Gaia-X FEDERATION SERVICES GXFS

gaia-x

Self-Description of Resources, Service Offerings and Participants within Gaia-X Ecosystems

Decentralised, cryptographically secure, interoperable, extendible, and future proof

Supported by:

Federal Ministry
for Economic Affairs
and Climate Action

on the basis of a decision
by the German Bundestag

Abstract

In Gaia-X, all providers have to describe themselves and their service offerings using standardised, machine-comprehensible metadata called Self-Descriptions. After an initial introduction of the role of Self-Descriptions within the Gaia-X ecosystem, this Whitepaper proceeds to explain the different levels on which trust in Self-Descriptions is established. Next, we illustrate what classes of service offerings can be described, and what attributes their Self-Descriptions can comprise. We introduce the governance process of the Gaia-X Working Group Service Characteristics, based on which the Gaia-X members agree on the specification of these classes and attributes. For the execution of this governance process, as well as any Gaia-X participant's work with Self-Descriptions, we offer practical tool support. Finally, we point out challenges that the specification and implementation of Self-Descriptions will continue to face as Gaia-X evolves and conclude with an outlook to next steps.

Table of Contents

1. Role of Self-Descriptions within Gaia-X Ecosystems

Participants in a Gaia-X ecosystem talk to each other in the language of Self-Descriptions. Providers use metadata to describe themselves, their service offerings, as well as the resources that their service offerings are composed of. These metadata records – that is, the Self-Descriptions – are comprehensible for the Federation Services that manage the Self-Descriptions through their lifecycle, as well as for consumers, including humans and the automated machine agents which assist them. This universal comprehensibility is ensured by building on the Semantic Web and Linked Data standards of the World Wide Web Consortium (W3C), which are widely established for data modeling, service definition and knowledge engineering. These standards are used not just to define schemas to structure data, but also to document the meaning of every term in such a schema, both in human-comprehensible natural language and in a lightweight machine-actionable logic. In this manner, Self-Descriptions enable interoperability in Gaia-X ecosystems.

2. Trust in Self-Descriptions

How do Gaia-X participants know whether they can trust the information that providers convey about themselves or their service offerings via their Self-Descriptions? The construction of a Self-Description follows the W3C Verifiable Credentials data model:

1. **Claims** are unverified statements about an entity without any guarantee of truth. For example, think of an ACME provider claiming, “My service is sustainable” and “My service is hosted in Germany” (and thus under EU legislation). There would be no guarantee that ACME’s statements about their service offerings are actually true. In Gaia-X, claims may only be made about entities defined in the Gaia-X Conceptual Model (see Section 3 below).
2. **Verifiable Credentials** are claims whose correctness has been checked and signed by a third party. A Verifiable Credential is a key concept with respect to trust in Self-Descriptions and the self-sovereignty of Gaia-X. As Gaia-X neither defines nor manages trustworthy third parties, it is instead up to the receiver of Verifiable Credentials to decide if its issuer is trustworthy or not. Let’s summon up the idea that TrueConfirm is a third party and authorised to verify ACME’s claims. A Verifiable Credential comprises one or more claims about ACME, verified and signed by TrueConfirm, indicating that these claims are true. Gaia-X participants who trust TrueConfirm can thus safely assume that, for example, ACME’s service is sustainable. Participants who do not trust TrueConfirm would consider TrustConfirm’s Verifiable Credential to be just a set of claims.
3. A **Verifiable Presentation** is a subset of the Verifiable Credentials regarding an entity selected for sharing with another entity for a certain purpose. Such purposes might include being admitted to some process; for example, a consumer might only be willing to purchase sustainable service offerings, while not being concerned about what country they are hosted in. The concept of Verifiable Presentations is important for digital sovereignty, which allows any Gaia-X participant to decide in a self-determined way what information about themselves should be shared and with whom.

Figure 1 visualises the different layers of trust regarding Self-Descriptions.

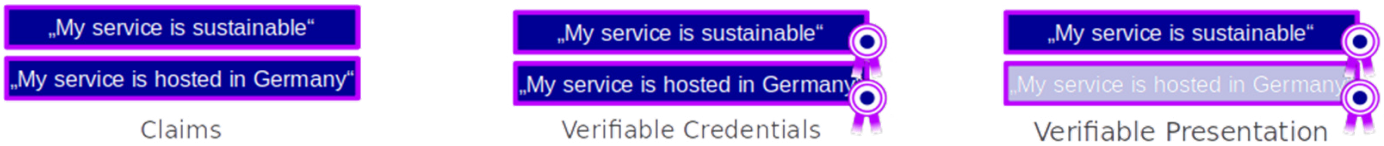

Fig. 1: Level of Trust within Self-Descriptions

The concept of claims, Verifiable Credentials and Verifiable Presentations is also known from the physical world. Imagine Bob wanting to buy alcohol. To do so, in many countries Bob would have to be at least 18 years old. Bob might go to a supermarket and simply claim “I am over 18”. If the supermarket staff do not know Bob and do not trust him, they will request a proof for his claim. The proof must be verified and signed by a third party the employee trusts – for example, Bob’s ID card issued by the government. Bob’s ID card is a Verifiable Credential, which, as a claim, includes a picture of Bob’s face, Bob’s address and Bob’s birth date. There is, however, one important difference between the physical and the digital world: In the digital world, Bob is empowered to create Verifiable Presentations by himself, which means he does not need to show his address when buying alcohol, as this information is not requested for that purpose.

Further details of Verifiable Credentials and Verifiable Presentations will not be covered in this Whitepaper. Instead, we will focus on the vocabulary of the claims and on the tools to create and validate pure, unchecked and unsigned claims.

3. Conceptual Model and Self-Description Attributes

The Gaia-X Conceptual Model, as published in the Gaia-X Architecture Document[1], defines all entities in the Gaia-X universe and their relationships with each other. The most important entities are those classified as “provider”, “service offering” and “resource”. Gaia-X defines a provider as a Gaia-X participant “who makes service offerings and operates resources”. A service offering is a description of a digital service available for order. Service offerings can be composed of resources and/or may depend on other service offerings by third party providers. A resource is a building block of a service offering that is not available for order. Figure 2 depicts the relevant subset of the Gaia-X Conceptual Model.

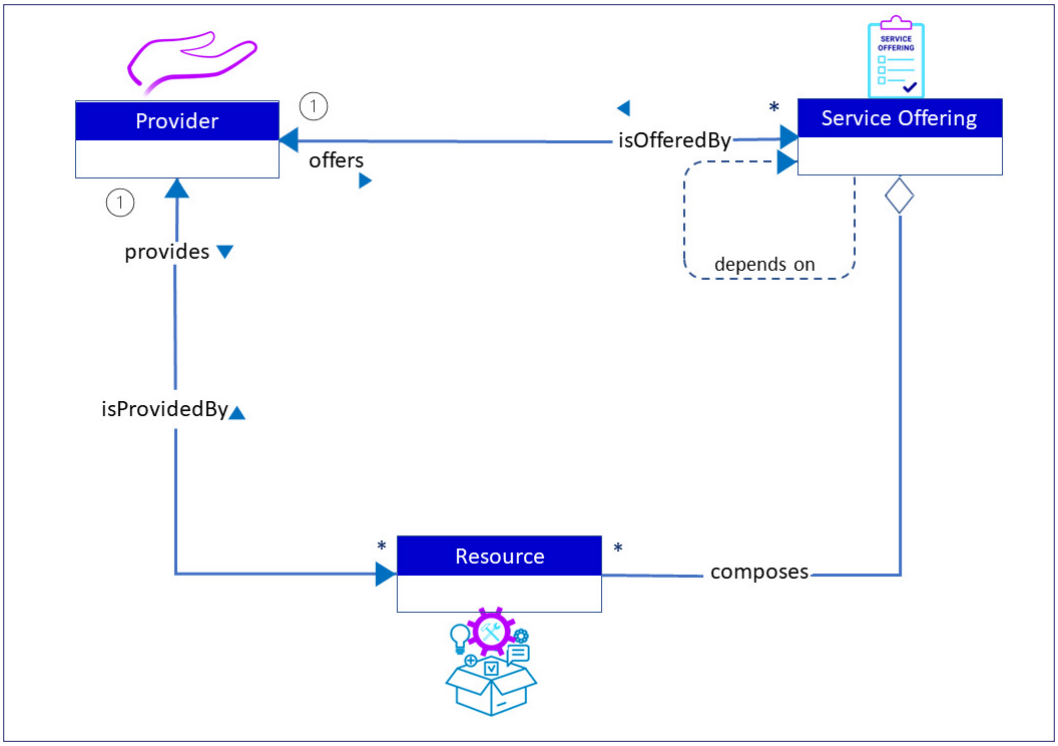

Fig. 2: Subset of the Gaia-X Conceptual Model

1 Gaia-X Architecture Document 21.12. https://gaia-x.eu/sites/default/files/2022-01/Gaia-X_Architecture_Document_2112.pdf

As explained above, a Self-Description is a set of claims about the properties of a provider, a service offering or a resource. The vocabulary in which claims are expressed is called a Self-Description Schema. It defines a hierarchy of classes corresponding to the Conceptual Model and going into further detail as explained below (in the "taxonomy"). For each class, it defines a collection of attributes. For each attribute, it specifies the name, the datatype, and, where applicable, a reference to underlying standards as specified by W3C, for example. Typical examples for Self-Description attributes are the address of the headquarter of a service provider, the geographical location of a data centre or the copyright owner of a data resource.

For compliance reasons, an attribute may be mandatory: that is, without this, a Self-Description would not be valid. For non-mandatory attributes, validity with regard to the Self-Description Schemas means, for example, the correct usage of datatypes. Mandatory attributes are specified in the Gaia-X Trust Framework; further requirements for the definition of attributes come from technical necessities and domain-specific circumstances. The Gaia-X Trust Framework defines a minimal set of attributes of the providers, service offerings and resources that have to be specified for these entities in order to be admitted as a compliant part of a Gaia-X ecosystem. For example, within their Self-Descriptions, providers must report their registration number within their respective national registry, such as the commercial register or the register of associations. Requirements coming from a technical perspective describe important information for the use of a service offering or resource, such as data formats. In addition, a service offering's or resource's domain may also dictate attributes. In a sustainable environment, for example, it may be recommended to publish information of a provider's and/or a service offering's carbon footprint.

4. Taxonomy of Provider, Service Offering and Resource Classes

To structure the definition of Self-Description attributes, Gaia-X has defined a three-level service taxonomy. The taxonomy's top-level elements are those of "provider", "service offering" and "resource", as defined in the Conceptual Model. These three classes are further refined by specialist subclasses, each of which has its own set of mandatory and optional attributes.

4.1 Provider

According to the Gaia-X Conceptual Model, a provider is a special role a participant can have within a Gaia-X ecosystem. A provider compiles service offerings and operates resources – see Figure 3.

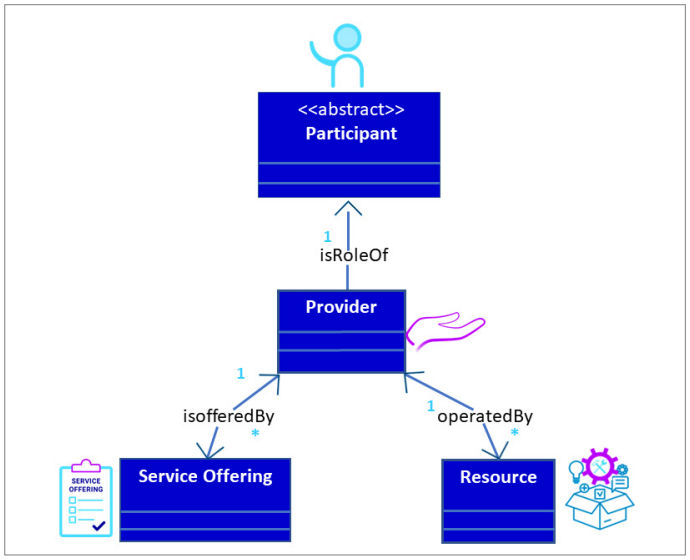

Fig. 3: Taxonomy of Providers

A Gaia-X participant is defined either as a legal person – including, but not limited to organisations or business entities – or as a natural person. Instances of a participant being neither a legal nor a natural person are prohibited. The class diagram in Figure 4 shows the respective section of the taxonomy below the participant entity.

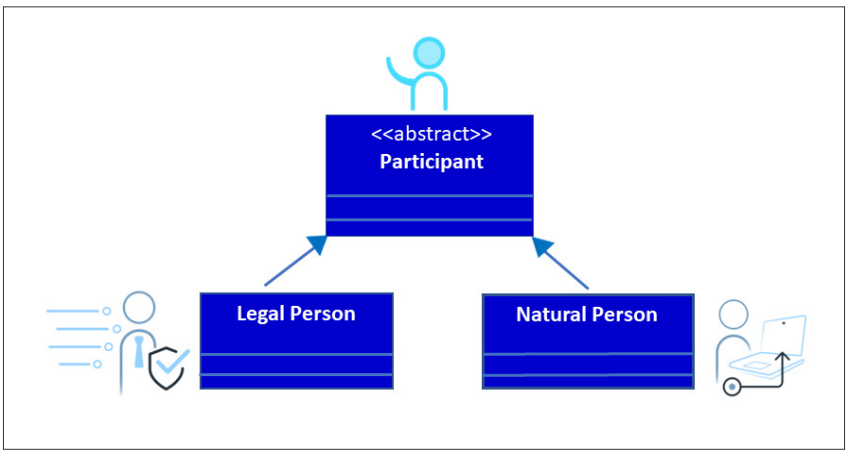

Fig. 4: Taxonomy of Participants

4.2 Service Offering

The world of digital services is heterogeneous and complex, as services differ in various points, such as their domain, functionality, requirements or policy restrictions. To handle this complexity, the Gaia-X taxonomy of Service Offerings is guided by the well-known pattern of Anything-as-a-Service (XaaS). Hence, the top-level subclasses of the service offering comprise software, platform and infrastructure. Infrastructure services provide computational capabilities, such as virtual or bare metal machines; platform services offer development and execution environments, such as web servers or online shops; and software services comprise all other services that do not offer infrastructure or platform capabilities, such as AI services. The class diagram in Figure 5 shows the taxonomy of service offerings.

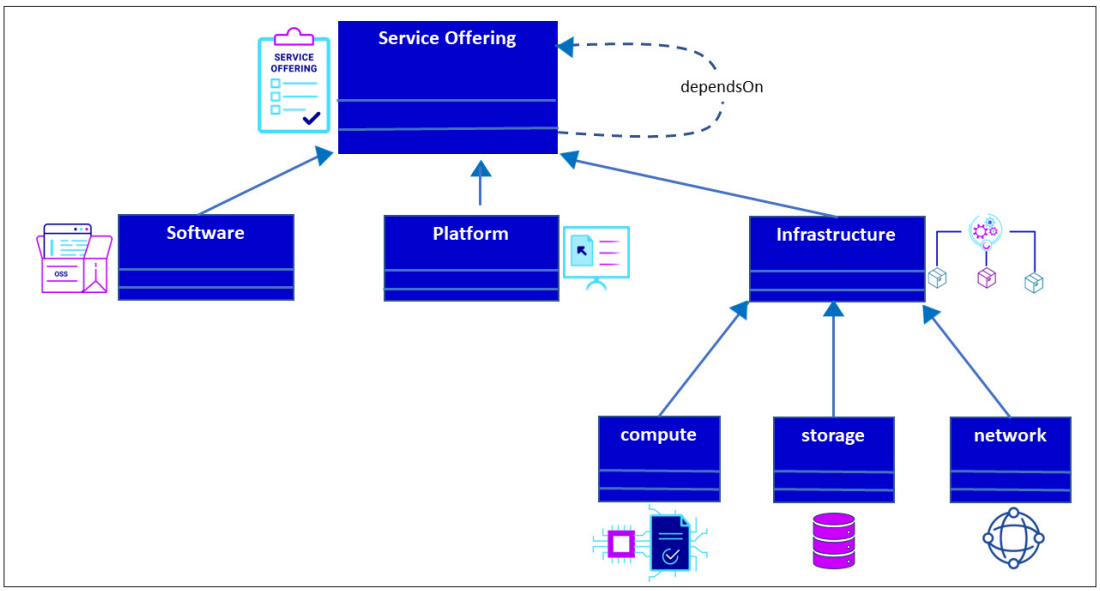

Fig. 5: Taxonomy of Service Offerings

4.3 Resource

Gaia-X defines a resource "… as an internal building block of service offerings". A resource is either a physical resource or a virtual one. Similarly, when it comes to the distinction between legal persons and natural persons for participants, Gaia-X does not allow any other type of resources. Resources can be composed by the other resources as well.

Physical resources have a weight and position in physical space, such as data centres, bare metal or appliances. Relevant properties of physical resources are the physical location and the resource owner. The owner of a resource is a person who is authorised to sell the resource.

Virtual resources, such as software or datasets, do not have an explicit physical location. However, they inherit this information from the physical resource on which they are hosted. The differentiation between physical and virtual resources is one important pillar of digital sovereignty. Knowing where data is stored or processed, by whom, and which national regulations may apply to data, increases transparency for consumers' decisions.

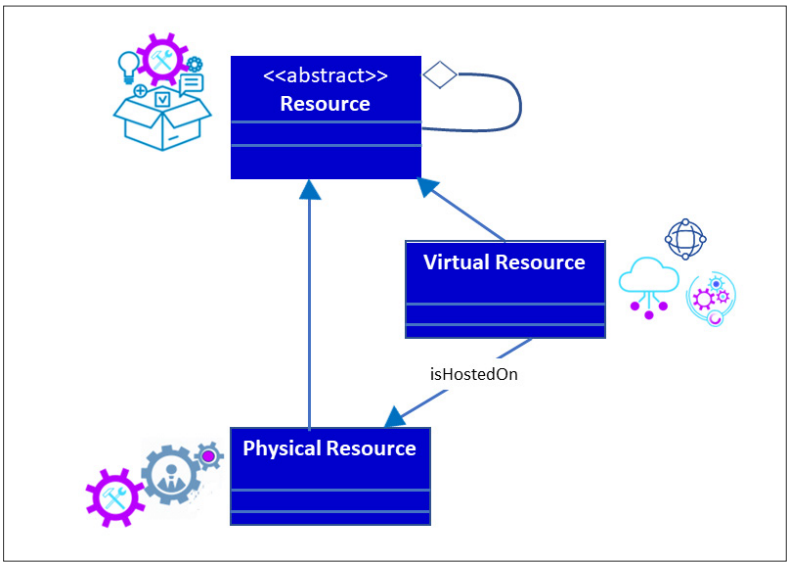

Fig. 6: Taxonomy of Resources

The properties of physical and virtual resources are heterogenous; hence, these classes are subclassed by a further layer. Data resources and software resources both specialise the concept of virtual resource. Gaia-X defines data resources as "...data in any form and including necessary information for data sharing", and software as a resource "... consisting of non-physical functions". For both data and software resources, the attributes comprise "copyright owner" and "license" to govern usage and access restrictions.

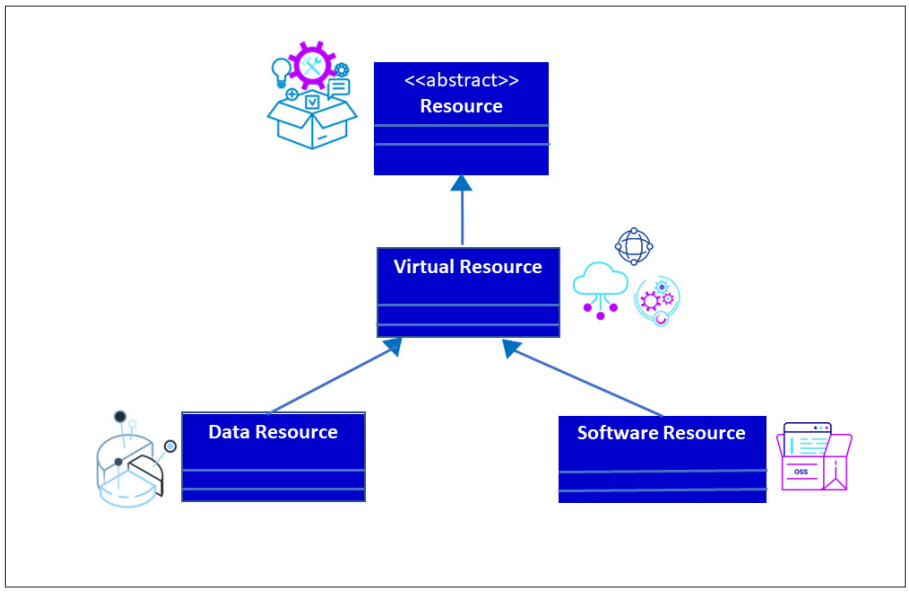

Fig. 7: Taxonomy of Virtual Resources

The operation of services requires two additional specialist subclasses of resources: nodes and interconnection. Nodes "… represent […] a computational […] entity that hosts, manipulates, or interacts with other computational […] entities". Interconnections "… include […] details of the connection between two or more nodes". Both classes may be virtual as well as physical; for this reason, the taxonomy distinguishes between virtual nodes such as virtual machines or containers, physical nodes such as bare metal machines, and virtual interconnections, such as VLANs, as well as physical interconnections such as optical fibres. Figures 8 and 9 show the taxonomy of resources.

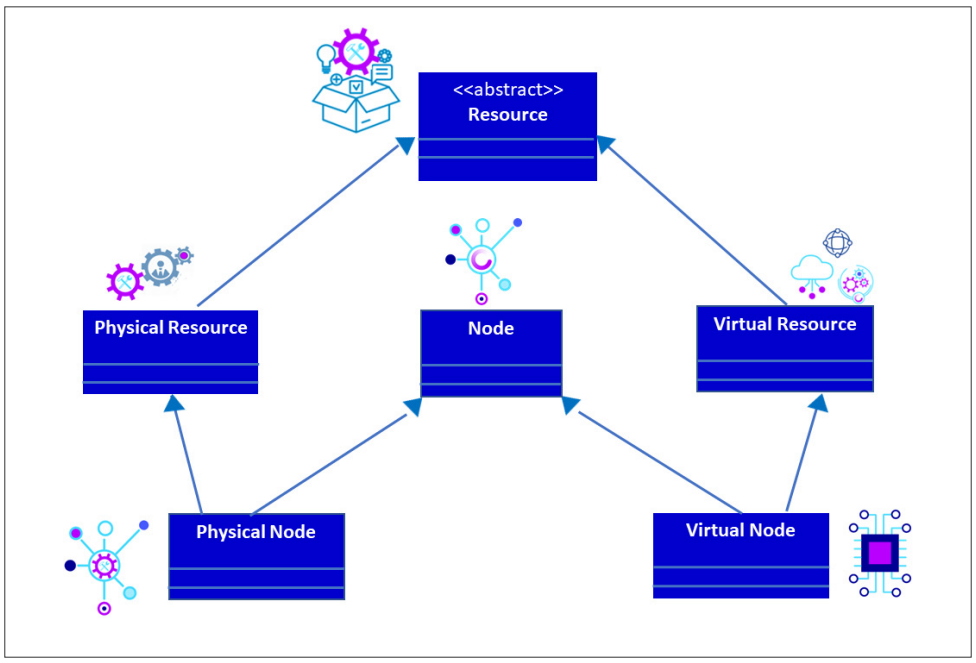

Fig. 8: Taxonomy of Nodes

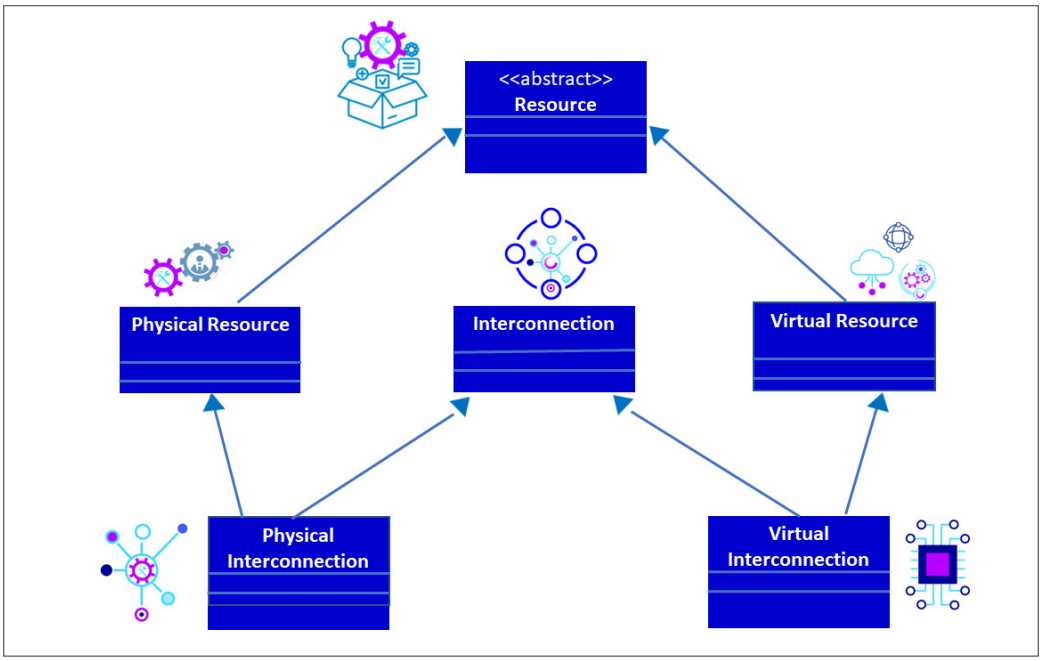

Fig. 9: Taxonomy of Interconnections

The following example illustrates the taxonomy in more detail. Think of a company called ACME that wants to offer a web shop as a service (WSaaS) in Gaia-X. The web shop is based on a software framework called webShopFram, version 23.05. To separate clients and their data, ACME creates a virtual machine for each client exclusively and operates these machines on top of their own data centre located in Germany.

In this example, ACME is a **provider**, offering a service of a **platform** nature. WSaaS, as a composed service, is aggregated from three **resources**: two nodes, a virtual and a physical one, and a software resource. The data centre is represented as a **physical node**, whereas the virtual machine is represented as a **virtual node**. Finally, the webShopFram (web shop framework) is represented as a **software resource**. According to the Conceptual Model, each taxonomy element identified above requires a Self-Description to be written. As such, there are five Self-Descriptions in total:

(1) ACME (provider)
(2) WSaaS (platform service offering)
(3) data centre (physical node)
(4) virtual machine (virtual node)
(5) software framework webShopFram (software resource)

Figure 10 depicts these Self-Descriptions as a UML object diagram, including some exemplary Self-Description attributes. A first definition of all attributes of Self-Descriptions within the three-level taxonomy is still under development and outside of the scope of this Whitepaper.

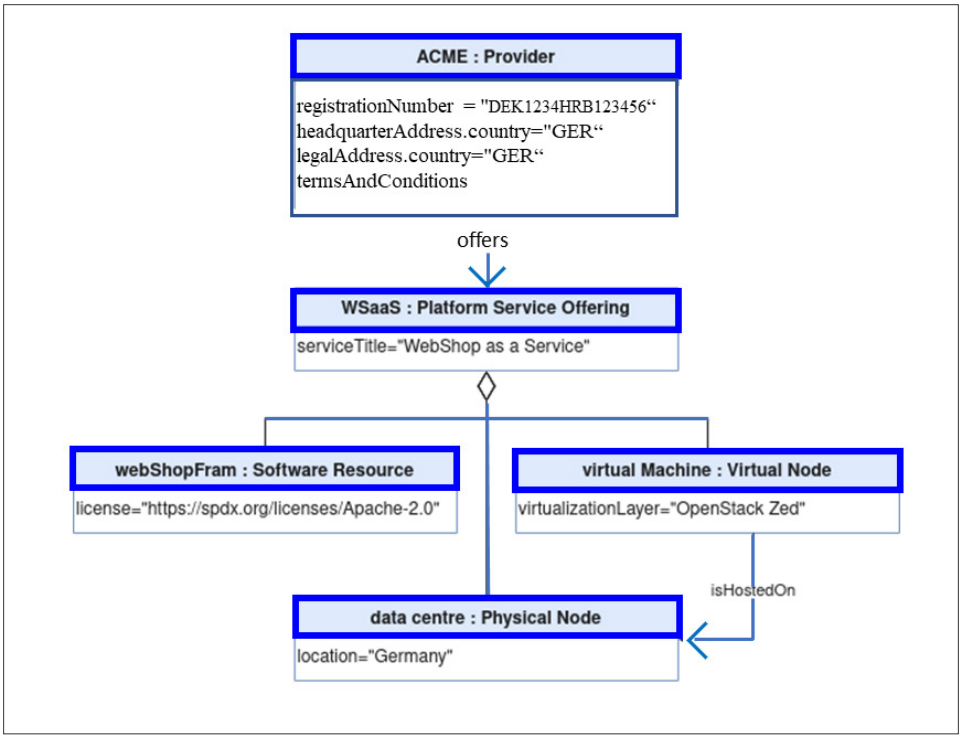

Fig. 10: Object Diagram of the Example

5. Governance Process

The Gaia-X Conceptual Model is defined by the Gaia-X Technical Committee Working Group Architecture. The taxonomy of the provider, service offering and resource classes introduced above, including the attributes for all classes, is defined by the Gaia-X Technical Committee Working Group Service Characteristics. The Gaia-X Policy and Rules Committee Working Group Compliance decides on whether an attribute should be mandatory. Decisions in these Working Groups are taken on the basis of four steps:

1. A member writing down a proposal for a change to one of the Working Group's deliverables – human-readable documents or technical representations of taxonomies or schemas,
2. The proposal being presented and defended in one of the Working Group's regular meetings,
3. Other, independent members approving the proposal after discussion,
4. The decision finally being included in the deliverable.

The history of any changes proposed and decisions taken is archived transparently. Gaia-X currently relies on merge requests in the GitLab version control software for this purpose. Any member of a Gaia-X member organisation can get on-boarded to these Working Groups.

6. Practical Tool Support

Once agreed upon, according to the governance process explained above, the Self-Description Schema, comprising taxonomy and attributes, needs to be made available in two versions: a human-readable documentation helps providers to understand how to write Self-Descriptions according to the best practices, and a machine-comprehensible representation supports the automation of tasks such as querying and

validation. The vocabulary of the queries posed, for example, to a Federated Catalogue, is represented in an ontology, which, for instance, gives technical identifiers such as http://w3id.org/gaia-x/resource#Node to concepts, and defines the taxonomy and attributes in a formal way. Technically, this ontology follows the W3C Resource Description Framework (RDF) Schema and Web Ontology Language (OWL) standards. To check whether the claims in a Self-Description follow all constraints, such as including all mandatory attributes, the claims are validated against a shape. Technically, these shapes follow the W3C Shapes Constraint Language (SHACL). The claims themselves are represented as an RDF graph, serialised in the W3C JSON-LD format, where JSON is a data interchange format widely supported by programming languages, and JSON-LD (LD = "linked data") makes it compatible with RDF. In addition, in order to gain the expected level of trust in the JSON-LD expressed claims, the W3C Verifiable Credentials Data Model standard in conjunction with the W3C Decentralised Identifier concept is used.

Figure 11 shows how the Working Group Service Characteristics has implemented its governance process by a technical toolchain, so that most of the agreed-upon definitions only have to be maintained in one central place, from which all the other artefacts – facing non-technical end users, domain experts or application developers – can be generated automatically. The technical Self-Description Schemas are used by a variety of further tools: for example, a creation wizard with interactive web forms which guides non-technical end users towards writing valid Self-Descriptions, or libraries that enable software to read, write and process Self-Descriptions.

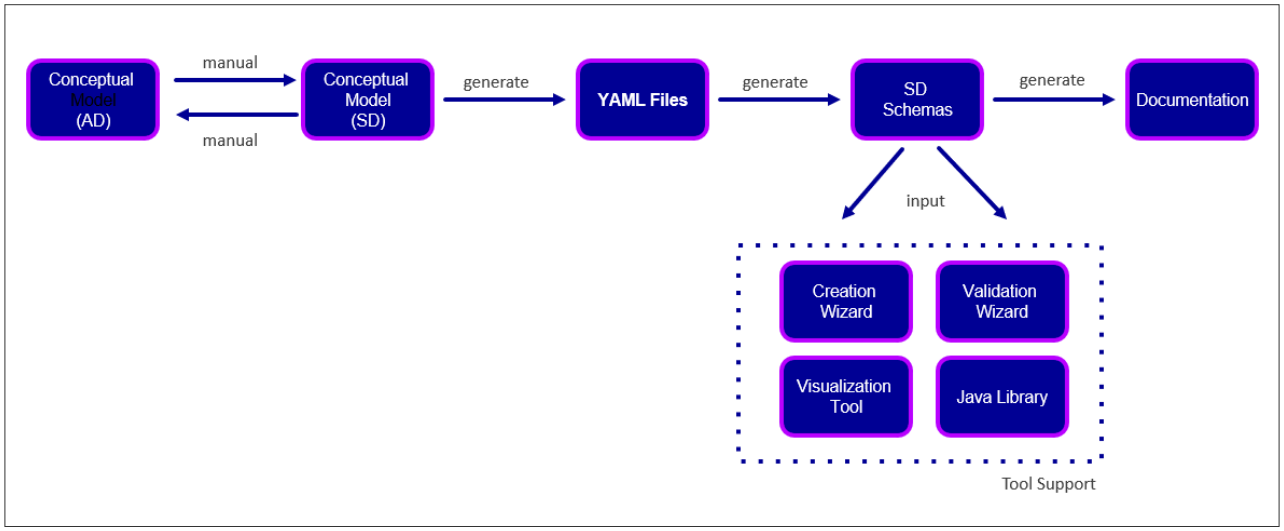

Fig 11: Automatic Generation of Technical Schemas and Documentation for Self-Descriptions

In this manner, this toolchain ensures that the claims in Self-Descriptions are always syntactically and semantically valid, and that most steps of the process are automated – from capturing an expert's understanding of how a certain class of services should be described, to the creation of Self-Descriptions of such services. Today, no single step of this toolchain – for example, generating documentation for a schema or validating a metadata input form – is rocket science, but is possible by simply applying stable, state-of-the-art tools. The innovation is in the integration of all of these parts into a compound solution that addresses all needs of those who create Gaia-X Self-Descriptions. Full coverage of the lifecycle of Self-Descriptions will be provided by an integration of the claim-focused toolchain available so far with the Notarisation Service, which issues Verifiable Credentials backed by trust anchors, implementations of the Gaia-X Trust Framework such as the Trust Services for their verification, and the Federated Catalogue for making Self-Descriptions discoverable and open to query. These further integration steps are scheduled for 2022.

7. Challenges

In the work with Gaia-X Self-Descriptions, we are facing three main challenges. Here is how we have identified and addressed these so far.

- Regarding **technological gaps and barriers**: for the Self-Description of simple, ubiquitous concepts – such as a company that acts as a provider, or a data resource that is composed into a service offering – ready-to-reuse schemas exist: for example, the W3C Data Catalogue Vocabulary DCAT. For describing the more technical infrastructure in a Gaia-X ecosystem, such as nodes or interconnections, standard description languages or communication protocols exist, but they lack semantic interoperability with the more abstract layers of data and AI. For AI services and software, on the other hand, semantically interoperable descriptions exist, but mainly in research prototypes. The nucleus of the Gaia-X Self-Description Schemas is therefore created by integrating mature web standards, but we enable service experts to connect their existing standards and to capture relevant service characteristics into extensions of that nucleus.
- Regarding **acceptance and adoption rate**: the level of interoperability, the extensibility to additional concepts from application domains, and the formal correctness that we need requires the usage of standards from the W3C Semantic Web and Linked Data family. People who just want to focus on developing and providing services do not necessarily like the cognitive overhead that these standards entail. From these standards, we selected exactly what is needed for Gaia-X and try our best to shield the inherent complexity from those who just want to get their job done.
- Regarding **uncertainties**: in capturing and representing knowledge about the world – in this context, this is about about providers and their service offerings – using a formal language may lead to over-engineered schemas that are too complex for describing 90 per cent of plain, everyday factors. On the other hand, enforcing too little complexity would not allow consumers to compare available service offerings by queries against a Federated Catalogue in a meaningful way, or to automate the composition and orchestration of multiple services. It will be a continuous challenge to strike the right balance.

8. Next Steps & Conclusion

The scope of this Whitepaper was devised to present an overview of the current state of Gaia-X Self-Descriptions. Section 1 explained the role of Self-Descriptions within Gaia-X ecosystems. Section 2 described trust in Self-Descriptions, followed by an introduction to the Gaia-X Conceptual Model in Section 3, and the further taxonomy of provider, service and resource classes in Section 4. This Self-Description taxonomy is just a first approach and not yet ready for all industrial applications. Three main requirements to a Self-Description taxonomy and schema are still unsolved: (1) configuration model, (2) release management and (3) extension points.

A **configuration model** allows parametrisation of Self-Descriptions for service offerings. Currently, the Self-Description Schema provides the vocabulary for describing fixed service offerings. Customers have no choice but to configure the offer according to their needs, such as selecting the number or size of virtual machines within an IaaS Offering, unless the provider created an exponential number of Self-Descriptions explicitly representing all possible configurations. One of our next steps will be to develop a model to support parameterised and configurable service offerings.

The Self-Description Schema, including the taxonomy and attributes, is still under development and will most probably undergo further changes in the near future. To guarantee upwards compatibility of older Self-Descriptions instances with newer schemas, the definition of a **release management** process will be an additional next step.

Gaia-X aims at defining general principles for data and service ecosystems, regardless of the specific domain of application, while fostering the application of these principles in any concrete domain. For Self-Descriptions, this means that the taxonomy of provider, service and resource classes defined by the Working Group Service Characteristics will remain limited. However, it should come with clearly defined **extension points** and a documentation of best practices on how to extend it for application-specific purposes. We will do so by following the principles of application profiles, which is common in information science. The Dublin Core Metadata Initiative defines an application profile as “a metadata design specification that uses a selection of terms from multiple metadata vocabularies, with added constraints, to meet application-specific requirements”.

9. Additional Material

Website: www.gxfs.eu

Authors:

Anja Strunk (Cloud&Heat Technologies GmbH)

Anja Strunk is Lead of Cloud Innovation Team at Cloud&Heat Technologies GmbH. She is the Interims Vice-Lead of Working Group Service Characteristics (previously known as Open Work Package Self-Description), and she is mainly involved in the definition of the current Self-Description Schema.

Dr. Christoph Lange (Fraunhofer FIT)

Christoph Lange is Head of the Data Science and Artificial Intelligence department at the Fraunhofer Institute for Applied Information Technology FIT. He is Lead of the Gaia-X Working Group Service Characteristics.

Association Name: Gaia-X European Association for Data and Cloud AISBL
Contact Person: Dr. Vassilia Orfanou, CMO
Phone: +306974825327
Email: vassilia.orfanou@gaia-x.eu
Address: Avenue des Arts 6-9, 1210 Brussels, Belgium
Website: www.gaia-x.eu

Company Name: eco – Association of the Internet Industry
Contact Person: Mareike Zeisig, Marketing Manager
Phone: +49 (0) 221 - 70 00 48 - 107
Email: mareike.zeisig@eco.de
Address: Lichtstr. 43h, 50825 Köln, Germany
Website: www.gxfs.eu

Supported by:

Federal Ministry
for Economic Affairs
and Climate Action

on the basis of a decision
by the German Bundestag