Data Trusts,
Data Intermediation
Services and Gaia-X

gaia-x
Hub Germany

Whitepaper 2/2023 Version 2.0
July 2023

**Dennis Appelt, Peter Kraemer,
Abel Reiberg, Adam Smoleń**

acatech
DEUTSCHE AKADEMIE DER
TECHNIKWISSENSCHAFTEN

Authors

acatech – National Academy of Science and Engineering
Peter Kraemer, Head of Technological Sovereignty, and Industrial Value Creation and
                    Coordinator Gaia-X Hub Germany
Dennis Appelt, Researcher Gaia-X
Dr. Abel Reiberg, Researcher Gaia-X
Adam Smoleń, Student Assistant

Table of Contents

Summary

Data trusts and data intermediation services are increasingly seen as useful instruments for promoting a data economy that is in line with European values. As a result, more and more voices from politics, business and academia are calling for a favourable environment for such services. In addition, initial funding programs and regulatory measures have been implemented.

As the debate intensifies, however, it often remains unclear what is meant by terms such as data trust and data intermediation service and how these terms are related. This is also true in the context of Gaia-X, one of the most important infrastructure projects to promote sovereign, open, and transparent data sharing. This white paper aims to help clarify these ambiguities by explaining the concepts of data trusts and data intermediation services and their relationship, particularly in the context of the Gaia-X project.

It will become clear that the main purpose of data trusts and data intermediation services is to enable the use of data. The core function of the data trusts is to manage data on behalf of and in the interest of its trustors. As actors that exert control over data, the data trustees are often supposed to seek a balance of interests between data providers, data creators, and data consumers. But there is a wide variation in the complementary functions and other aspects of data trusts.

Initial requirements specifically for such services that can be considered data trusts were recently defined in the Data Governance Act (DGA). The legislation serves as the basis for a fair and transparent use of data that prevents monopolies and establishes fair competitive conditions in the data economy. To this end, the DGA defines the term data intermediation service. One of the key characteristics of such data intermediation services is that they enable a commercial relationship between data providers and data consumers. The DGA sets out minimum standards and requirements for providers of such services. Among other things, providers are required to register, not to discriminate against users with regard to competing services, and not to use the shared data for their own commercial purposes.

In the Gaia-X project, numerous stakeholders from different nations are working to realize goals that are similar to those underlying the DGA. While the Gaia-X concepts of embedded fraud prevention, interoperability, and transparent open-source software do not guarantee compliance with the DGA, they do provide a suitable basis for achieving it. Some examples of projects that practically implement the data trust concept and rely on Gaia-X are EuroDaT, the Mobility Data Space (MDS), and HEALTH-X dataLOFT.

1. General and Political Background

With the digital revolution, our society is experiencing a fundamental change. In the course of this change, data is becoming both an important resource in value creation processes, and a means to create and optimise such processes.

A lack of data, in turn, is a key obstacle to value creation and innovation. Currently, many data sources remain untapped, and potentials of the data economy remain unused. This is especially true in Europe, which is in danger of falling behind Asia and America in the development of the data economy.

One possible cause for this situation is a lack of trust. In the current platform-driven data economy, problems have arisen from the concentration of big data and the value created from it. This poses challenges for the functioning of markets and the protection of fundamental rights.

A possible solution is now seen in services described by terms such as data trust, data intermediation service and federator. They serve as tools that allow data providers to exercise maximum control and exactly determine when and how their data is shared. Such services thus promise to provide a widespread incentive for data extraction and data sharing, and an impetus for making data available and creating value from it.

Accordingly, significant efforts are currently being made to support the creation of such services. Both distributive and regulatory measures are being used.

In Germany the current governing parties stated in their coalition agreement that they would like to "support data trusts, data hubs and data donation in alliance with business, science and civil society" (Coalition Agreement 2021, p.17). Accordingly, projects for the development and implementation of data trusts are currently being funded – for example, within funding programs of the Federal Ministry of Education and Research (BMBF) (BMBF, 2022) and the Federal Ministry of Economic Affair and Climate Action (BMWK).

At the European level, data trusts and similar instruments have been the focus of attention for some time. For example, in its European Strategy for Data, the EU Commission explained, that it wants to create a "supportive environment" for "trusts acting as novel neutral intermediaries in the personal data economy" (European Commission, 2020).

In this context it is also worth to consider the efforts for European data spaces and Gaia-X – one of the first comprehensive projects to promote data spaces.[1]

Meanwhile a legal basis for the regulation of data trusts has been created with the Data Governance Act (DGA) which introduces the term data intermediation service.

But what exactly is behind terms such as data trust and data intermediation service and what does the regulation and practical implementation of corresponding services look like? This paper aims to provide answers to these questions.

---

[1] For an overview of Gaia-X see for example: Gaia-X Hub Germany (2023), Person and Schütrumpf (2023).

To this end, the following section will first deal with the concept of data trusts. Data trusts are defined according to their core function, the management of data or the rights to data on behalf of and in the interest of a group of stakeholders. Subsequently, the variety of data trusts is discussed on the basis of their complementary functions. It becomes clear that data trusts are services that have different characteristics and can serve – or balance – different interests.

The second section explains how the DGA lays the foundations for the regulation of data trusts. For example, the DGA provides guidelines on the additional functions that a data trust can offer. In doing so, the DGA does not mention the term data trust but defines the term data intermediation service. Due to this legal definition, which is now available for the first time, the term data intermediation service is particularly suitable for describing concrete projects.

Examples for such projects from the context of Gaia-X are discussed in the third section of this White Paper. As can be seen here, Gaia-X offers concrete possibilities for creating DGA-compliant services. Some examples for such services can be found in the Gaia-X projects Mobility Data Space, EuroDaT and HEALTH-X dataLOFT.

2. The Term Data Trust

The term data trust is increasingly used in discussions about the opportunities and challenges of the data economy. There it seems to stand for the vision of a decentralized data economy that does not rely on centralized platforms. The concept of data trusts is often associated with hopes for a functioning competitive market and a just and diverse economy. In the context of the term "platform", on the other hand, problems such as the concentration of data and the values derived from data are increasingly being addressed alongside opportunities.

The concept of data trusts is thus associated with a specific approach to promoting the data economy and protecting users' rights (see e.g., BMBF, 2022). But how exactly can the term data trust be defined? To date, the term has been used increasingly and in different ways and with different connotations, which is why it makes sense to propose a definition of the term in this White Paper.

2.1. Core Function of Data Trusts and Definition of Terms

Such a definition can be based on various aspects of data trusts. A common approach is to define data trusts with reference to the their function (Specht-Riemenschneider & Kerber, 2022). Similarly, we derive a general definition by referring to the core function of data trusts. Based on this general definition, narrower definitions for specific needs can be developed with reference to sub-functions and complementary functions.

Core Function: Data Stewardship or Exchange

The core function of data trusts is described in different ways in the literature, but many of the approaches focus either on the stewardship or the exchange of data so far.

An example of approaches with a focus on data stewardship is the approach of the Open Data Institute (ODI), which has made a significant contribution to the understanding and implementation of data trustees. The ODI's 2018 definition places data stewardship as the core function of a data trust (Hardinges, 2018). The ODI further narrowed this definition in 2020 by describing stewardship as "independent and fiduciary" (Hardinges, 2020).

An example of approaches focusing on data sharing is the approach of Blankertz and Specht-Riemenschneider (2021), who consider data trusts as data intermediaries and see their function in the exchange of data or in facilitating data sharing. Accordingly, data trusts are considered as institutions that enable data sharing by either establishing contact between data providers and data recipients, or by transferring data (Blankertz & Specht-Riemenschneider, 2021).

One advantage of defining the term data trust with a focus on data stewardship is that this comes closer to common definitions of the more general concept of trusteeship. In (German) everyday language, trusteeship is also understood as the stewardship or exercise of third-party values or rights (Duden.de, 2023). Similarly, data trusteeship can be understood as the stewardship or management[2] of certain data or the exercise of rights in relation to the data in question.

For this paper, a definition is chosen that takes up both, the focus on data stewardship as well as the focus on data sharing: **Data trusts are defined as institutions that manage data or rights to data on behalf of and in the interest of a group of stakeholders.[3]** In the course of their activities, the trustees obtain control over data and then use it immediately or at a later point in time to enable access by the data provider or third parties. The latter case reflects data sharing, thus both variants of definitional approaches explained above are taken into account.

The main advantage of such a definition based on data management – compared to a definition solely based on data sharing – is that it is broader and therefore more likely to cover all the current application of the term. To illustrate: It is easy to think of examples where a data trust manages data but does not enable data sharing. For example, the main task of a personal information management system could be to keep data secure for a specific data provider and access would only be enabled for the data provider after appropriate authentication. Conversely, only a few examples can be imagined in which a data trust enables data sharing but does not manage data in any way.

---

[2] It should be taken into account that trusteeship generally refers to tangible goods. Data, on the other hand, is an intangible good. Connected to intangibility (at least in the case of data) are many other properties, such as non-rivalry in use. Such properties are relevant for the transferability of the concept of trusteeship to the object of data. However, the extent to which they allow for transferability cannot be discussed in detail here. In the following, the term trusteeship usually refers to data trustees.

[3] In the translation of this paper, "to manage" was deliberately chosen over "to steward" to include the possibility of a data trustee that is less focused on representing the interests of specific actors and more on enabling data exchange. In the German version of this paper, the term "verwalten" is used. The definition is based on a broad understanding according to which management can also include control, administration, supervision, and care. Specifically, data management can also involve an actor having a significant degree of control over access to data without holding the data itself.

Representation of Interests by Data Trusts

The general definition presented above can be narrowed down further (also according to the respective needs), whereby further attention can be paid to the complementary functions or other aspects of a data trust. For example, the ODI's 2020 definition focuses more closely on the interests that can be taken into account when managing or sharing data (Hardinges, 2020).

The "neutrality" of data trusts is also often called for in this context (Buchheim et al., 2022). It should be noted at this point that neutrality — at least if the data trust is structured as an independent organisation — does not mean that it has no interests of its own, nor that it disregards or is unaffected by the interests of others.

Rather, the representation or balancing of certain interests is always linked to a trusteeship. Data trusts are usually structured (by technical, legal, or organisational means) in such a way that they can represent the interests of certain (especially structurally disadvantaged) groups of actors or balance these interests with the interests of other groups of actors.

Which interests are to be taken into account by the data trustees is an essential question that deserves and receives much attention but defies a one-size-fits-all answer.

For example, the interests of data providers may need to be considered. In a hypothetical example involving the exchange of medical data, data providers could be manufacturers of medical devices who generate medical data through their devices and the software installed on them. Furthermore, the interests of those to whom the data refers must often be considered. In said example, this could be patients who are examined by means of the medical devices. Finally, the interests of third parties may also have to be taken into account. In the example, these could be suppliers of medical devices who would like to acquire the data for their product development or people who could benefit from this product development as patients.

Usually, the interests of such different groups of actors will only partially coincide. In the hypothetical example, for instance, patients might be interested in having "their" data traded as little as possible in order to minimise the invasion of their privacy. The data providers, on the other hand, might be interested in limited data trading at high prices to maximise revenue from the provision of data. Data being traded widely and at low prices, thus minimising the cost of acquiring the data, would in turn be more in line with the goals of the data consumers.

The question of which of these interests a data trustees should represent or balance cannot be answered in a blanket manner. On the one hand, it may be required that certain (structurally disadvantaged) actors are represented against other (structurally advantaged) actors. On the other hand, it may be desired that a balance is achieved between (structurally similarly placed) actors. Furthermore, it may be desired that neutrality is ensured insofar as the pursuit of certain self-interests or certain interests of others is structurally excluded. The question of whose interests a data trust represents and to what extent can often only be clarified with a view to its concrete design. In particular, the variety of forms of data trusts must be taken into account.

2.2. Complementary Functions and Variety of Forms of Data Trusts

In general, but especially if one uses a broad definition, as is the case here, the range of forms resembling the term "data trust" is wide. This is evident, for example, with regard to the scope of functions of data trusts.

Complementary Functions of Data Trustees

On the one hand, the core function is implemented by data trusts in different ways (and different sub-functions). On the other hand, different trusts offer different complementary functions.

Both can relate, on the one hand, to the actors who want access to data. The data trustees can, for example, make data findable for them through cataloguing, they can take over the identification and authentication of interested parties, they can bring interested parties together (so-called matching) and make a pre-selection regarding supply and demand through more or less elaborate procedures.

On the other hand, functions can also relate to the data itself — that is, the trustees can offer or enable various data processing services (from pre-processing to analysis). Examples would be the adaptation of data formats, data cleansing, anonymisation or pseudonymisation of data and others.

In these steps of data processing, the trustees can either become active themselves or realise the processing through service providers. In both cases, the data trustees must ensure the desired level of control over the data (for themselves or others). It can also be a special function that the data processing cannot be supervised by any of the actors — not even by the data trustees themselves — which also means that no insight into potentially sensitive data is possible.

Classifications of Data Trusts

In addition to the functional scope, there are also large variances between current data trusteeship models regarding other features. These differences can also be used to distinguish classes of data trusts.

Blankertz and Specht-Riemenschneider (2021), for example, classify data trusts about two characteristics. They distinguish between data trusts with centralised or decentralised storage and data trusts who provide for mandatory or voluntary use. According to Blankertz and Specht-Riemenschneider, these characteristics of trusts have implications for the data protection risk associated with their use.

Blankertz et al. (2020) consider trusts according to the target group addressed by the service. They distinguish between "B2B models" and "B2C models" depending on whether the trustee represents consumers or companies.

Arlinghaus et al. (2021) in particular highlight the characteristic of belonging to the public or private sector and address differences in the business models of trusts, in terms of revenue mechanisms, organisational structure and resources.

Finally, Buchheim et al. (2022) distinguish three forms of data trustees. These include the "transaction-based data trustee", the "siloed data trustee" and private information management system (PIMS). For their categorization they address several distinguishing features of data trusts and explain the 'transaction-based data trustee' in great detail with reference to the EuroDaT project as an example. The key distinction for the classification being whether data is stored permanently with the trustee (in the case of silo data trustees) or not (in the case of transaction-based data trustees). In the case of transaction-based data trustees, it is ruled out that data of users is permanently stored and that users have to waive their rights to data (Buchheim et al., 2022).

These categorisations show that there is a need for a differentiated consideration of various data trusteeship models, which is also due to the wide range of forms of trusts developed so far. The field of data trusts currently shows itself to be diverse. This is true not only in terms of how the basic function of the trust is implemented (for example, whether the management of data involves permanent storage or not) and what other functions are offered (for example, whether simple storage of the data takes place or more complex processing is enabled). Rather, the diversity is also evident in relation to many other aspects, such as whether a trust is state-owned or non-state-owned.

The expectations placed on data trusts are correspondingly diverse. On the one hand, there are expectations that data trustees will enforce certain interests over other interests — for example, that the goal of data minimisation for end users will take precedence over the goal of data availability for service providers. On the other hand, it is hoped that data trustees will be able to balance such different interests, for example by facilitating specific agreements between end-users and service providers.

It is hoped that it is precisely through such balancing of conflicting interests that current obstacles to the data economy can be at least partially overcome. To this end, distributive and regulatory policies are also of particular importance. One of the most important European measures in this context is the Data Governance Act.

3. The Data Governance Act and the Term Data Intermediation Service

The aim of the Data Governance Act as part of the European Strategy for Data is to strengthen the data economy in Europe and, in doing so, to avoid undesirable developments that have become apparent or may become apparent as a result of the dominance of large platform operators.

In doing so, the DGA takes up the concept of data trusts. However, the central term ultimately defined and used, is that of data intermediation service, which is why it gains in clarity and relevance compared to the term data trust.

3.1. The Term Data Intermediation Service

In the DGA a data intermediation service is essentially considered to be a service "[…] which aims to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means, […]" (Art. 2 No. 11 DGA).

Explicitly categorised as data intermediation services are also such services where data sharing is carried out "[…] for the purpose of exercising the rights of data subjects in relation to personal data […]" (Art. 2 No. 11 DGA). This is a reference to widespread data trust models.

Not to be considered as data intermediation services are, inter alia: services involving the sharing of data without a commercial relationship being established between the provider and the user of the data (Art. 2 Nr. 11 (a) DGA); services that are used in a closed group; and "services that focus on the intermediation of copyright-protected content" (Art. 2 Nr. 11 (b) DGA).
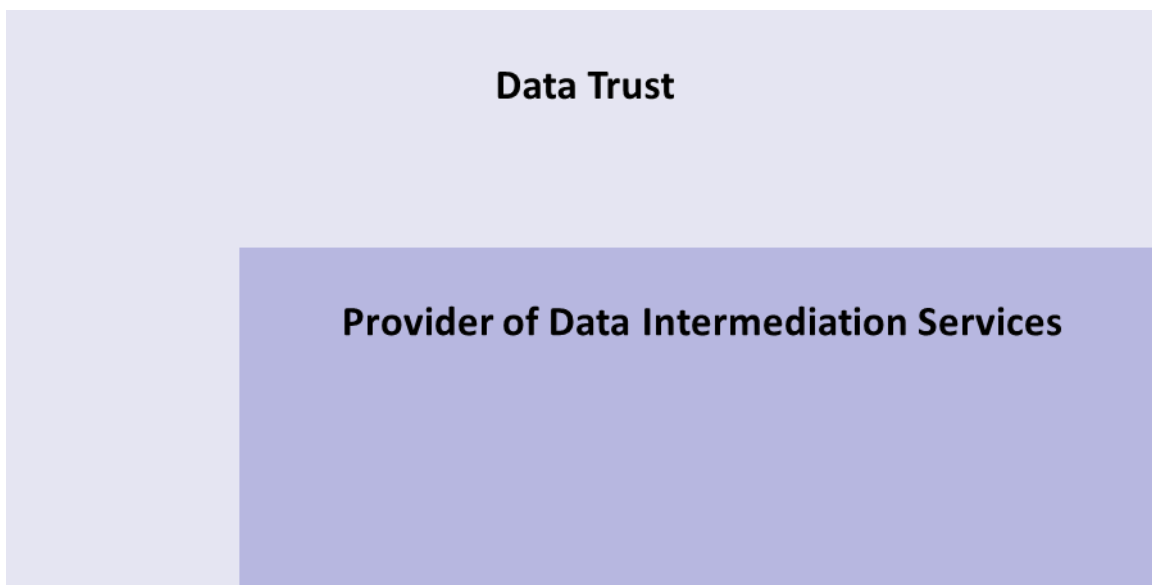
The term data intermediation service, which had previously been scarcely used, had prevailed over the term provider of data sharing services / data intermediary in the course of the legislative process. It thus joins that of data trust as another central term in many contexts.

There is a great overlap between the terms. For example, according to the above definition, data trusts are essentially acting as data intermediaries under the DGA if a business relationship for the sharing of the data managed by the data trustee is established by the data trustee.

This may also be the case, according to the definition used above, if the data trustee / provider of data intermediation services in question does not have direct access to the data. An example could be a so-called federator in the sense of the Gaia-X framework, an organisation that operates a data space and thus facilitates the exchange of data between data providers and data users. The federator could take over the management of the data so that it gives data providers the opportunity to precisely define their preferences for data sharing and ensures that data is only actually shared if these preferences are met. Here, although the data trustee / provider of data intermediation services does not have direct access to the data, it does have significant influence over who gets access.

However, the terms data trust and data intermediation service are not entirely congruent. For example, a data trustee would not be considered a provider of a data intermediation service if it only interacts with one person who is then both a data user and a data provider. Similarly, this would not be the case if the data provider and data user are in a closed group for the purposes of the DGA. Furthermore, a data trustee is not to be considered provider of data intermediation services if its activities result in the management of data, but no business relationship is established between the data provider and the data recipient (cf. Art. 2 No. 11 DGA).

Based on these considerations, the correspondence between the terms data trust and data intermediation service can be simplified as shown in Fig. 1.

Fig. 1: Relationship between the terms data trustees and data intermediation services. Source: Own representation

However, the extent of the intersection between data trust and providers of data intermediation services can only be described provisionally and approximately as long as the diversity of definitions for the term data trust remains and the present legal definition of the term data intermediation service has not been used in judicial decisions and only rarely by corresponding literature.

In view of this situation, it may be advantageous to resort to only one of the two concepts for the description of concrete projects. In this case, the term data intermediation services may be more useful since a legal definition for the term exists and clear legal requirements apply.

3.2. Specifications for Data Intermediation Services

Since the DGA sets binding requirements for providers of data intermediation services, it is worth taking a closer look at the contents of the DGA. The purpose of the DGA is not to make the dissemination of data intermediation services more difficult, but to promote such services by creating minimum standards that strengthen the trust of users. The requirements of the DGA serve both to promote competition in the data economy and to protect users' rights. How closely the two aspects are linked has been shown in the past, especially with regard to large platform operators. The market dominance of these operator has shown less by problematic pricing and more by the restriction of users' rights. In the area of social media, examples can be found in which, due to market dominance, prices for users were low and often zero, but user rights — such as the right to informational self-determination — were restricted. The DGA is now intended to create the basis for data sharing in general to rely less on vertically integrated large platforms and more on data intermediation services that act as neutral instances of sharing.

Most of the components of the DGA fall into three main areas:

The first area deals with the utilisation of data which are held by public authorities, and which are particularly worthy of protection for various reasons. The DGA provides, among other things, that these data remain effectively protected on the one hand and are made accessible

for further use under certain conditions and, if necessary, requirements on the other. For example, exclusivity agreements between data-providing public bodies and data-receiving bodies are largely prohibited.

The second area concerns the use of data for altruistic purposes. This is essentially understood as the use of data for non-commercial purposes of general interest based on voluntary consent. Here, the DGA provides that data altruistic organisations can register in a European register. Such organisations also have to comply with certain requirements, for example regarding consent management, which should promote trust in and use of the services in question.

Finally, the third area, which has particular relevance to this White Paper, concerns requirements for data intermediation services. These fall into three main categories:

Firstly, there are requirements that oblige providers of intermediation services to register. The necessary requirements and processes, including deadlines and fees, are defined for this registration.

Secondly, specifications are made as to which services may be offered by a provider of intermediation services. These specifications essentially determine which underlying business models are permissible. It is, for example, largely prohibited that data that was shared through an intermediation service is used for the own commercial purposes of the intermediation service provider. Essentially, the only processing steps that are allowed for the provider of data intermediation services are those that are carried out on behalf of the client and are intended to facilitate data sharing, such as pseudonymisation of data.

A functional separation between providers of data processing services and providers of data intermediation services is thus established. For this purpose, an organisational separation of such providers is prescribed.

In particular, the aim of the functional separation is to ensure that there are no strong commercial incentives for providers of data intermediation services to enable data use that does not correspond to the interests of the data providers and other entitled parties.

Thirdly, specifications are made that oblige providers to offer their services in a certain way. Among other things, providers must offer their services in a fair, transparent and non-discriminatory manner, while meeting security requirements. Interoperability must also be ensured with regard to various aspects. These requirements, which will be discussed in more detail in the next section in relation to Gaia-X, are also intended to ensure that the market for data intermediation services develops into a level playing field and that tendencies towards market concentration are discouraged.

In order to meet these various legal requirements in practice, numerous organisational and technical measures must be taken. In practical implementation, it is particularly worthwhile to look at advanced projects to promote sovereign data sharing, such as Gaia-X.

4. Gaia-X and the Data Governance Act

Gaia-X is currently one of the largest European projects for the utilisation of decentralised infrastructures for sovereign data sharing. The Gaia-X framework provides specifications and proposals for the technical and organisational design of a federation of data spaces. This framework also includes a definition of fundamental roles in the emerging ecosystem, of which the role of the federator is of particular importance for the construction of the ecosystem as well as for the topic of this White Paper. Federators are essentially responsible for providing the services that are important for a functioning data exchange in Gaia-X data spaces.

Although it is not possible to make a conclusive statement yet, it can be assumed that the vast majority of organisations that will act as federators in the sense of the Gaia-X framework will also be considered providers of a data intermediation service in the sense of the Data Governance Act. The primary purpose of building and operating a Gaia-X data space is to enable and facilitate data sharing between participants in the respective data space. This largely corresponds to the core function of the provider of a data intermediation service in the sense of the DGA, the establishment of a business relationship between data provider and data consumer. Otto (2022) summarises accordingly:

> *„Data spaces and their underlying software infrastructures must support trust, interoperability, and portability of data and data sovereignty and must be non-discriminatory. Thus, data spaces can be understood as intermediaries and data sharing service providers to which the EU Data Governance Act applies which is currently under review."* (Otto et al., 2022, p. 7 f)

It should be noted that Otto (2022) made the above statement before the final version of the DGA was published. However, the statement is still relevant, as it can be assumed that all participants in a data space are likely to be data owners, data users, or both in the sense of the Gaia-X framework, and in the sense of the DGA.

However, it must be taken into account that the Gaia-X framework is designed to be open with regard to federators, which is why federators that do not offer data intermediation services are also conceivable. For example, an actor could facilitate the provision of data as a federator without establishing a business relationship between data holders and data users. It would thus not offer a data intermediation service according to Art. 2 No. 11 lit. d) DGA. Based on these assumptions, the relationship between data trustees, providers of a data intermediation service and federators can be visualised as in Fig. 2.
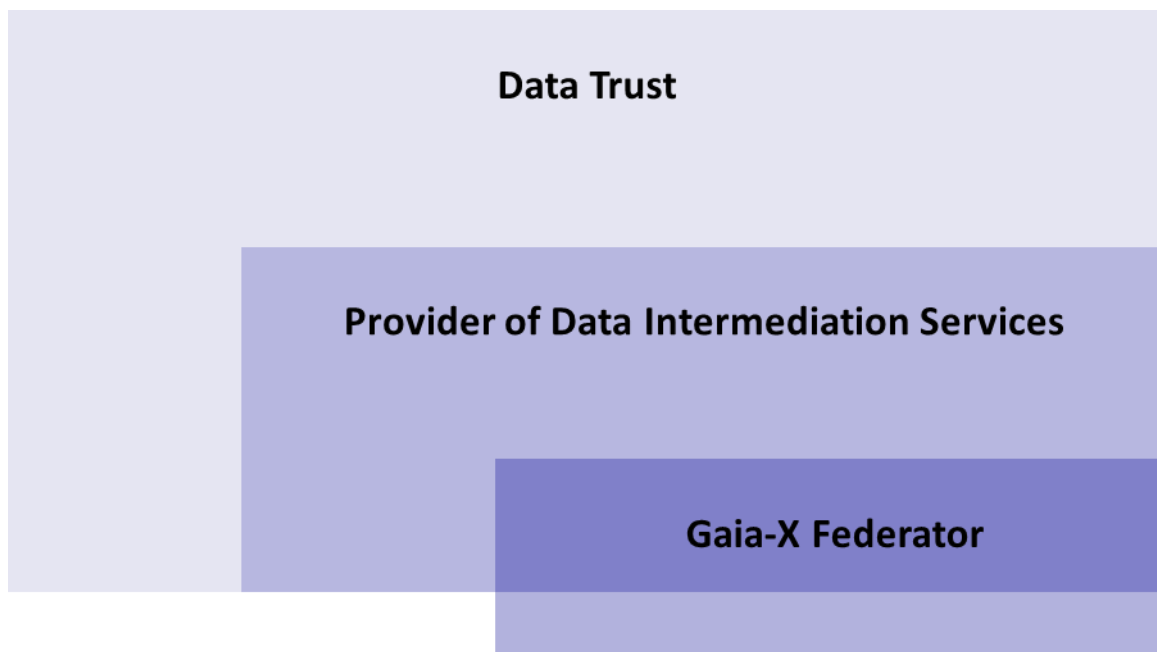
Fig. 2: Relationship between the terms data trustees, data intermediation services and federator. Source: Own representation

With regard to the relationship of the three concepts to each other, however, it must be taken into account that for a certain period of time after the entry into force of the DGA, it will not yet be conclusively clarified how corresponding services must concretely meet the requirements of the DGA. This is comparable to the years after the publication of the General Data Protection Regulation, in which the wording "appropriate technical and organisational measures" (Article 32 (1) of the GDPR) was still subject to a comparatively broad scope of interpretation until a series of court rulings made it clearer how the legal requirements can be met. Therefore, it is to be expected that, analogous to the transitional period since the introduction of the GDPR, clarity will also be achieved step by step after the introduction of the DGA and similar legal acts.

Even if it has not yet been conclusively clarified how the requirements of the DGA are to be implemented in practice, it can be assumed that public authorities will consider the use of voluntary commitments, which is accompanied by Gaia-X compliance and conformity with more far-reaching requirements in Gaia-X data spaces.

The reason for this is the strong convergence in the objectives of the policy DGA and the project Gaia-X. Put simply, the DGA attempts to achieve the regulatory goal that Gaia-X set itself as an infrastructure project several years ago and has continued to consolidate: sovereign data sharing in open and transparent ecosystems.

That is why several of the requirements of the DGA also coincide with those of Gaia-X. Accordingly, while Gaia-X compliance is not the same as DGA compliance, meeting the requirements of the DGA can be made much easier by participating in Gaia-X.

First indications on how Gaia-X helps to reduce legal risks when complying with the requirements of the DGA follow in the coming sub-chapters.

4.1. Interoperability

Art. 12 (i) of the DGA

"the data intermediation services provider shall take appropriate measures to ensure interoperability with other data intermediation services, inter alia, by means of commonly used open standards in the sector in which the data intermediation services provider operates"

Considering Art. 12 (i) of the DGA, the participation in a Gaia-X data space in a specific industry sector can be seen as an important measure for the required guarantee of interoperability with other data intermediation services — it could even be argued that precisely this interoperability between data intermediation services is a core principle of Gaia-X. Also, the use of commonly used open standards (of a sector) is an important building block of any (sector-specific) data space.

Also, the retention of original data formats as stated in Art. 12 (d) — unless conversion is legally necessary, serves interoperability or harmonisation or is requested by the data user — strongly coincides with the interests that are likely to be present in most Gaia-X compatible data ecosystems.

4.2. Fraud Prevention

Gaia-X compatibility can be a basis for the use of "procedures […] to prevent fraudulent or abusive practices in relation to parties" as required by Art. 12(g) of the DGA. Components of such procedures are being developed with the Gaia-X Federation Services (GXFS) work packages (eco – Verband der Internetwirtschaft, 2023).

Art. 12 (g) of the DGA

"the data intermediation services provider shall have procedures in place to prevent fraudulent or abusive practices in relation to parties seeking access through its data intermediation services"

One important cornerstone of the GXFS is the work package "Identity & Trust", in which the possibilities are created to manage decentralised identities in a trustworthy manner. For this purpose, there are services for authentication/authorisation, authentication managers and

trust services that contain a wealth of functions. This will make it much more difficult for fraud to be attempted in Gaia-X with false identities, which should facilitate law enforcement.

In combination with the other GXFS, especially the "Compliance" work package, this also lays a foundation for the prevention of fraud and abuse by means of Compliance as Code (Gronlier, 2022), although this does not yet guarantee compliance with Art. 12 (g) of the DGA.

4.3. Fairness, Transparency, and Non-discrimination

Infobox 3

Art. 12 (f) of the DGA

"the data intermediation services provider shall ensure that the procedure for access to its service is fair, transparent and non-discriminatory for both data subjects and data holders, as well as for data users, including with regard to prices and terms of service"

The fairness, transparency and non-discrimination required in Article 12 (f) of the DGA are also taken into account in the Gaia-X framework. The openness of the Gaia-X ecosystem, for example, prevents discrimination, as no actors are per se excluded from participation. In principle, discrimination at another level could also be implemented in Gaia-X, for example by only allowing participation in a data space if specific services of a provider are already being used. However, such a practice would at least be transparent in the requirements for participation in this data space. In general, many components of the GXFS contribute to transparency; for example, the so-called labelling framework from the work package "Compliance" discloses with which (legal) requirements a service or data space is compliant (Gaia-X AISBL, 2021): For the highest level three, among other things, immunity against non-European access must be ensured resulting from national legislation like the US Cloud Act. In Gaia-X, these and other requirements strengthen trustworthiness and reduce obstacles to data sharing — and contribute to compliance with Art. 12 (f) of the DGA.

4.4. Pricing

Other requirements of the DGA — for example on the topic of independent pricing from Art. 12 (b), as well as fair prices, terms, and conditions in Art. 12 (f) of the DGA— are not fulfilled by mere Gaia-X compliance. However, the status quo is made transparent. This is also intended to offer Gaia-X services and data spaces the greatest possible scope for design. Here, it is up to the individual data spaces which further requirements are placed on their participants. However, due to the pioneering role of individual projects and the governance for and by the Gaia-X Digital Clearing Houses (Gaia-X AISBL, 2023), it can be assumed that Gaia-X data spaces will facilitate DGA compliance by means of further requirements.

5. Examples from the Projects

As can be seen, there is a great overlap between the three concepts mentioned: data trusteeship, data intermediation services, and the Gaia-X federator. This can be seen firstly at the conceptual level, with a view to the definition approaches for data trusteeship, legal definitions and specifications for data intermediation services and architecture descriptions for the Gaia-X federator. Secondly, the connection between the concepts is also evident at the practical level — especially with regard to concrete projects. In the following, three projects of the Gaia-X context, namely EuroDaT, Mobility Data Space and HEALTH-X dataLOFT, will be discussed to illustrate the intersection once more:

The EuroDaT project ("European Data Trustee") has the term data trustee in its name and, according to its own statements, aims to create a "transaction-based data trustee". This is to be characterised by the fact that there is no holding and reuse of data, but only a transfer of analysis results after distributed algorithmic processing (Buchheim et al., 2022). Each transaction is also "encapsulated", so neither the trustee EuroDaT nor the analysis recipients can access the raw data. The using party only receives the result of the processing. Thus, EuroDaT is also a data trustee according to the above definition, because not data, but the analysing access to data is managed in the interest of the data providers. As a funded Gaia-X project, the spin-off EuroDaT GmbH will find its role in the Gaia-X ecosystem and can either offer services as a DGA-compliant component of an existing data space or create a new data space then.

The Mobility Data Space (MDS) is a Gaia-X data space. Accordingly, the operating company "DRM Datenraum Mobilität GmbH" can be classified as a federator in the sense of the Gaia-X framework. Moreover, as described above, it can probably be regarded as a provider of data intermediation services on the basis of the DGA. Conformity with the regulations of the DGA is accordingly one of the goals of the operating company. In the case of the Mobility Data Space, although data sharing contracts are concluded between the exchanging parties and not with the MDS itself, the MDS can nevertheless be assigned to the broad definition of data trust above. This is the case because data sharing contracts only are concluded with the substantial assistance of the MDS, which, on the basis of the majority shareholder acatech (a non-profit association), realises a neutral representation of the interests of the stakeholders participating in the data space. This includes, among other things, the reduction of trust deficits through a standardised technical and legal framework, the reduction of information deficits, and the provision of a matchmaker function by describing the data offer in the metadata catalogue, as well as ensuring the correspondence of data usage rights and actual data usage through standardised ID management.

A Gaia-X-compliant data space is also to be created during the HEALTH-X dataLOFT project — in this case in the health sector. For example, a spin-off yet to be created in HEALTH-X dataLOFT could act as a federator and, as such, will probably have to comply with the specifications of the DGA as a data intermediation service. An important difference to the MDS is that HEALTH-X dataLOFT will not be a pure B2B data space, but patients will make up a large

part of the participants — as providers of the health data related to them. This leads to several technical, legal and organisational implications, of which a detailed explanation would exceed the scope of this paper. However, the Gaia-X framework offers the possibility for the creation of applications of various kinds. While HEALTH-X dataLOFT does not hold any data at the organisation itself, like the MDS, an operating model of this project can also be called a data trust based on a broad definition. This is because, especially in the context of such a health data space, the conflict of interest between the data providers, the data users, and the persons to whom the data relate must be addressed. The interests that need to be balanced are illustrated by the example from chapter 2.2, in which data from manufacturers of medical devices about patients could be used by third parties. To balance such, at least partially conflicting interests, HEALTH-X dataLOFT provides corresponding participation opportunities for relevant stakeholders. There is a particular focus on citizens. Among other things, citizens must be asked for granular consent for certain uses by certain data users and must be informed transparently about any further use of data in accordance with the consent given.

6. Conclusion

On both a conceptual and practical level, a wide overlap can be seen between the concept of data trusts, the concept of data intermediation service as defined in the DGA, and the Gaia-X concept of a federator. The broadest concept considered in this White Paper is the one of data trust. It should be noted that this term has been used in many ways. In this White Paper, different approaches to defining the term have been identified and taken into account in a separate definition. Following this definition, data trusts are institutions that manage data or rights to data on behalf of and in the interest of stakeholders.

Another term that can be used in many cases as an alternative to the term data trust is data intermediation service. The concept of data intermediation service is narrower than that of data trust and essentially includes those services that establish a business relationship between data providers and data recipients. The term data intermediation services may be more suitable for describing concrete services than data trust, since the DGA provides a first legal definition of the former. Also included in the DGA are concrete requirements for providers of data intermediation services. These are intended to strengthen the trust of users and the dissemination of data intermediation services. Participation in Gaia-X can be a helpful step in implementing the requirements, as many of the intentions behind the DGA are congruent with those of the Gaia-X ecosystem. Some of the data trustees and providers of a data intermediation service, active in the Gaia-X ecosystem, will be able to be referred to by a third key term, the federator as defined in the Gaia-X framework. Federators are essentially those organisations that ensure the provision of Gaia-X federation services.

How the understanding of the three aforementioned concepts can be interpreted in concrete terms and which roles different data trustees can assume in Gaia-X, becomes clear when looking at the three projects EuroDaT, Mobility Data Space and HEALTH-X dataLOFT. It also shows how projects for the realisation of data trusts, the Gaia-X project, and regulatory measures such as the DGA intertwine in a meaningful way thanks to the very similar underlying goals and requirements.

Bibliography

Arlinghaus, T., Kus, K., Kajüter, P., & Teuteberg, F. (2021). *Datentreuhandstellen gestalten: Status quo und Perspektiven für Geschäftsmodelle. HMD Praxis der Wirtschaftsinformatik*, *58*(3), 565–579. https://doi.org/10.1365/s40702-021-00727-x

Blankertz, A., Kuzev, P., Richter, F., Richter, H., Schallbruch, M., & Von Braunmühl, P. (2020). *Datentreuhandmodelle*. https://doi.org/10.13140/RG.2.2.24915.30246

Blankertz, A., & Specht-Riemenschneider, L. (2021). *Neue Modelle ermöglichen – Regulierung für Datentreuhänder*.

BMBF. (2022). *Datentreuhandmodelle: BMBF fördert Pilotvorhaben - BMBF Digitale Zukunft*. Bundesministerium für Bildung und Forschung - BMBF Digitale Zukunft. https://www.bildung-forschung.digital/digitalezukunft/de/technologie/daten/datentreuhandmodelle_pilotvorhaben/datentreuhandmodelle_pilotvorhaben.html

Buchheim, J., Augsberg, S., & Gehring, P. (2022). *Transaktionsbasierte Datentreuhand.* JuristenZeitung (JZ), 77(23), 1139–1147. https://doi.org/10.1628/jz-2022-0368

Duden.de. (2023). *Treuhandschaft: Rechtschreibung, Bedeutung, Definition, Herkunft.* https://www.duden.de/rechtschreibung/Treuhandschaft

eco – Verband der Internetwirtschaft (2023). *GXFS Toolbox. GXFS.Eu.* https://www.gxfs.eu/set-of-services/

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), (2016). http://data.europa.eu/eli/reg/2016/679/oj/deu

European Commission. (2020). *Communication from the Comission to the European Parliament, the Council, the European Economic and Social Committee and the Commitee of the Regions—A European strategy for data* (COM(2020)66). https://eur-lex.europa.eu/legal-content/de/ALL/?uri=CELEX:52020DC0066

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance), (2022). http://data.europa.eu/eli/reg/2022/868/oj/deu

Gaia-X AISBL. (2021). *Gaia-X Labelling Framework*. https://gaia-x.eu/wp-content/uploads/files/2021-11/Gaia-X%20Labelling%20Framework_0.pdf

Gaia-X AISBL. (2023). *GXDCH - Gaia-X: A Federated Secure Data Infrastructure*. https://gaia-x.eu/gxdch/

Gronlier, P. (2022). *Gaia-X: Compliance as Code - Gaia-X: A Federated Secure Data Infrastructure*. https://gaia-x.eu/news/latest-news/gaia-x-compliance-as-code/

Hardinges, J. (2018). *Defining a 'data trust.'* Open Data Institute. https://www.theodi.org/article/defining-a-data-trust/

Hardinges, J. (2020). *Data trusts in 2020.* Open Data Institute. https://www.theodi.org/article/data-trusts-in-2020/

Otto, B., ten Hompel, M., & Wrobel, S. (Eds.). (2022). *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage.* Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5

Sozialdemokratische Partei Deutschlands, Bündnis 90/Die Grünen, & Freie Demokratische Partei (Eds.). (2021). *Mehr Fortschritt wagen Bündnis für Freiheit, Gerechtigkeit und Nachhaltigkeit.*

Specht-Riemenschneider, L., & Kerber, W. (2022). *Analysen & Argumente: Datentreuhänder—Gesellschaftlich nützlich, rechtlich größere Anforderungen erforderlich*. 8.