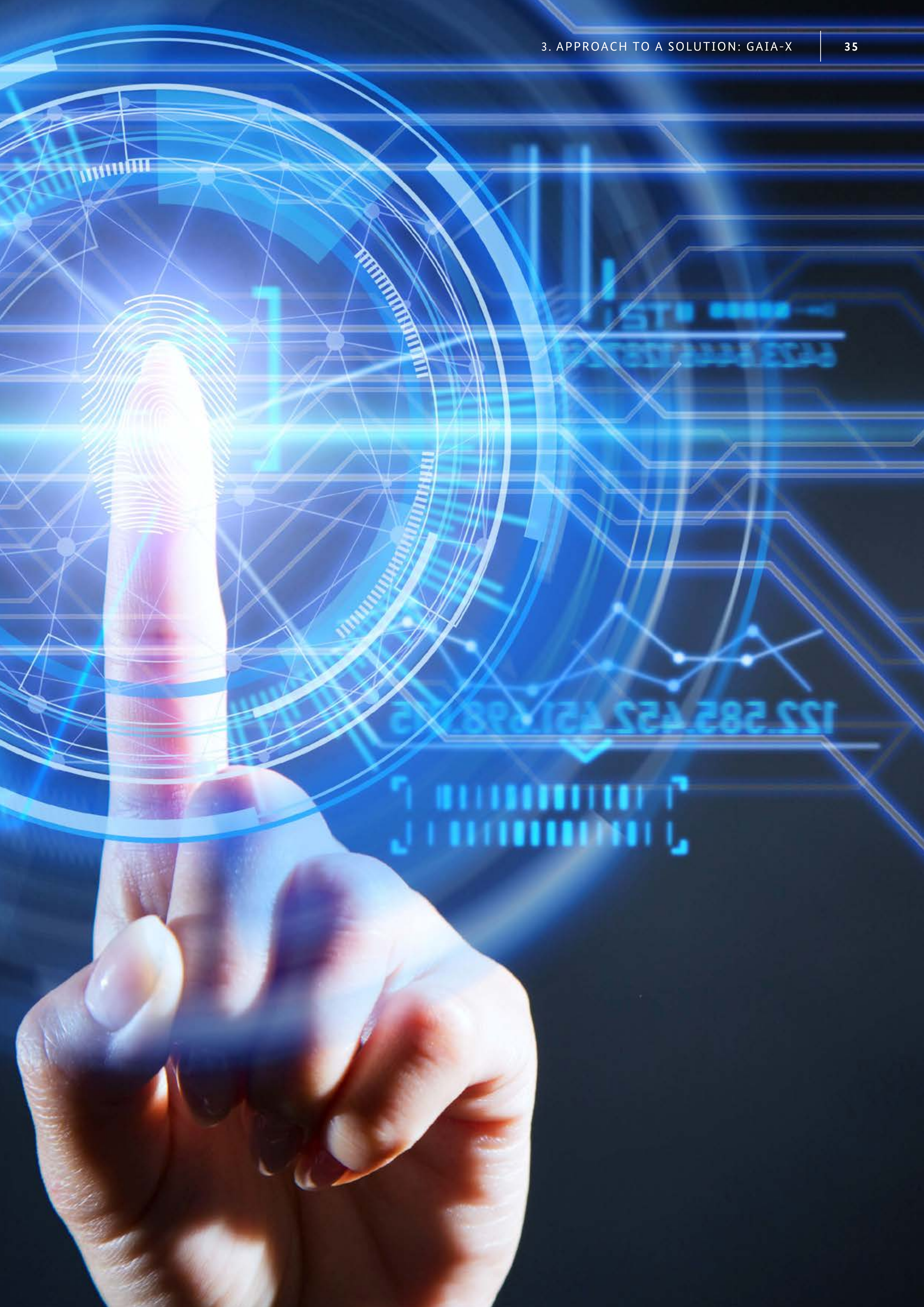
1. There must be compliance with the regulations drawn up by the Federal Office for Information Security (BSI), especially the 'minimum standards for use of external cloud services'. GAIA-X cloud-service providers thereby grant the Federal Office for Information Security (BSI) a right to check and audit.

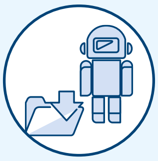
2. GAIA-X cloud-service providers fulfil the 'Federal Office for Information Security (BSI) Catalogue of Requirements on Cloud Computing (C5),' including basic and additional criteria, and provide proof of this by a suitable attestation.

3. Practices adhere to and implement the relevant modules of IT basic protection.

4. For certain application cases, ISO 27001 certification is recommended (based on IT basic protection). Whether this is necessary must be checked in the individual case.

The state is a player in the digital world, not only as a regulator, but also as a user. For applications in public administration, security, reliability, trust and transparency play a decisive role. Applications in science primarily need to have an infrastructure of high-performance capability and broad data availability.



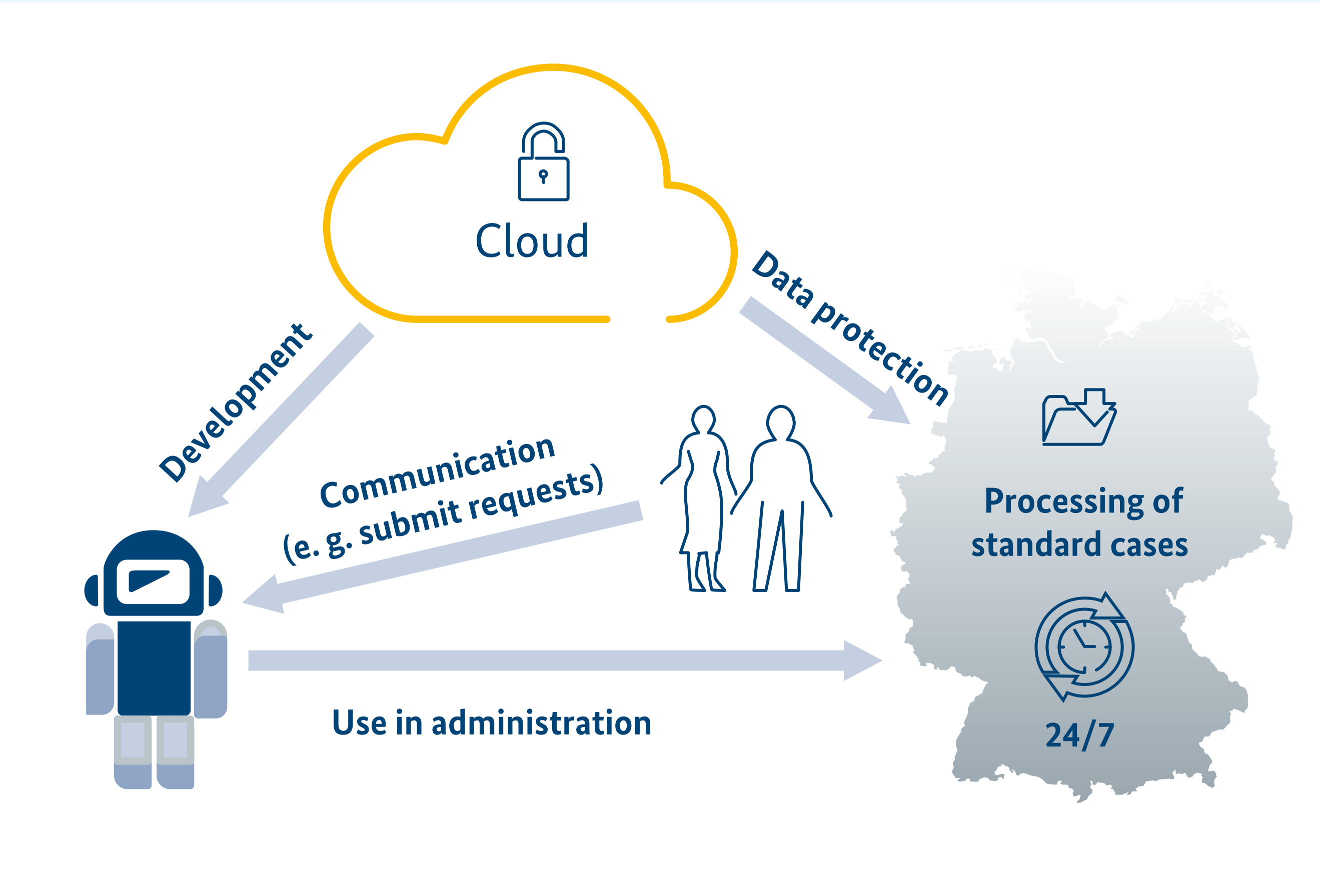


I: Citizens' service around the clock: the chatbot in public administration

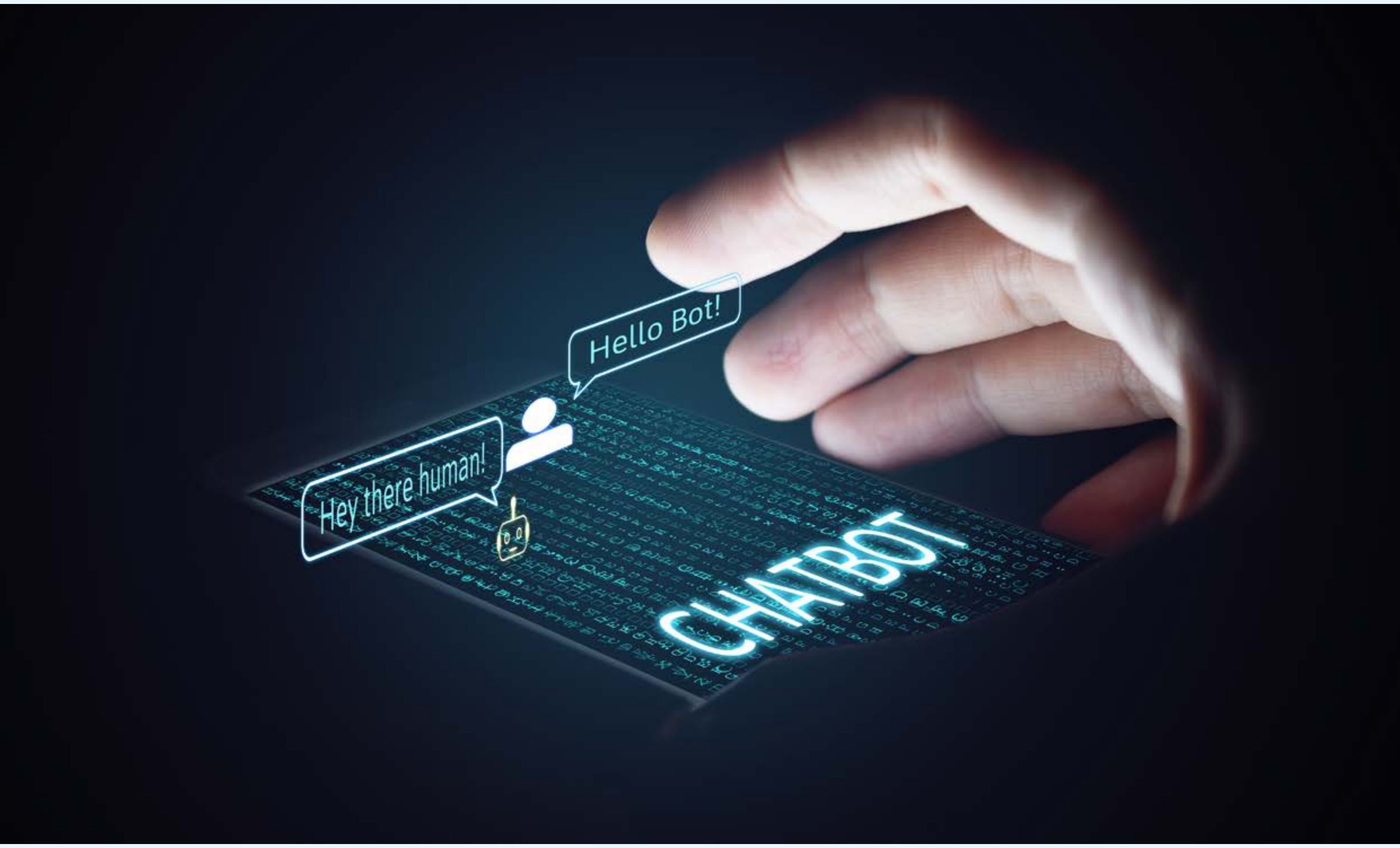
Example from practice; current challenges faced

- Chatbots are digital, text-based or spoken-word-based conversation partners. In the future, their role is to open up easy access to public administration services for the public, around the clock, seven days a week. Primarily, their task is to process standard cases, thereby also reducing the burden on personnel in public offices and other administrative institutions. This means that applications can be processed more quickly. In turn, this makes it possible to focus staff working time on more demanding cases.
- High-calibre language-recognition and text-recognition services are needed to develop chatbots; at present these are primarily being offered by US hyperscalers.
- Public administration is subject to strict requirements on data protection and confidentiality when using cloud services. For instance, citizens' data is not permitted to be stored or administered outside the infrastructure of the relevant State *(Land)* or the country as a whole. Public administration has sovereignty over access to the data; access to it must remain blocked off from the service providers.

Figure 10: Use Case: Chatbot in Public Administration



Source: Dataport



What added value does 'Project GAIA-X' offer?

- The project can support the deployment of chatbots by creating the necessary basis for a legally compliant cloud environment; that environment develops and uses chatbots in administration in which functional AI building blocks can be modularly linked up (in the context of 'function as a service') with the public authority's infrastructure for data storage and processing.

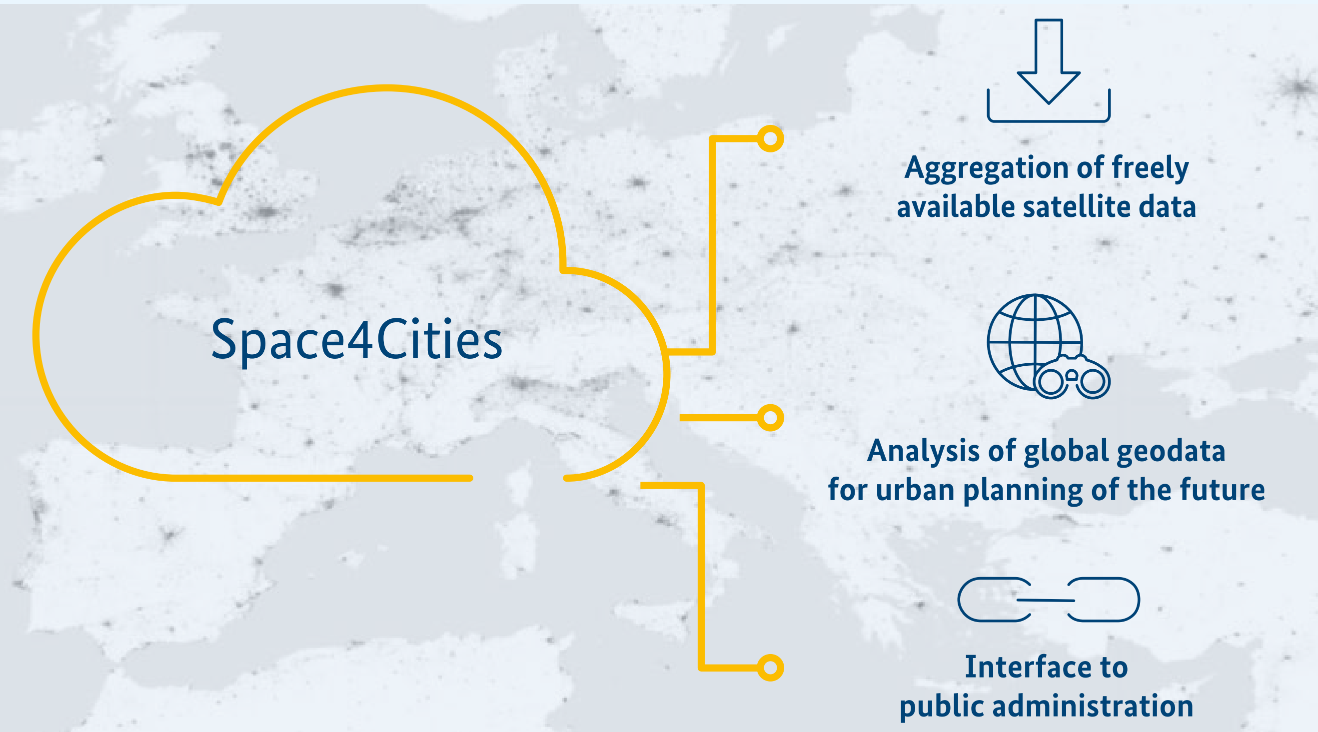
Patrons
Dr. Marianne Wulff and Dr. Derek Meier – Dataport

J: Shaping sustainable cities, using Big Data from space

Example from practice; current challenges faced

- The world population is growing; climate change is advancing, while social and demographic developments demand answers.
- By the year 2030, 70 per cent of the global population will live in urban areas. Satellite-based observation of the earth is an important source of information for new solutions to the sustainable development of cities and spaces.
- Thanks to freely available satellite data, it is possible in principle to record the dynamics of global urbanisation.
- And yet, for their analyses and for processing of the vast data quantities, German research establishments depend on US commercial cloud-service providers.
- This dependency carries risks. If the providers were to discontinue their services, the research would lose high-performance data management and specific tools for efficient processing of data on observations of the earth. Research institutions are hardly able to protect their intellectual property in such a situation: their data analysis algorithms are on the cloud provider's servers.
- To be able to offer specific, effective solutions for urban developers and decision-makers, it is often necessary to combine and assess the combination of satellite information and public administration data. Yet access proves to be difficult due to regulatory obstacles regarding data storage. There is also a lack of suitable interfaces.

Figure 11: Use Case: Space4Cities



Source: DFD, DLR

What added value does 'Project GAIA-X' offer?

- The project cannot solve the problem of necessary data storage capacities; however, it can render support to the link-up of various cloud participants within a common ecosystem, one in which a secure and standardised exchange of (European) data, algorithms, functionalities and results can take place.
- The real added value is found in the synergised assessment of stocks of data from earth observation and from public administration; this is done to break up data silos and to offer, in compliance with the GDPR, tailor-made information for urban development and/or for new digital products and business models (e.g. in the realm of the Sharing Economy or of local public transport). This contributes to sustainable solutions for future urban development.
- Ideally, the ecosystem could generate alternatives which also make data storage possible with European service providers, thereby safeguarding redundant access to particularly relevant research data. This could also reduce the risk of loss of intellectual property.

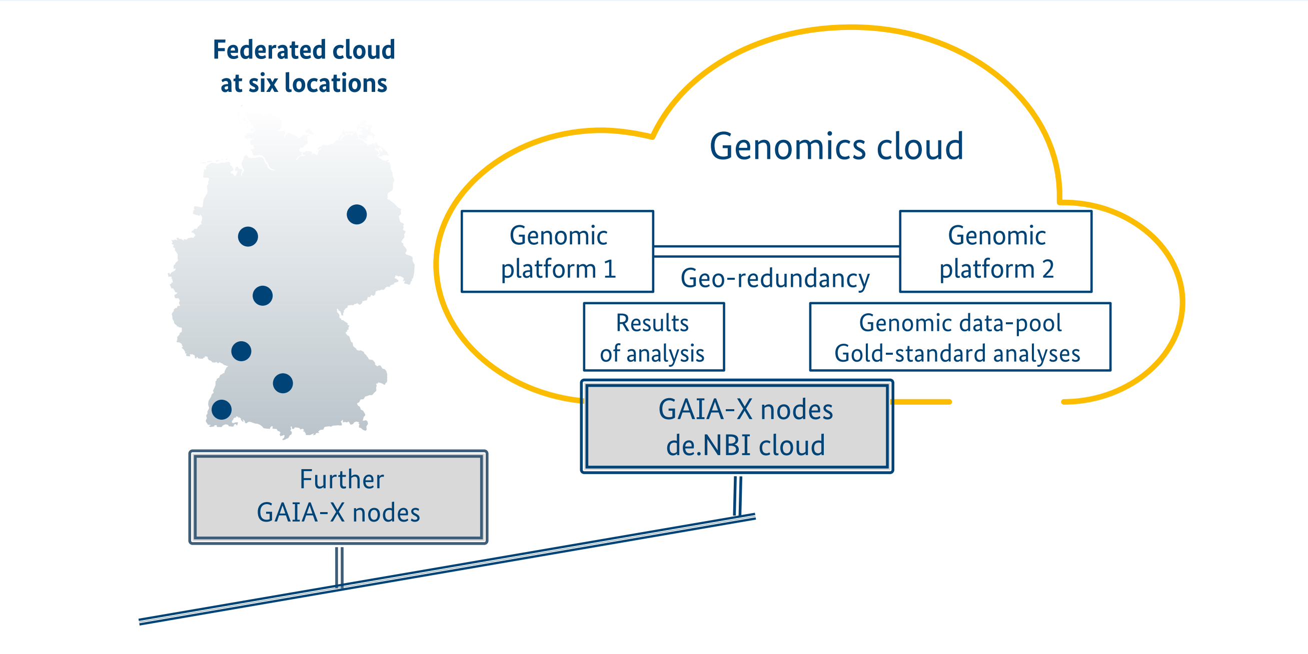
Patrons
Dr. Thomas Esch and Julian Zeidler –
German Remote Sensing Data Center (DFD)
German Aerospace Center (DLR)

K: Defeating cancer – a research cloud for genome data

Example from practice; current challenges faced

- Biomedical research has developed into a data-intensive science: in a few days, genome sequencing generates project data in the terabyte range.
- Research depends on a technical infrastructure that both enables large data quantities to be stored securely and also makes available a high-performance computation architecture for the laborious analysis of data in the petabyte range.
- To be able to better predict the emergence of cancer in the future and to support the development of new treatment methods on the basis of data, the German Cancer Research Center in Heidelberg and the Berlin Institute of Health/Charité are currently establishing a cloud platform for storage and analysis of genome data.
- To do so, this research platform uses the cloud from Germany's national bio-information sciences initiative (de.NBI), funded by the Federal Ministry of Education and Research (BMBF). This cloud offers a federated and academic infrastructure for German life-science professionals.
- One challenge will be to connect the platform to further research and health domains, via cloud and edge technologies, and to integrate the project into international initiatives.

Figure 12: Use Case: A Research Cloud for Genome Data



Source: Berlin Institute of Health, Charité – University Medicine Berlin, German Cancer Research Center

What added value does 'Project GAIA-X' offer?

- The project creates the opportunity for secure, GDPR-compliant access to various protagonists' data in healthcare, by means of centralised checks conducted on the corresponding GAIA-X nodes.

- Providers of high-performance (infrastructure) components, and also of effective functions of calculation and analysis, can be integrated into the GAIA-X network; doing so, combined with reachability of these resources for various users, offers users the prospect of savings on time and costs as well as efficiency advantages through the use of economies of scale.

- The openness and the resulting flexibility make it possible to connect existing (data) platforms to further research and health domains and also to international initiatives. Accordingly, it is also possible to achieve a simpler access to the *de.NBI* cloud and to use it more, for instance in connection with future funding projects placed upon the GAIA-X architecture.

- The possibility to integrate data across individual domains (e.g. visual-image data, clinical information) gives the GAIA-X network the potential to conduct integrative, more complex analyses within personalised medicine, for patients' benefit.

Patrons
Christian Lawerenz, Prof. Dr. Roland Eils and Jürgen Eils –
Berlin Institute of Health and Charité – University Medicine Berlin
Peter Lichter and Ivo Buchhalter –
German Cancer Research Center

L: Innovation based on data: High Performance and Quantum Computing 'as a service'

Example from practice; current challenges faced

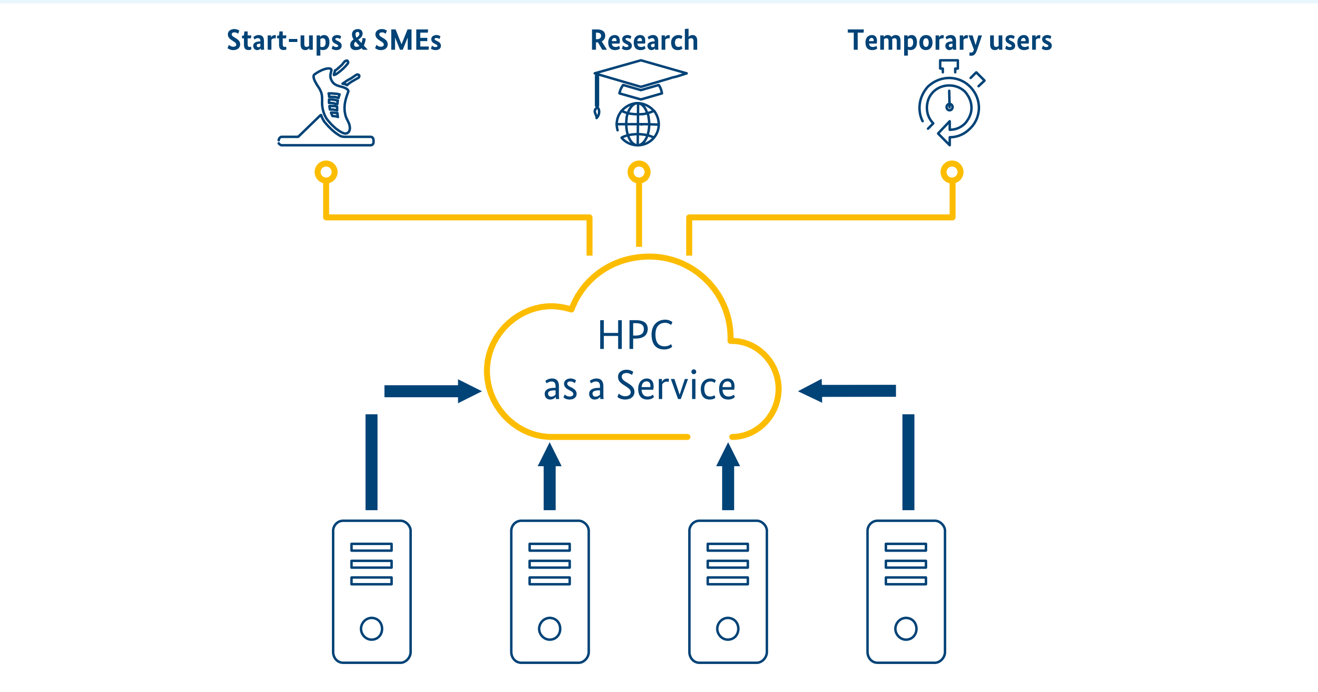
- The advance of digitalisation across all activity areas dramatically increases the availability of data. Analysing this data, with the help of AI and simulations (among other resources) makes it possible, for instance, to present climate or traffic models, or other scenarios of great complexity and a high data volume, such as in the sectors of finance (e.g. discovery of fraud scenarios), health (genome analysis) or industry (supply-chain optimisation).

- To conduct these computations, what is needed are very high levels of calculation performance at short notice, such as those made available by 'High Performance Computers' (HPC). A technology that is developing in this regard is quantum computing (at present: quantum simulators); this holds out the prospect of a multiplication of performance capability in tackling certain problematic issues.

- HPCs and quantum computers open up new possibilities in many areas; yet for many potential users it is not possible to invest in relevant systems – because these would only have partial use of capacity and are only in operation for a short time, due to the rapid innovation cycles.

- An option worth examining is joint use by various users via an infrastructure and service model that guarantees a high degree of (data) security and directs the inward and outward flow of the necessary data and also the separation from other services. First approaches to such 'sharing models' are under way in research (shared 'supercomputers'); yet these capacities are hard to access for users from industry, for instance, and no commercial operating model exists.

Figure 13: Use Case: High Performance and Quantum Computing 'as a service'



Source: ATOS

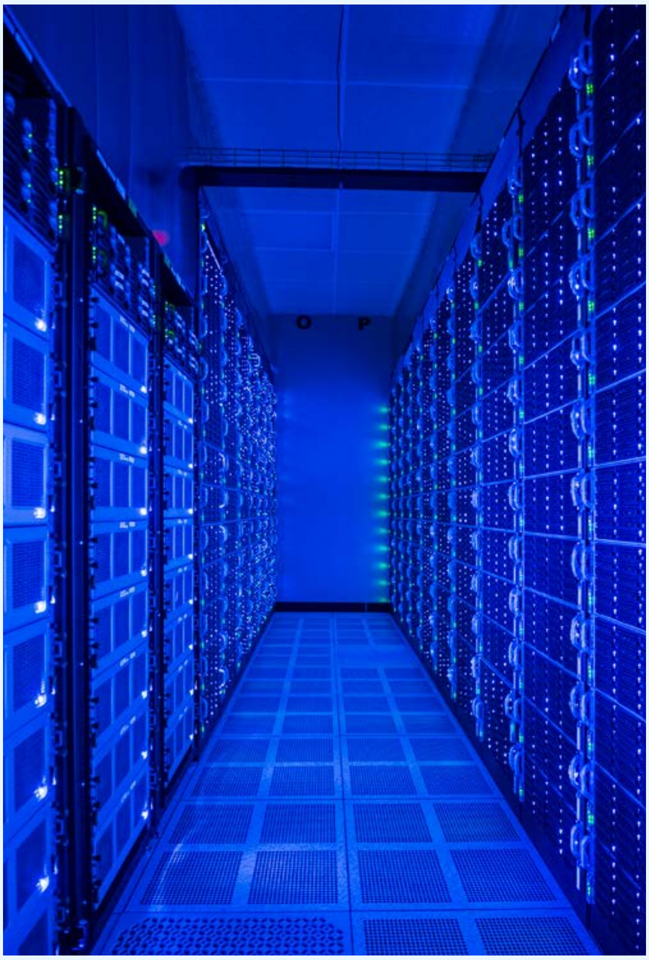
What added value does 'Project GAIA-X' offer?

- The project can offer broad access to HPC and quantum capabilities 'as a service'. This means that additional users (operating outside public research) can use HPCs for AI, modelling and simulations and can store the results securely. The service provider takes over responsibility for updates and maintenance.

- Because of increasing use, HPC services can be offered less expensively - the volume of capacity use goes up and the contributions to cost coverage can be distributed among a larger group of users.

- As a basis for secured marketplaces for data, spanning across individual business sectors, the project can set up an ideal environment for operation of HPCs, one where users always retain sovereignty over their data.

Patron
Klaus Ottradovetz – Atos



3.4 Added value from the provider's perspective

Project GAIA-X uses its federated data infrastructure to strengthen German and other European providers' chances of further developing, expanding and scaling up their cloud and computer-centre product offerings. The improved opportunities to introduce innovative, tailor-made solutions which are also well-suited to the market will increase investment activity. European providers will gain from the market growth that is forecast.

The basis for this is the establishment of a standardised data exchange between various security domains and data sovereignty beyond the boundaries of individual users' own systems. Modular integration of solutions and services, flexible use of the infrastructure, and the opportunity to bring functions to the data in different domains: these factors favour the emergence of new, innovative products and services. In this ecosystem, European start-ups and other companies will be able to develop new activities. The open-source approach and open access, even for the smallest service-providers, also enable the establishment of niche products and services.

GAIA-X combines individual solutions and transfers them into a homogeneous structure. The various offerings gain visibility on the market and can thus also distinguish themselves better, based on their competitive strengths.

The three levels of the value creation chain have the following categories of service-provider, in particular:

- Infrastructure: computer centres, edge computer centres, Internet Service Providers, software-defined wide area networks (SD-WANs), internet nodes and interconnections, in addition to carriers,

- Cloud and IT systems: cloud service providers, hosters, managed service providers (MSPs), and

- Services and platforms: platform providers, IDS, AI as a service, system integrators.

The federated data infrastructure enables substantial synergy effects to be attained on all three levels. In this context, decisive factors are the decentralised structures with quality-assured connection paths and mechanisms for the integrating and networking of industry locations and service locations. Innovative, scalable *'Made in Europe'* service solutions will emerge, with the demand of European business and industry as their orientation. At the same time, the open nature of the federated data infrastructure also offers access for niche service providers. Providers can transform the following advantages into reality:

- new client groups can be accessed, thanks to greater reach,

- their turnover potential is increased by the building up of a partner network, by strategic positioning and marketing, as well as by extending their service portfolio to include new, innovative solutions,

- their cost efficiency is improved thanks to common use of the federated infrastructure and

- their processes are improved with regard to the handling of orders, by using GAIA-X services.

4. Future Prospects

This project, Project GAIA-X, was drawn up by the ministries, companies and institutions named in the Appendix; in the process it has been widely endorsed. We are convinced that Project GAIA-X offers the right answers to the challenges presented by the new, global data economy. This openness to new, research-intensive concepts, characterised by risk, serves as a substantial contribution to our model's future competitiveness.

Our goal is to develop the next generation of a European data infrastructure, for Europe, its states, its companies and its citizens. Hence, we regard the federated data infrastructure as the cradle of an open, transparent digital ecosystem, where data and services can be made available, collated and shared in an environment of trust.

Over the last months, we have jointly steered this project, driving it forward in three individual workstreams. The user perspective and use cases of Workstream 1 are mostly to be found in Chapter 4 of this document. The conception of the technical foundations was mainly elaborated in Workstream 2; likewise, the results can be found in Chapter 4. The topics and the organisational structure are also to be found in the Appendix. Workstream 3 tackled the task of the project's communication activity and also of interlocking the project with European partners and the European Commission; the organisational proposal for project implementation was also drawn up in this workstream.

We now want to carry the project over into fixed structures. Together, we invite interested European partners to take part in the project and to further develop it, or to flank its activities with measures of their own. The same applies to those international partners who share our goals of data sovereignty and availability. To achieve this, we are striving to put Project GAIA-X onto a permanent basis, in the form of an organisation with legal capacity whose objective will be to advance and implement a European data infrastructure. In particular, this organisation will drive forward the elaboration of the reference architecture that the federated data infrastructure uses; it will also be responsible for determining and specifying the data infrastructure's technical requirements and its body of rules. The founding partners are also to decide on the name of the organisation, and of its data infrastructure and products.

To this end, the following criteria must be fulfilled: in particular, consistent with the goals of the project, the organisation must be attractive and open for partners from many different countries. It must work in an efficient and transparent way. Its tasks are to develop a reference architecture, to define standards, and to specify criteria for certifications and product quality seals. The organisation must be able to work in an economically viable way and implement a sustainable business model. For this, it must also be able to set up sub-organisations to pursue specified business purposes.

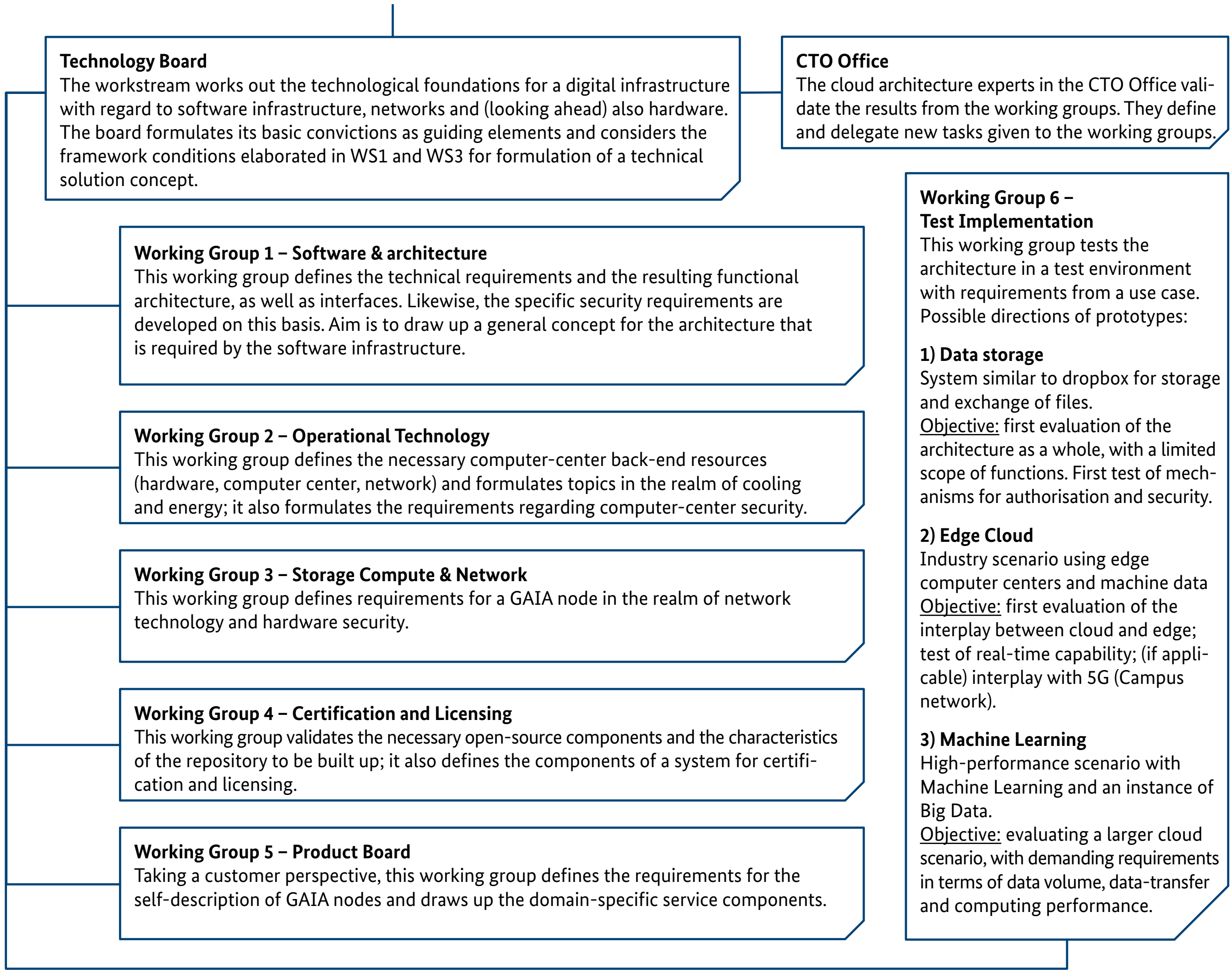
One possibility would be the formation of a European cooperative society (SCE), of which interested parties can become a member and to which they can commit resources. On this basis, over the coming weeks, and consistent with this project's openness and integrability, we invite interested European partners, notably those from France, to think further ahead with us. The organisation should be founded as soon as possible (spring of 2020). A governance model is needed for the participants' cooperation in the organisation and on the infrastructure. We will drive forward the coordination and conceptualisations necessary for this.

The first validation of the technical implementation, based on the areas of application, has already taken place. We will robustly drive the technical implementation forward and put the first use cases into practice as soon as possible. Alongside this task of forming this new entity, we are planning to promptly complete a first test of the technical concept in the second quarter of 2020 ('Proof of Concept'), as well as to start live operation by the end of 2020, with the first service-providers and users.

5. Appendices

Additional background materials and information on Project GAIA-X are available on the homepage: www.data-infrastructure.eu. Companies and organisations interested in getting involved in the project can also make direct contact with the responsible persons, at data-infrastructure@bmwi.bund.de.

Figure 14: Work structure – technology



Source: BMWi

6. Contributors

The foundation for building up a federated data infrastructure, based on European values, was developed at the initiative of the ‘Plattform Industrie 4.0’ and on behalf of the following people:

- **Peter Altmaier** (Federal Minister for Economic Affairs and Energy)
- **Dr. Roland Busch** (Siemens AG)
- **Jörg Hofmann** (IG Metall)
- **Prof. Dr. Henning Kagermann** (Global Representative and Advisor Plattform Industrie 4.0)
- **Anja Karliczek** (Federal Minister of Education and Research)
- **Prof. Dieter Kempf** (Bundesverband der Deutschen Industrie e.V. / The Federation of German Industries)
- **Christian Klein** (SAP SE)
- **Bernd Leukert** (Deutsche Bank AG)
- **Prof. Dr. Friedhelm Loh** (Friedhelm Loh Stiftung & Co. KG / German Edge Cloud)
- **Dr. Frank Melzer** (Festo AG & Co. KG)
- **Rolf Najork** (Robert Bosch GmbH)
- **Claudia Nemat** (Deutsche Telekom AG)
- **Prof. Dr.-Ing. Reimund Neugebauer** (Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.)
- **Henrik A. Schunk** (SCHUNK GmbH & Co. KG. Superior Clamping and Gripping)
- **Karl-Heinz Streibich** (acatech – Deutsche Akademie der Technikwissenschaften e.V. / National Academy of Science and Engineering)

This document was conceived and drawn up by the following people, in close cooperation:

- **Dr. Maximilian Ahrens** (Deutsche Telekom AG)
- **Fabian Biegel** (SAP SE)
- **Marco-Alexander Breit** (Federal Ministry for Economic Affairs and Energy)
- **Dr. Andreas Fier** (Deutsche Telekom AG)
- **Michael Jochem** (Robert Bosch GmbH)
- **Mirco Kaesberg** (Robert Bosch GmbH)
- **Dr.-Ing. Fabian Kohler** (Federal Ministry of Education and Research)
- **Dr. Thomas Lange** (acatech – Deutsche Akademie der Technikwissenschaften e.V. / National Academy of Science and Engineering)
- **Prof. Dr. Boris Otto** (Fraunhofer Institute for Software and Systems Engineering)
- **Dr. Carsten Polenz** (SAP SE)
- **Prof. Dr. Peter Post** (Festo AG & Co. KG)
- **Dr. Sebastian Ritz** (German Edge Cloud, a Friedhelm Loh Group company)
- **Ernst Stöckl-Pukall** (Federal Ministry for Economic Affairs and Energy)
- **Harald A. Summa** (DE-CIX Group AG)
- **Dr. Herbert Zeisel** (Federal Ministry of Education and Research)

and with the collaboration of:

- **Joachim Astel** (noris network AG)
- **Dr. Roman Bansen** (Bitkom e.V.)
- **Dr. Kevin Bauer** (TechQuartier, FinTech Community Frankfurt GmbH)
- **Dr. Stephan Bredt** (Hessian Ministry of Economics, Energy, Transport and Housing)
- **Jochen Breh** (Robert Bosch GmbH)
- **Dr. Ivo Buchhalter** (German Cancer Research Center Heidelberg)
- **Prof. Dr. Klemens Budde** (Charité – University Medicine Berlin)
- **Bundesdruckerei GmbH**
- **DE-CIX Management GmbH**
- **Prof. Dr. Claudia Eckert** (Fraunhofer Institute AISEC)
- **Günter Eggers** (e-shelter services GmbH)
- **Jürgen Eils** (Charité – University Medicine Berlin)
- **Prof. Dr. Roland Eils** (Berlin Institute of Health and Charité – University Medicine Berlin)
- **Dr. Thomas Esch** (Deutsches Zentrum für Luft- und Raumfahrt e.V. / German Aerospace Center)
- **Andreas Fauler** (arago GmbH)
- **Thomas Feld** (Strategion GmbH)
- **Dr. Marius Feldmann** (Cloud&Heat Technologies GmbH)
- **Dr. Dirk Franke** (DLR e.V. / German Aerospace Center)
- **Dr. Friedrich Gröteke** (Federal Ministry for Economic Affairs and Energy)
- **Heiko Großkopf** (Federal Office for Information Security)
- **Gerd Hoppe** (Beckhoff Automation GmbH & Co. KG)
- **Kai Kalusa** (VDMA e.V.)

- **Dr. Markus Ketterl** (msg systems ag)
- **Oliver Klein** (Bundesverband der Deutschen Industrie e.V. / The Federation of German Industries)
- **Lukas Klingholz** (Bitkom e.V.)
- **Johannes Krafczyk** (T-Systems International GmbH)
- **Thomas Kriesel** (Bitkom e.V.)
- **Christian Lawerenz** (Berlin Institute of Health and Charité – University Medicine Berlin)
- **Henning Lesch** (eco – Verband der Internetwirtschaft e.V. / Association of the Internet Industry)
- **Prof. Dr. Peter Lichter** (German Cancer Research Center Heidelberg)
- **Dr. Kai Lindow** (Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.)
- **Dr. Philipp Ludewig** (Federal Ministry for Economic Affairs and Energy)
- **Dr. Kai Martius** (secunet Security Networks AG)
- **Matthias Marx** (Federal Ministry for Economic Affairs and Energy)
- **Dr. Derek Meier** (Dataport AöR)
- **Dieter Meuser** (IoTOS GmbH)
- **Dipl.-Ing. Lars Nagel** (International Data Spaces e.V.)
- **Dr. Andreas Nauerz** (Robert Bosch GmbH)
- **Thomas Niessen** (Kompetenznetzwerk Trusted Cloud e.V. / Competence Network Trusted Cloud)
- **Clemens Otte** (Bundesverband der Deutschen Industrie e.V. / The Federation of German Industries)
- **Klaus Ottradovetz** (Atos SE)
- **Markus Quicken** (SupplyOn AG)
- **Alexander Rabe** (eco – Verband der Internetwirtschaft e.V. / Association of the Internet Industry)
- **Ulli Tobias Reitz** (Deutsche Telekom AG)
- **Thomas Riegler** (VDMA e.V.)

- **Dr. Martin Sauer** (Robert Bosch GmbH)
- **Dr. Thomas Schmidt** (acatech – Deutsche Akademie der Technikwissenschaften e.V. / National Academy of Science and Engineering; Plattform Lernende Systeme / Germany's Platform for Artificial Intelligence)
- **Dr. Christina Schmidt-Holtmann** (Federal Ministry for Economic Affairs and Energy)
- **Arne Schmieg** (German Edge Cloud GmbH & Co. KG.)
- **Tim Schneider** (Federal Ministry of Education and Research)
- **Marco Schuldt** (Federal Ministry for Economic Affairs and Energy)
- **Joachim Sedlmeir** (acatech – Deutsche Akademie der Technikwissenschaften e.V. / National Academy of Science and Engineering)
- **Nabi Siefken** (Federal Ministry for Economic Affairs and Energy)
- **Konrad Sippel** (Deutsche Börse AG)
- **Thorsten Sommer** (Deutsches Zentrum für Luft- und Raumfahrt e.V. / German Aerospace Center)
- **Sebastian Steinbuß** (International Data Spaces Association e.V.)
- **Dr. Christoph F. Strnadl** (Software AG)
- **Frank Trautwein** (RAYLYTIC GmbH)
- **René Walter** (Federal Ministry of the Interior, Building and Community)
- **Andreas Weiss** (EuroCloud Deutschland_eco e.V.)
- **Dr. Christian Weiss** (Deutsche Telekom AG)
- **Sascha Wessel** (Fraunhofer-Institut AISEC)
- **Lisa Witte-Stremmel** (Federal Ministry of Health)
- **Dr. Marianne Wulff** (Dataport AöR)
- **Torsten Wunderlich** (DATEV eG)
- **Julian Zeidler** (Deutsches Zentrum für Luft- und Raumfahrt e.V. / German Aerospace Center)

www.bmwi.de