Governance, Architectures and Business Models for Data and Cloud Federations: the EOSC and GAIA-X Case Studies[1]

Version 0.5, 10-06-2021

Contact authors:

Mark Dietrich, EGI Foundation (Mark.Dietrich@egi.eu)

Tiziana Ferrari, EGI Foundation (Tiziana.Ferrari@egi.eu)

Abstract

Data and tools to realize the data value chain are enablers of digital transformation in all sectors of the economy and society, boosting economic development, increasing the ability to tackle societal challenges and excel in science. To realize this, collaboration, across business sectors, scientific domains and organizational boundaries, countries and continents, is widely seen as essential to innovation, productivity and discovery. Federation of data and services coupled with interoperability frameworks for compliance to federation-level policies, technical standards and service delivery processes, have been successfully adopted in many domains. This paper compares governance, architectures and models of federation for cloud-based data exploitation focusing on two use cases: the European Open Science Cloud and GAIA-X, the Federated Data Infrastructure for Europe.

[1*] This work was supported by the European Commission H2020 grant number 101017567 EGI-ACE (https://www.egi.eu/projects/egi-ace/), and H2020 grant number 871920 H-CLOUD (https://www.h-cloud.eu/).

Contents

Executive Summary

Various federation models have been successfully demonstrated in sectors where data sharing produces common benefits to the members of the federation. There have been many efforts over decades to facilitate collaboration, from development of networks and communities of practice, to creating technology platforms and formal business agreements. In recent years, discussion has turned to the opportunities and challenges associated with exchanging data between collaborating parties. The document explores two major initiatives for cloud-based data exploitation in Europe: the European Open Science Cloud (EOSC) and GAIA-X.

EOSC federates existing and emerging research data infrastructures with the objective of offering a virtual environment in Europe to share and re-use research data across borders and disciplines. On the other hand, GAIA-X aims at creating a proposal for the next generation of a European data infrastructure as a secure, federated system that meets the highest standards of digital sovereignty while promoting innovation.

Both EOSC and GAIA-X describe themselves as federations: they are alliances of multiple organizations, where the participating organizations collaborate for common goals and agree to conform with various technical standards and operating procedures that enable interoperation, collaboration and sharing. Each initiative has a federating entity at its core as a virtual or a real organization separate from any member, and participation can involve a degree of sharing resources.

This study highlights how both EOSC and GAIA-X need to elaborate their respective business models and their solutions to the problem of sustainability, including service delivery models and economic models for the federating entity and the participants. Models that have been successfully adopted in existing federations include the Open or Structured Marketplace, Reseller, Assembler, One-Stop Shop and Full Integration.

With regards to compliance, both initiatives recognize the need to establish common “rules of engagement” for participants in their respective initiatives. These rules need to align with technical solutions but primarily focus on semantic, organizational and legal topics. The two sets of rules both emphasize transparency, but otherwise address largely complementary issues.

Various initiatives are tackling the issue of data and service integration in an ecosystem. Besides EOSC and GAIA-X, Destination Earth addresses the problem of data system interoperability, advocating to minimize data movement by relying on a federated cloud and data architecture where data can be discovered remotely and processed locally at the system providing data access. Both EOSC and GAIA-X seek to integrate both data and infrastructure, enabling safe and secure access to relevant data, and access to a range of relevant data processing resources and systems, enabling the creation of new insights and knowledge and support for data-driven activities and business. Against this vision, both initiatives must contend with practical limits to this integration: (1) the responsible “owners” of both data and infrastructure may need to limit both the visibility and the use of their data and infrastructure, (2) owners may have their own reasons to want to establish similar limits on visibility and use, and (3) even when owners are interested in making their data and infrastructure widely visible and usable, effort is usually needed to prepare that data and infrastructure for listing and use.

Regardless of the reasons, this variation in visibility translates into the need to create separate environments for data and infrastructure that allow visibility and use to be limited within a given environment and that allow entity owners to make sovereign decisions about to which environments

they wish to expose their data and infrastructure. This is a strong requirement in both the corporate and the research domains. In the EOSC 'system of systems' coexisting communities and research projects with service and/or content provider role, retain the ability to make decisions on various topics, from visibility, use and access policies, however these share the opportunity to leverage common services. We can identify each of these separate environments as a data ecosystem that will be supported through EOSC.

Similarly, in GAIA-X resources and assets compose service offerings, the primary entities that will be listed in Federated Catalogues. Self-asserted descriptions define certifications and other non-functional attributes appropriate to each type of entity. The GAIA-X Architecture does not refer to the FAIR Data Principles, but its architectural approach aligns with those principles. A Federation in GAIA-X is defined as a loose set of interacting actors that directly or indirectly consume, produce, or provide Assets and related Resources; however, federations are not defined as entities within the GAIA-X Conceptual Mode.

Both EOSC and GAIA-X services are grouped into services for the federation participants and services for the end-users, however they put emphasis on different components. The first group includes similar functionalities such as: discovery and selection, ordering and procurement, usage control, monitoring, accounting and metering, interoperation, trust and identity, consultancy and support. GAIA-X has included a "sovereign data exchange" service in its suite of federation services since its inception. Both in EOSC research data and GAIA-X industrial data participants retain control on what data is shared and with whom. For both bilateral and multilateral data sharing, it will be important to establish clear policies, and possibly provide supporting technology, to track Use Conditions when multiple data objects are processed together. Although multi-provider service provision is an explicit objective of GAIA-X, the initiative has not yet proposed support for Service Coordination, Integration, or Operations Management.

The paper concludes by recommending collaboration in data sharing, compliance, governance and technical interoperability to:

(1) Share good practices on cross-domain/organization data sharing policies and FAIR best practices that enable data exploitation and validate these with concrete cross-domain and cross-initiative use cases.
(2) Adopt compatible models for data ecosystem business models, federation architectures, and service integration and delivery models.
(3) Align the definition of technical and operational interoperability frameworks that make data exchanges between EOSC and GAIA-X possible and share experience on existing technical solutions.

1 Introduction

Collaboration across organizational boundaries, countries and continents, is widely seen as essential to innovation, productivity and discovery. Collaboration in research has been encouraged for decades, and international collaboration is the hallmark of key scientific efforts, from the discovery of the Higgs Boson to the mapping of the human genome. In research it has been recognized that research data that is FAIR ("Findable, Accessible, Interoperable and Reusable"), play an essential role in enabling Open Science and accelerating research workflows [1]. Likewise, collaboration in business has been identified as critical to superior business performance and sustainability, offsetting natural incentives to compete [2]. Collaboration and integration in sectors such as mobility [3], energy [4], health, agriculture [5], finance, and sustainable development [6] have been identified by the European Commission as essential to enabling improved efficiency and effectiveness in those sectors while also enabling achievement of the EU's policies on reduction of greenhouse gas emissions, and increased resource sustainability [7].

Both EOSC and GAIA-X are defined as federated systems. Various federation models have been successfully demonstrated in sectors where data sharing produces common benefits to the members of the federation. There have been many efforts over decades to facilitate collaboration, from development of networks and communities of practice, to creating technology platforms and formal business agreements. In recent years, discussion has turned to the opportunities and challenges associated with exchanging data between collaborating parties, enabling multi-lateral data sharing within communities of practice, and creating "open data" that can be widely accessed and used to power data-driven business and discovery and create good for society.

This document explores two major initiatives for cloud-based data exploitation in Europe: the European Open Science Cloud (EOSC)[2] – focusing on research, and GAIA-X[3], the Federated Data Infrastructure for Europe – driven by industrial use cases.

There are differences between the initiatives, but also striking similarities in areas of governance, technology and architecture and policies. Similarities and differences will be explored in these and related areas. For cloud-based data exploitation in EOSC the technical developments of the EOSC related projects are also considered. The analysis will be further expanded by considering additional approaches emerging from the EOSC thematic cloud projects.

1.1 EOSC

EOSC is defined as the web of data and services that '*will allow researchers to find, exploit and combined linked datasets, providing a basis for artificial intelligence (AI) tools leading to new discoveries and research paradigms*' [8]. EOSC '*federates existing and emerging research data infrastructures with the objective of offering a virtual environment in Europe to share and re-use research data across borders and disciplines. EOSC is expected to serve approximately 2 million researchers in Europe and progressively expand its user base to include the wider public sector and the private sector*' (business organisations) [9].

[2] https://www.eosc.eu/

[3] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

An EOSC that offers added value to researchers was taken as a starting point with its scope as described in the Strategic Implementation Plan [10]: *"the EOSC should be a federation of existing and planned research data infrastructures."* EOSC responds to the need to integrate efforts on findable, accessible, interoperable and reusable data with the provisioning of data-oriented services of research infrastructures and e-Infrastructures [11].

The European research infrastructure landscape includes networking, computing facilities and data infrastructures, which rely on the federation on national, regional and institutional Infrastructures. It should be noted that the pan-European e-Infrastructure services are often being provided by national e-Infrastructures in a collaborative manner, and the European initiatives are dependent on the existence of strong, persistent and synergic national infrastructure nodes.

EOSC intends to operate across the European research community and to expand to the wider public and private sectors. EOSC is explicitly designing its architecture to support dozens of existing, productive scientific and research communities that want the benefits of working with EOSC, but which remain hesitant to change the way they operate or cede control to EOSC. In this sense, EOSC is a system of systems, a federation of federations.

1.2 GAIA-X

As stated by the most recent GAIA-X Architecture Document [12] "*The GAIA-X initiative, gathering representatives from business, science and politics on a European level, aims at creating a proposal for the next generation of a European data infrastructure as a secure, federated system that meets the highest standards of digital sovereignty while promoting innovation. The mission of GAIA-X is to design and implement a data sharing architecture that consists of common standards for data sharing, best practices, tools, and governance mechanisms. It also constitutes an EU federation of cloud infrastructure and data services*".

GAIA-X addresses "stakeholders from industry, the public sector and science" and is consulting with multiple industry sectors, for example automotive manufacturing, transport, energy and financial services, in order to meet their needs. GAIA-X is working to clarify its value proposition for both individual organizations and for ecosystems that operate in those sectors. Within each sector there are multiple business ecosystems, each with multiple participants, that could potentially use GAIA-X as a common technical foundation, but which also have concerns about giving up any autonomy or control to GAIA-X. Like EOSC, GAIA-X might be regarded as a system of systems or a federation of federations.

2 Organization and Governance

Both EOSC and GAIA-X describe themselves as federations and exhibit the essential characteristics of federation, which the Horizon Cloud project[4] studied in [13] highlighting the following essential characteristics:

- *A federation is an alliance of multiple organizations.* This means that a federation is a collective entity that is not "owned" by any single organization.

[4] H-CLOUD, h-cloud.eu

- *Participating organizations are "members" of the federation and collaborate for common goals*. In many federations, members "join" the federation by acknowledging its common goals and agreeing to collaborate with other members, participate in governance and abide by agreed standards, policies and procedures.
- *Each federation has a "federating entity" at its core that can be either virtual or a real organization separate from any member*. The federating entity supports federation governance, collaboration and a range of possible coordination activities agreed by the members.
- *Members agree to conform with various technical standards and operating procedures that enable interoperation, collaboration and sharing, appropriate to the type and purposes of the federation*. Agreeing on these standards and procedures is enabled by agreement on federation goals and governance. An important subject of federation governance is the very process of agreeing to these standards and procedures.
- *Participation can involve a degree of sharing resources (including services, data, metadata or other assets)*. At minimum this requires members to make their shareable resources discoverable and accessible to other federation members. Cloud federation is often seen as a mechanism for sharing physical IT infrastructure with other federation members, but such a shared approach is not a universal feature. Service and data interoperability is a more common feature, and this depends both on agreement to technical standards and procedures and on a willingness to make those services and/or data available through the federation.
- Federations typically have *two service portfolios*, an internal that benefits the members and keeps the federation running, and an external, which faces shared customers.

Both in EOSC and GAIA-X a marketplace plays the important role of service and resource registry and European delivery channel towards the customers.

2.1 Organization

Federations are often composed of 'members' and a 'federating entity' that is tasked to coordinate the activities of the federation. Governance models are usually participatory to ensure the involvement of federation members in decision making. Formal aspects of governance, e.g., nominating and voting for directors, are usually legally defined for members as part of the statutes of the federating entity, as are requirements for membership.

This paper refers to individuals and/or organizations that participate in the work of the federation as 'participants' (participants may or may not be members of the federation). While requirements for membership are defined in the statutes and by-laws of the federating entity, requirements for participation are defined as policies of the federation. Note that participants do not have any role in the governance of the federation, unless they join it as members.

The federating entity may take responsibility for certain tasks beyond governance and corporate activities. These "core" tasks can be performed by the federating entity, delegated/subcontracted to one or more organizations (which may be members and/or participants), and/or performed collectively by members/participants. Table 1 below compares how each initiative is organized.

Table 1 Organizational comparison of EOSC and GAIA-X

Organization	GAIA-X	EOSC
Federating Entity	GAIA-X Foundation AISBL[5]	The European Open Science Cloud Association (abbreviated EOSC Association) is a Belgian AISBL [14]. EOSC Association participates in a European Partnership [15].
Birth of formal organization	2021	2020
Execution	Executive team plus staff. Policy and Rules Committee; Technical Committee; Data Spaces Business Committee; Working Groups of Members	A Secretariat is responsible for the day-to-day management, for the implementation of decisions and the performing of administrative acts. It advises and supports the General Assembly and the Board of Directors.
Members	Membership is open to legal entities (such as corporations or associations) incorporated in any jurisdiction, as well as to Member States of the EU. Membership is limited to just one entity from a group of corporations controlled by an "ultimate parent company"[6].	Members and Observers can be research funding organizations, research performing organisations, service providing organisations or belong to 'other' categories. A Member or an Observer is a legal entity established in accordance with the laws and customs of the country of origin or be constituted as an intergovernmental organisation pursuant to an international treaty in accordance with principles of international law. They cannot be a department of national governments or ministries.
"Categories" of Members	Members headquartered in an EU Member State ("European Members") can each nominate 1 candidate Director for the Board.	Member States may appoint one Member to act as its Mandated Organisation, to represent national interests. Observers may send Representatives to attend the General Assembly, but they cannot vote on any matter.

[5] 'Belgian Association Internationale Sans But Lucratif' (an international not-for-profit association).
[6] With a specific exception described in the Articles of Association.

Number of Members	Over 200 and growing as of May 2021	A growing number of organizations has been joining EOSC, which in Q1 2021 includes about 200 members.
Participants	Participation requirements being defined by GAIA-X Policies and Rules Committee. Initial draft in the "Policy Rules and Architecture of Standards" [16], with a new draft [17] in consultation as of April 26, 2021. GAIA-X defines Participants as having one or more roles (at the same time): Provider, Consumer, Federator.	EOSC is based on the principle of openness in matters that concern user and resource participation. EOSC is open to any user with a focus on research in Europe. Resources federated in EOSC are listed in a publicly accessible registry which complies with policy requirements. Services have to align with the EOSC architecture and interoperability guidelines [18].
"Core" Tasks: Coordinating Activities Beyond Governance and Corporate Functions	GAIA-X intends to participate in various national and EU funding initiatives, notably one or more Important Projects of Common European Interest (IPCEI) as well as (currently) funding calls from the German and French governments. GAIA-X currently intends to fund the development of initial open-source versions of planned Federation Services (further detailed under Architecture). GAIA-X is considering operating one or more Federation Services, acting as either an initial "root of trust" for the federation, or possibly providing common capabilities to augment ecosystem-specific services that might be available only to ecosystem members. GAIA-X intends to select/define a variety of compliance standards that would apply to Participants and Services/Assets, and GAIA-X would delegate certification of compliance to competent organizations or authorities.	The Association may form operational bodies such as working groups and task forces to address issues or concerns relating to the mission and operations of the Association. Advisory bodies may also be formed to provide advice on fulfilment of the Association's mission.

2.2 Governance

A unique subject of federation governance is the very process of agreeing to the standards and procedures with which participants (many of whom are or will be members) will be asked to comply. Both the GAIA-X Association and EOSC Association are implementing new processes associated with such collaborative decision-making. The table below compares how each initiative is governed.

Table 2. Comparison of governance models in EOSC and GAIA-X

Organizational Aspect	GAIA-X	EOSC
General Assembly	The General Assembly (GA) is the supreme authority of the Association and is composed of one Delegate per Member with voting rights.	The General Assembly (GA) is the supreme authority of the Association and is composed of one Delegate per Member with voting rights and one Representative per Observer without voting rights.
Responsibilities of the General Assembly	The GA decides upon statute amendments, adoption and amendment of bylaws, the appointment and discharge of Directors and the auditor, the approval of the annual budget and accounts, the dissolution of the Association, the admittance and termination of Members and the fees.	The GA decides upon statute amendments, adoption and amendment of bylaws, the appointment and discharge of the Board and the auditor, the approval of the annual budget and accounts, the dissolution of the Association, the admittance and termination of Members and Observers and the fees.
Board of Directors	The Board prepares the meetings of the GA, the budgets and the annual accounts, proposes amendments, provides recommendations and directs the implementation of decisions. The number of Directors will be set at the first GA scheduled for June 7.	The Board (a minimum of 7 members) prepares the meetings of the GA, the budgets and the annual accounts, proposes amendments, provides recommendations and directs the implementation of decisions.
Selection of Directors	The GA appoints Directors from amongst the candidates nominated by European Members and from the candidates for Independent Board Member proposed by the Board	The GA appoints Directors from amongst the Delegates of the GA.

	itself.	
Board	The Board is composed of Directors, who are appointed in their individual capacity. Although Member-nominated Directors are appointed in their individual capacity rather than as a representative of that Member, their holding a seat on the Board is subject to that Member's continuation as a Member in good standing of GAIA-X.	The Board is composed of Directors, who are appointed in their individual capacity,
Other governance bodies	Two bodies provide advice to the Board: Government Advisory Board composed of representatives of the European Member States that are Members of GAIA-X, and a General Advisory Board composed of experts and stakeholders.	The Association may form operational bodies such as working groups and task forces to address issues or concerns relating to the mission and operations of the Association. Advisory bodies such as committees may also be formed to provide advice on fulfilment of the Association's mission. Complementing this, in the context of the EOSC co-programmed European Partnership, a Partnership Board defines the cooperation in the Partnership. A Steering Board including delegates from Member States and countries associated with the Horizon Europe Framework Programme of the EC, plays an advisory role to the EC and the Partnership Board.
External affiliated organizations	National GAIA-X Hubs have formed, and are being formed, in at least 9 countries across Europe. National Hubs parallel the objectives and activities of GAIA-X itself, and their activities are currently coordinated with GAIA-X at the executive level.	Observers have no voting rights in the GA and cannot propose candidates for the Board
Decision- making processes	Most of GAIA-X' activities to date have focussed on leading a community-based process to identify	GA Members have voting rights and the GA deliberates according to rules that define a quorum. The GA strives

	and develop consensus on policies and standards for adoption by GAIA-X. Formal document management tools (such as GitLab) have been used to manage the development of high-level architecture, service specifications, and related policies. A transparent and consultative approach is employed, although the federation has yet to formally adopt any policy or specification (e.g. through the Board or GA).	to adopt its decisions by consensus. Policies and the interoperability framework [18] are a community effort coordinated by a dedicated group. The framework defines technical, semantic, organizational and legal interoperability.

2.3 Business models and sustainability

Both GAIA-X and EOSC must elaborate their respective business models and their solutions to the problem of sustainability. Several aspects of this topic must be defined:

- The "service delivery model" of the federation.
- The “economic model” for the services delivered by federation participants to end users.
- The “economic model” for the activities of the federating entity, including the common functions or services essential to the operation of the federation.

Horizon Cloud’s Cloud Federation report [13] identifies several alternative service delivery models for cloud federations, ordered by the degree to which offered services need to be coordinated and integrated to meet a customer’s needs. These models rely on two important roles: the Marketplace Operator and the Integrator.

The Marketplace Operator is responsible for the management of a marketplace platform that supports various functions: ‘(a) matching buyers and sellers; (b) facilitating the exchange of information, goods, services and payments associated with market transactions; and (c) providing an institutional infrastructure, such as a legal and regulatory framework’ [7].

The Service Integrator can be defined to be a ‘single, logical entity held accountable for the end-to-end delivery of services and the business value that the customer receives. This removes the requirement for the customer to concern itself with the management overhead of looking after the complex web of service providers. However, it relies on the customer empowering the service integrator and giving it the responsibilities of day-to-day coordination and control of service providers’[8].

1. *Open Marketplace*. Services are discoverable and accessible through a marketplace run by the Marketplace Operator, who does not limit how services can be listed and is not responsible for the provisioning of such services.

[7] https://www.dsi.unive.it/~marek/files/02%20-%20e-marketplaces
[8] https://www.scopism.com/service-integration-a-different-service-management-art-form/

2. *Structured Marketplace*. Services can be discovered and accessed like in the previous case, but here the Marketplace Operator chooses the regulatory policies according to which services are listed and described. Such policies may be maintained by an external entity such as a federator. After selection, the customer is fully responsible for service integration.
3. *Reseller*. Services can be discovered and selected as in a *Structured Marketplace*, and integrated payment and contracting services are provided by the Marketplace Operator to enable procurement of the selected services. Procurement might include a variety of prepaid or non-monetary arrangements. The customer is then responsible for integration.
4. *Technical Integrator*. Services can be discovered, selected, and procured and the Marketplace Operator or an external entity then integrates some of the services. In this model the customer is still responsible for the end-to-end delivery for the complete package of selected services, as well as their ongoing management and maintenance.
5. *Full Integrator*. The Marketplace Operator provides a complete range of services, from selection to service integration, delivery and management, and performs the integrator role.

The scope of each initiative's service delivery model defines what remains to be done by the initiative's users or their communities/ecosystems before they can successfully use the initiative's services. A more limited scope of the initiative's service delivery model leaves more work for the customer. A greater scope requires less work from the user. This scope is a key factor in each initiative's value proposition for its users.

Once services are selected by a customer, many economic models are possible. In a commercial environment, typical options are "pay as you go" or contracted access to an agreed quantity of services for an agreed price.

In the research environment, economic models for service access and use are more complicated. The norm is "free at the point of use" access to resources that have been specifically funded for the purpose of serving researchers. The European Charter for Access to Research Infrastructures [19] describes and classifies the most commonly adopted economic models that are applied to European infrastructures to grant access to potential users from academia, business, industry and public services. These include:

1. *Policy-based access*: in this model resources are made available for free by a sponsor, typically a research funding organization that defines the conditions for access, for example limiting free access to the members affiliated to a specific scientific collaboration or to the researchers based in each country in case of nationally sponsored research.
2. *Excellence-driven access*: use of a resource is conditional to the ability of a research collaboration to be selected according to the excellence, originality, and quality of its research idea. Excellence-driven access is an example of policy-based access.
3. *Market-driven access*: use of a resource is conditional to the establishment of an agreement against payment of a fee.
4. *Wide access*: no restriction is applied to the use of a resource. This model is suitable for resources whose capacity is non-depletable, such as scientific data. This access model maximizes availability and impact on research.

In both the commercial and research domains, service providers must find sustainable economic models for the delivery of their services.

Finally, but crucially, both GAIA-X and EOSC must themselves find sustainable economic models for the delivery of their federating services, the range of common functions and services that will be essential to the operation of the federation[9]. Finding a sustainable "economic model" for such activities is challenging since these core services have an indirect, enabling and possibly regulatory nature that is hard to price, and since any fees or costs set for these services are often perceived as "taxes" by participants.

2.3.1 EOSC

2.3.1.1 EOSC Service Delivery Models

EOSC intends to deliver services using a mixed service delivery model which is not yet fully defined. The intent is that all services (and data) will be discoverable through EOSC's "marketplace" mechanisms. Some of these services are offered "free at the point of use", so users get immediate access to these services through the marketplace. Other services require additional authorization steps, which may be automated or may require some manual involvement by the service provider personnel. Of these services, some can be "assembled" (through orchestration or composition tools), and a range of service integration and service coordination services are also available.

The EOSC Portal was launched in November 2018 to aggregate, in a common catalogue, information about resources from hundreds of different organizations thanks to the adoption of common interoperable vocabularies by the participating providers. The portal offers users a European-level platform, the Marketplace, that aggregates supply and adds value by supporting discovery, comparison, ordering and secure access to listed resources. Ordering capabilities have been successfully demonstrated[10] as well as the ability to offer a central platform for order management, accounting integration and support through a common helpdesk.

EOSC provides service ordering and authorization capabilities that work with a range of procurement options appropriate to the research world. EOSC includes a hybrid procurement mechanism, called "virtual access", where financial compensation for a service provider is arranged in advance through European Commission grants, creating a "credit balance" that individual researchers can draw against to pay for specific service usage. This mechanism requires trusted usage metering and accounting, as well as trusted identity and authorization management. Virtual Access aims at promoting cross-organization and cross-country provisioning in those settings where services provided by publicly funded organisations may have a mandate and a budget to serve only a well-defined set of users (perhaps limited by research discipline or geographical boundaries), so additional funding is needed to enable wider access to those services. However, virtual access is limited by the availability of the projects which fund it, and is not currently a long-term model

2.3.1.2 EOSC-Exchange Service Economic Models

The definition of economic models and sustainability in EOSC is work in progress. The Sustainability working group of EOSC classifies EOSC capabilities and recommends the application of different

[9] For GAIA-X, which has indicated it does not intend to operate such services directly, equitable mechanisms for compensating any delegated service providers must still be found.

[10] https://opsportal.eosc-portal.eu/home

sustainability strategies for each class [9]. EOSC resources are divided into three areas: (1) the EOSC Core, (2) data and (3) the EOSC Exchange. The EOSC Core provides central capabilities for discovery, sharing, access and re-use of resources, while the EOSC Exchange delivers the services and the infrastructure for transferring, storing, processing and preserving research data. Different economic models can be applied to sustain the EOSC Core and the EOSC Exchange.

According to the EOSC study, different economic business models can coexist for the provisioning of the services of the Exchange. This is particularly suitable in multi-supplier environments, and it is a common approach adopted in distributed infrastructures where different federation members support heterogenous funding models and different access policies.

According to the Sustainability working group study, '*the scale and diversity of the services and resources to be federated implies that the operational and financial responsibility of federated services and data will remain with their existing operators and funders. The investment in federated services and resources by Member States needs to be measured and acknowledged as an in-kind contribution to the overall EOSC funding model*'. In addition, '*the funds for developing, operating and maintaining the services included in EOSC-Exchange is principally the responsibility of the service providers that operate them. Services made available via EOSC-Exchange may be available free of charge or against payment but remain free at the point of use*.

Specifically, the science clusters ENVRI-FAIR, EOSC-LIFE, ESCAPE, PaNOSC, SSHOC and the European e-Infrastructures EGI, EUDAT, GÉANT and OpenAIRE, in a joint position paper [20], claimed the need to sustain the costs of re-use of data by integrating data visualization, analysis and physical resources to store and re-use data for open science. EOSC should play a role in sustaining the cost of open data re-use; the paper advocates that the MS need to play a key role in the long-term service provisioning funding to avoid complex commercial transactions.

Ultimately, funding for services in EOSC will come primarily from national or European support, whether through free at the point of use services, or by giving funds or tokens to researchers in order to purchase them. The development for shared models to allow this and to support and persuade providers to widen access to their services is a key question for EOSC as it develops.

2.3.1.3 EOSC-Core service economic models

The Sustainability Working Group conducted an initial operational cost assessment of the EOSC-Core. Based on the findings of the study on the EOSC-Core operational costs, it is estimated that the cost of operating EOSC-Core is approximately 7 Million €, however a more accurate estimate will be possible once full costs of ownership and the list of capabilities to be provided in the EOSC-Core are defined. The study also concludes that the long-term funding of the EOSC-Core should be addressed by the members of the EOSC Association. Currently Members and Observers of EOSC pay an annual fee, contributing to the determination of the budget of the EOSC Association and support its operational activities. The fee scheme will be revised once the business model has been agreed.

2.3.2 GAIA-X

2.3.2.1 GAIA-X Service Delivery Models

As currently specified (in its Architecture Document [12] and the specification of its Federation Services [unpublished]), GAIA-X is developing a *Structured Marketplace* for GAIA-X compliant data and infrastructure services, along with a broader policy and standards framework intended to support its

mission of creating a "federated open data infrastructure based on European values regarding data and cloud sovereignty." It is hoped that digital ecosystems will adopt the GAIA-X approach in order to facilitate easier and more controlled sharing of data within each ecosystem, as well as facilitating expanded collaboration between, and potentially merger of, those ecosystems.

GAIA-X is aware of the value of supporting a *Reseller* business model, identifying service metering, billing and contract management as functions that might be provided by GAIA-X in the future. GAIA-X also aspires to offer some *Assembler* tools to help customers integrate the GAIA-X services they select with orchestration tools, which are covered in a draft specification for an Orchestration Federation Service.

2.3.2.2 GAIA-X Service Economic Models

Individual Service Providers within GAIA-X must set their own prices and contracting arrangements with GAIA-X Consumers. GAIA-X Federated Catalogues will display this information, but no "ordering" functionality is currently planned.

2.3.2.3 GAIA-X Federation Service Economic Models

There has been no public discussion within GAIA-X of how Federation Services and other common functions will be paid for. As noted in the Organization section, the GAIA-X AISBL is hesitant to be responsible for the operation of these services, but it is also not clear how Federators (GAIA-X Participants defined in the Architecture as operating Federation Services) would be compensated for this activity.

GAIA-X Members pay an annual fee ranging from 5,000 to 75,000 € for for-profit organizations (based on their consolidated revenues) and 2,500 € for not-for-profit organizations. With a membership base of over 200 members, and an assumed average membership fee of 10,000 Euros, GAIA-X' total member revenues are on the order of 2 Million €. It is unclear if this fee structure can sustainably support the AISBL's operations plus any services to be operated by (or on behalf of) the AISBL.

2.4 Policy and compliance

Both initiatives recognize the need to establish common "rules of engagement" for participants in their respective initiatives. These rules need to align with technical solutions but primarily focus on semantic, organizational and legal topics. The two sets of rules both emphasize transparency, but otherwise address largely complementary issues.

2.4.1 EOSC

Current EOSC Rules of Participation [21] are very high level, and act more as principles than rules or criteria. At present they are organized into eight areas:

1. EOSC is based on the principle of openness.
    a. Users: Use of EOSC is open to anyone, regardless of role or geography. However, the EOSC is envisaged primarily for use by researchers in Europe and beyond.
    b. Resources: EOSC resources are listed in a publicly accessible registry, comply with policy requirements (for example on openness of publicly funded data) and are findable without charge by all users.

i. Access to certain resources may require personal or organisational registration, authentication or authorisation including for the purpose of compensating the resource provider.

ii. Terms of use must be made available by resource providers, including information about whether access requires authentication and authorisation; licencing; and any quotas or charges which may apply. Open access is the default and any departure from this must be justified. Resources that are open will be tagged as such so that they can be easily found among the EOSC resources.

c. Resource providers will be considered for onboarding to the EOSC according to transparent, published criteria. Adherence to the Rules of Participation forms a central element of the onboarding process of resources to the EOSC.

2. EOSC resources align with FAIR (Findable Accessible Interoperable Reusable) principles. Data in EOSC should be made available in accordance with FAIR principles, the requirements of which extend to all EOSC resources as data may require software and other resources to yield reproducible research.

3. EOSC services align with EOSC architecture & interoperability guidelines.

   a. EOSC aims to build a coherent infrastructure that removes silos and provides integration of data and services within and across geographical and disciplinary boundaries. Examples of relevant technical standards adopted by EOSC include Six Recommendations for Implementation of FAIR Practice [22], A Persistent Identifier (PID) policy for the European Open Science Cloud [23], the EOSC Interoperability Framework [16] and the AARC blueprint architecture for Authentication and Authorisation Infrastructure (AAI) [24].

   b. EOSC architecture and interoperability guidelines shall enable user environments to be built across (multiple) EOSC resources for enriching the user experience.

      i. Services requiring authentication or authorisation should support the use of relevant credentials for federated AAI as defined by the EOSC AAI taskforce, and, where access requires authentication and/or authorisation, data providers should make clear what body will make these decisions (e.g., a Data Access Committee) along with appropriate details.

      ii. Services shall be described by a commonly agreed metadata scheme and service providers shall define and publish the terms of use for their services, including, for example, licensing, authentication and authorisation requirements, and any cost implications. These terms of use must comply with the EOSC principles and any relevant legal and ethical conditions on how data can be accessed, processed, analysed, changed and redistributed by others.

4. EOSC is based on principles of open science, ethical behaviour and research integrity.

5. EOSC users are expected to contribute to EOSC.

6. EOSC users adhere to terms and conditions associated with the resources they use.

   a. Ownership of resources is not changed when they are made available through EOSC. As such, EOSC users agree to adhere to the terms of use for the specific resources they use. Service providers can define and publish the terms of use for the service they are provisioning, but these terms themselves should be consistent with the EOSC Rules of Participation. This includes licensing and conditions of use, and whether access requires authentication and/or authorisation.

   b. Access to some resources may explicitly require users to accept Terms & Conditions before access is granted. In order to establish a trust framework on established standards and compliance with EU regulations (e.g. certification that a service conforms with GDPR or other restrictions), resource providers have to indicate, through resource metadata, the compliance and conformity with the standards and regulations. The resources are not accessible through EOSC until these conditions are met.

7. EOSC users reference the resources they use in their work.
8. Participation in EOSC is subject to applicable policies and legislation.

More detail is expected from the next phase of RoP, currently under development.

2.4.2 GAIA-X

The GAIA-X Policy and Rules Document [17] "*defines High Level Objectives safeguarding the added value and principles of the GAIA-X ecosystem. GAIA-X's Policy Rules intend is to identify clear controls to demonstrate European values of GAIA-X, such values including Openness, Transparency, Data Protection, Security and Portability. Each and every service offering to be provided under the umbrella / via the GAIA-X framework shall comply with all of the following objectives. In general, full adherence to applicable EU legislation (e.g. in areas such as data protection and Security) is a prerequisite and thus not waived or affected by the following policies and rules. ... participation within GAIA-X and providing GAIA-X compliant services, shall not prevent any provider to also provide non-GAIA-X service offerings outside the GAIA-X ecosystem.*"

[18] focuses on two main areas:

1. For Cloud Service Providers (not defined, but presumed to be a subset of Providers in the GAIA-X domain):
   a. Compliance with GDPR
   b. Transparency of Contracts, Assets, Compliance, Sub-contractors
   c. Compliant practices regarding Cybersecurity (referring also to the ENISA Cloud security certification framework, once it is available)
   d. Portability of services and data in accordance with the Free Flow of Data Regulation
   e. Required attributes of Contracts
2. Data Sharing within Data Spaces
   a. Data providers will define machine-readable usage policies
   b. Data consumers will respect the defined usage policies, taking appropriate technical/organizational means.

3 Architecture and Technology

This section explores the architectural approaches of GAIA-X and EOSC and the functional areas addressed. A general overview is provided, followed by a decomposition of architectural elements (such as roles, entity model, relationships, functions) and functional technology comparison.

3.1 General overview

Various initiatives are tackling the issue of data and service integration in an ecosystem. Besides EOSC and GAIA-X, which are illustrated in the following sections, Destination Earth is addressing the problem of data system interoperability, advocating to minimize data movement by relying on a federated cloud and data architecture where data can be discovered remotely and processed locally at the system providing data access [25].

3.1.1 EOSC

EOSC Future project further detailed the EOSC architecture by modelling it around: (1) roles (users and providers), (2) access channels (the EOSC Portal), (3) EOSC Core and EOSC Exchange services and (4) Interoperability Framework as illustrated in Figure 1.

- **EOSC-Core** is defined by the internal services which allow EOSC to operate as a federation. It includes a Core technical platform which facilitates EOSC delivery upon which the researcher-facing resources in the EOSC-Exchange can rely and integrate with as appropriate. It also includes non-technical coordination including onboarding and security coordination.
- **EOSC-Exchange** provides services and other resources registered into the EOSC to serve the needs of research communities. Generic services and resources which target multiple scientific domains and research communities are identified as *Horizontal Services*. Resources which target users from a specific scientific domain, community and/or regional domain are identified as *Thematic* and/or *Regional* Resources. The capability to compose resources across horizontal and thematic and/or regional resources relies on the EOSC Interoperability Framework [16].
- **EOSC Interoperability Framework** (EIF) [16] is a framework of standards and guidelines to support the interoperability and composability of resources in the EOSC-Core and EOSC-Exchange. It allows EOSC to integrate services and research products (e.g. publications, datasets, software) across resources and providers. Providers have the freedom to develop and operate provider specific implementations while conforming to the EIF guidelines and standards. Data ecosystems delivering thematic capabilities are independently operated outside EOSC for their reference targets groups.
- **EOSC Support** activities sit alongside the EOSC-Core and EOSC-Exchange, and comprise the training, engagement, and other human-centric activities which make EOSC more attractive and easier to use, and help users benefit from it more easily once engaged. They include Training, support and the EOSC Digital Innovation Hub for engagement with the commercial sector.

The diagram in Figure 1 illustrates the nested nature of the architecture. In particular, the more detailed diagram depicts the EOSC Portal, as well as “Thematic portals” serving each community/cluster operating within the EOSC framework. Thematic portals can be enabled by local

resources that are not integrated with EOSC, as well as resources published in EOSC. Access policies and service provisioning models in each thematic ecosystem are specific to the community. This aspect of the architecture is specifically required to accommodate the differences between domains, while at the same time encouraging adoption of the common capabilities of EOSC itself. An implementation example of this high-level architecture is illustrated at the bottom of Figure 1 representing how multiple thematic services are supported by multiple clouds adhering to the EOSC interoperability.

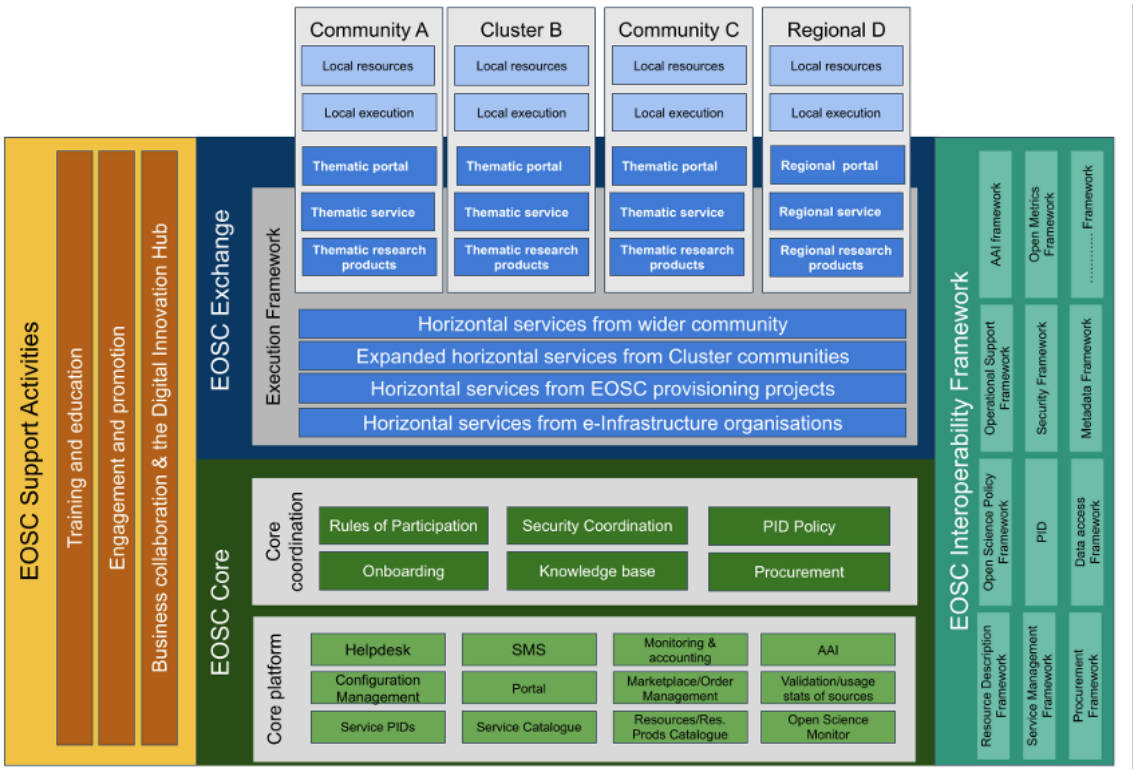

Figure 1. High-level architecture for EOSC (credits: EOSC Future Consortium [not yet published]).

3.1.1.1 EGI-ACE

Multiple EOSC implementation projects are adopting the EOSC architecture WG high-level architecture to realize thematic and regional ecosystems providing specialized services and resources shared in EOSC. EGI-ACE[11] (EGI Advanced Computing for EOSC) delivers the EOSC compute platform as a federated cloud-based infrastructure connecting multiple private and public cloud providers, and it relies on dedicated capacity provided by federation members of the EGI cloud and HTC infrastructure.

EGI-ACE deploys a generic set of horizontal services to support multiple data ecosystems for fundamental science, health (including drug discovery, containment of infectious diseases), and the Green Deal tackling various societal challenges (e.g., biodiversity conservation and water quality).

[11] https://www.egi.eu/projects/egi-ace/

The project realizes this by federating data, facilities and thematic applications involving tens of cloud hosting nodes from the EGI Federation[12], international communities of practice and data providers such as LSGC, WeNMR, IS-ENES, SeaDataNet, Galaxy.eu, DMCC; and research infrastructures including LOFAR, MeerKAT, EMSO, GBIF, ITER, EISCAT, VIRGO, e-RIHS, and PHIRI.

The resulting EOSC data ecosystems include:

1. Components that enable federation of distributed cloud and high-throughput computing systems and services,
2. Compute and storage capacity,
3. Federated data and compute management enabling data transfer, orchestration and workload management,
4. Federated trust and identity management, (5) access to artificial intelligence (AI) and machine learning (ML) on demand services and (6) platform-as-a-service (PaaS) capabilities for web-based data analytics.

The EGI-ACE architecture conforms to the principles of the EOSC high-level architecture. As data and compute federation, it integrates infrastructure, distributed data and applications, organizing capabilities in a tiered fashion. As illustrated in Figure 2, the *Federated Resource* tier delivers hybrid IaaS and HTC compute facilities and storage for hosting research data and tools.

The *Federated Access* tier complement this and provides distributed access services:

- Federated Identity Management to enable secure, trustworthy, and seamless access to all services
- Federated Data Access services to support exposing discoverable datasets and staging data into/out of the EOSC Cloud. These services manage the raw storage capacity delivered by the resource tier to deliver and transfer data between the EGI-ACE providers and external data repositories.
- Federated Compute Access services to orchestrate the execution of user workloads on the resources. These services exploit data locality by executing computing tasks next to the data and facilitate application portability across a diverse range of computing platforms (Cloud IaaS, HTC, HPC) by managing various software distribution mechanisms (VM images, container images, binaries).

On top of the Federated Access, the *Platforms* tier delivers higher level abstractions matching the needs of the data ecosystem operators, simplifying the execution of their data-driven workloads in the federation. The Platform Tier integrates the compute and storage resources so that they can be easily reused by different communities to build final end-user services.

The Data Ecosystems and Analytics tier includes data, discipline specific data analytics, as well as general-purpose simulation, machine learning and data analytics capabilities tailored to the needs of a specific research domain. Services from all four tiers are exposed to EOSC users via the EOSC Portal. Resource providers at all levels can access and use Federated Access services to deliver their own services to EOSC users.

[12] https://www.egi.eu/federation/egi-federated-cloud/

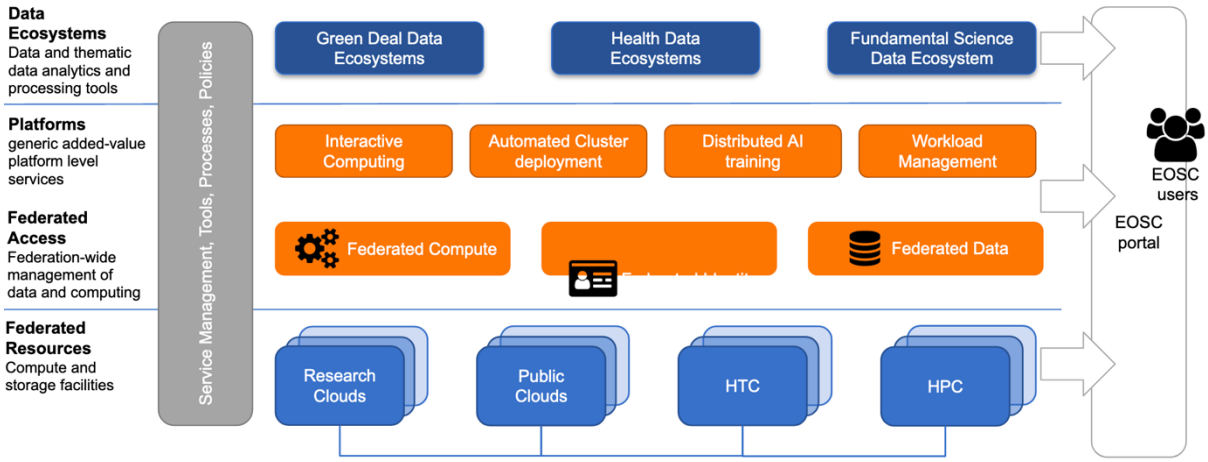

Figure 2. The EGI-ACE data ecosystem and service architecture is organized in three tiers, from bottom to top: Federated Resources, Federated Access services and Platforms, and Data Ecosystems.

In agreement with the Sustainability recommendations [9], EGI-ACE supports three different economic models depending on the type of service and the amount of capacity requested:

- *Sponsored access to the Compute Platform capacity (EC funding)*. Access to the EOSC Compute platform is centrally managed. EC funded capacity is used to promote the hosting of research data and its exploitation through thematic tools that are provided to the user as a fully secure, trustworthy and integrated data ecosystem. In this scenario, e-Infrastructures (EGI) and partner research communities play the role of integrator entering into a single Service Level Agreement (SLA) with the customer while interacting with service providers of the federation through separate Operating Level Agreements (OLAs). This will cover separate, unintegrated services, where the customer wants to rely on a single organization to manage multiple information services.
- *Sponsored access to additional capacity (MS funding)***.** This model is suitable to acquire compute and storage capacity to expand the capacity delivered thanks to EC adoption funds. In this scenario the customer is responsible for the implementation of custom data ecosystems, playing the role of system integrator to address the needs of a specific research community.
- *Pay-for-use access* to procure the necessary services with commercial providers. Like in the previous model, the customer plays the role of integrator.

3.1.2 GAIA-X

The GAIA-X architecture describes the concepts required to set up the GAIA-X data and infrastructure ecosystem and realize the vision established for GAIA-X. It integrates the Providers, Consumers, and Services needed for this interaction. These Services comprise ensuring identities, implementing trust mechanisms, and providing usage control over data exchange and Compliance – without the need for individual agreements.

Details about implementing the GAIA-X ecosystem are still to be defined by the *Architecture of Standards Open Work Package* group.

Automated contracts, legal binding, monitoring, metering as well as billing mechanisms, amongst others, are not currently defined in the GAIA-X architecture.

The architecture centres on the GAIA-X Conceptual Model (see Figure 3):

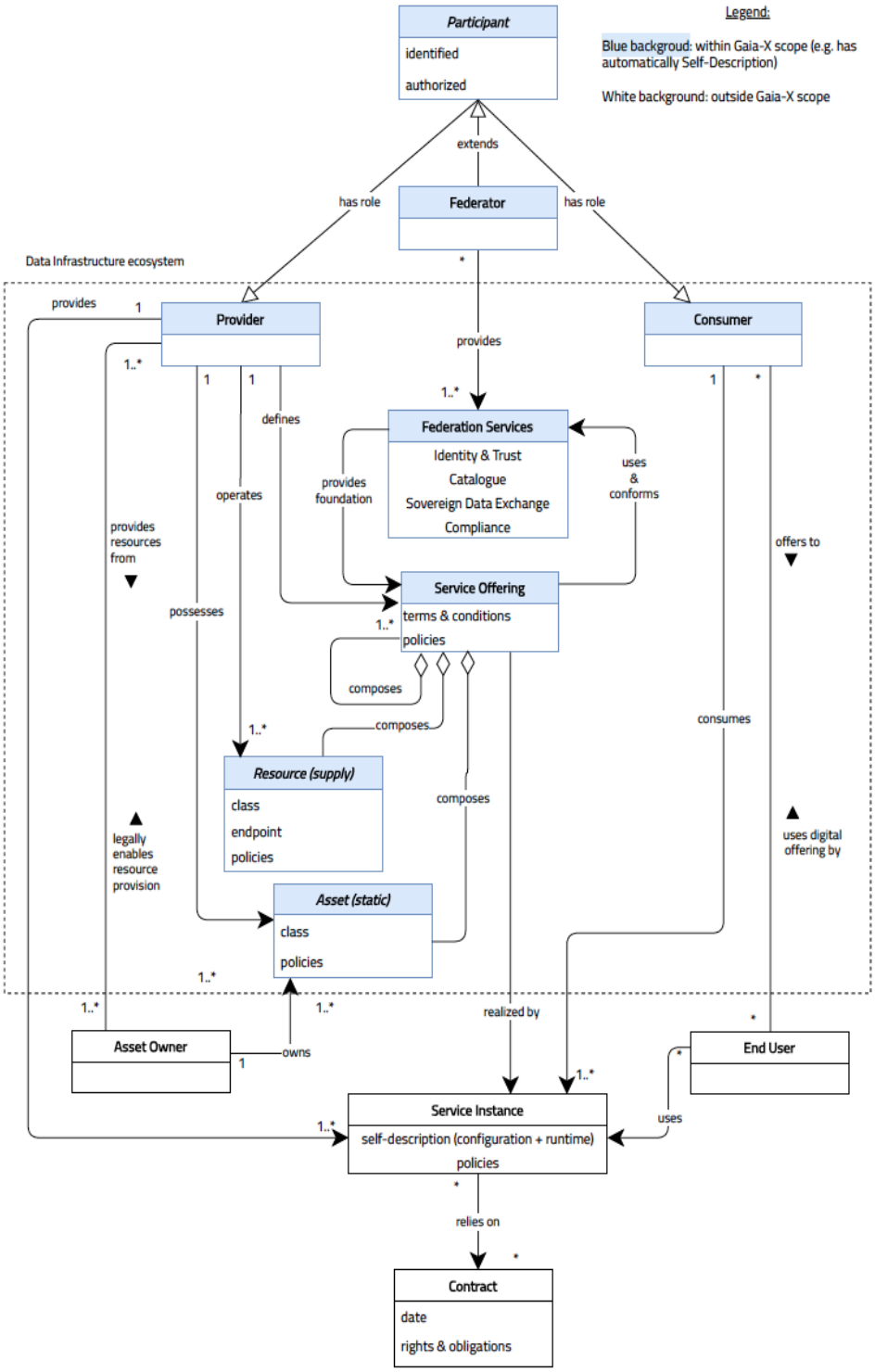

Figure 3. GAIA-X conceptual model

3.2 Integrating Data and Infrastructure within Data Ecosystems, Using FAIR to Enable Data Sharing

Both EOSC and GAIA-X seek to integrate both data and infrastructure, enabling safe and secure access to relevant data, and access to a range of relevant data processing resources and systems, enabling the creation of new insights and knowledge and support for data-driven activities and business.

Against this vision, both initiatives must contend with practical limits to this integration:

- The responsible "owners" of both data and infrastructure may need to limit both the visibility and the use of their data and infrastructure to specific users or defined groups of users, for a variety of legal, regulatory and ethical reasons.
- Owners may have their own reasons (e.g. confidentiality, commercial advantage) to want to establish similar limits on visibility and use. This can apply even to research data, which may need to be kept confidential pending publication or pending patent approval.
- Even when owners are interested in making their data and infrastructure widely visible and usable, effort is usually needed to prepare that data and infrastructure for listing and use.

Regardless of the reasons, this variation in visibility translates into the need to create separate environments for data and infrastructure that allow visibility and use to be limited within a given environment and that allow entity owners to make sovereign decisions about to which environments they wish to expose their data and infrastructure. This is a strong requirement in both the corporate and the research domains. For this analysis we equate these separate environments to "data ecosystems" and more generally to the "systems" or "federations" that will be brought together in any "system of systems" and "federation of federations".

Where there is a desire to share data, the FAIR Data Principles [22] have been developed specifically to guide the preparation of data for listing and use – primarily in a public context such as for research – but the principles apply to any exchange of data. Steps include:

- Storing the data in a place (such as a repository) and assigning an identifier to it so that it can be accessed,
- Annotating data with metadata so that it can be discoverable using search technology,
- Coding the data to ensure its consistent interpretation and understanding, and
- Defining the conditions of use in a consistent way so that data users are fully aware of their rights and obligations with respect to the use and protection of the data.

FAIR principles are highly applicable to data, with the limitations noted above, but are of limited applicability to services, which require other properties in order to effectively support open research. These include professional delivery of services to benefit from data, effective management and clear service levels, but these are still under definition.

3.2.1 EOSC

The EOSC vision is to allow for a seamless integration of resources including data and services, all of which can be discovered through Portal services.

The extended research community consultation process that produced the EOSC Architecture (see Section 3.1.1) determined that separate and nested environments were clearly required, from limited "local", to community-based "thematic", all the way to fully public environments. Such a nested and/or segregated architecture allows the architecture to map to existing communities, which assists with adoption and implementation, but more strategically, gives coexisting communities and research projects the ability to make decisions on various topics, from visibility and use, to membership, purpose and common policies, as well as the opportunity to leverage common services. We can identify each of these separate environments as a data ecosystem that will be supported through EOSC.

As depicted in Figure 1, each of these separate environments – or separate data ecosystems – contains its own portal/access channel, allowing the relevant data and services to be listed and potentially selected for use by members of that community. These community portals would include data and infrastructure specific to that portal, as well as entities from larger ecosystems of which they are a part, as well as finally, public entities available to all EOSC participants.

This segregated/nested approach is mirrored by EOSC's approach to identity and trust, which embraces a federated structure that recognizes the level of trust that has already been established within each community AAI scheme.

EOSC recognizes the value of applying FAIR principles and working to improve FAIR "readiness" or "maturity", but work is needed:

- There is a general agreement that persistent globally unique identifiers (PIDs) are needed, not only for data but also for a range of other research artefacts such as services, software, workflows, as well as interoperability artefacts, such as metadata schemas and ontologies. At the same time, today there are many types of PIDs and PID-generating services, with varying levels of accessibility and sustainability. EOSC will define PID policies which govern or advise on which PID systems are EOSC compliant.
- Providing all artefacts, or even just all research data, with consistent, complete and "useful" metadata is itself a challenge: numerous well-established metadata standards have been developed that are only partially interoperable themselves, and within existing data repositories the extent of annotation for each repository's holdings sometimes falls far below what might be expected, much less what is needed in order for metadata-based searches to return relevant results from the repository. Based on its review of multiple metadata standards, EOSC has proposed a minimum metadata format as a possible "minimum viable metadata standard" that might be adopted [22].
- Even in a research environment, where FAIR is understood to mean "as open as possible, as closed as necessary", the various limitations on visibility and use mentioned above translate into the need for the clear definition of acceptable Use Conditions. For example, personal health information can be used in health research, provided it has been safely anonymized or specific, informed consent has been obtained to enable the proposed use. These requirements need to be managed, ideally through standardized approaches that allow data use to be automated as well as tracked and monitored. There are various promising efforts to model these types of Use Conditions, for example, the Data Use Ontology [26] of the Global Alliance for Genomics and Health, but more work is needed to capture the needs of the community, even when "open data" is the assumed norm.

EOSC defines the entities that are available for use using EOSC Profiles [27]. These are '*specifications that define common data models for EOSC entities (Providers, Resources, etc) and related taxonomies. They contribute to the unified framework for describing and offering EOSC Resources to end-users in a harmonised way, guaranteeing the interoperability of resources metadata with open APIs'*. Providers register their own EOSC Profiles as well as the EOSC Profiles of any resources they want to list through the EOSC Portal. The EOSC Rules of Participation [21] define a *Resource* as a digital object or process such as data and metadata, publications, software, workflows, services, and training materials. "Resource" can refer to both services and data, as well as other artefacts needed by the research community.

In addition to information needed to populate the EOSC Service Portfolio entry corresponding to the new service, the EOSC Profiles provide a detailed description of the service, documenting licenses, helpdesk contact, accessibility and other aspects needed to support not only "display" of the service in the Portal, but also operational use of the service, from selection and interoperability to support for service use.

3.2.2 GAIA-X

GAIA-X also envisions a seamless integration of data assets/services and infrastructure assets/services.

Resources and Assets describe the goods and objects of a GAIA-X Ecosystem and are defined as follows:

- **Assets**. An Asset can be a Data Asset, a Software Asset, a Node or an Interconnection Asset. The different categories of Assets are defined below:
    - A Data Asset is an Asset that consists of data in any form and necessary information for data sharing.
    - A Node is an Asset and represents a computational or physical entity that hosts, manipulates, or interacts with other computational or physical entities.
    - A Software Asset is a form of Asset that consists of non-physical functions.
    - An Interconnection is an Asset that represents the connection between two or more Nodes.
- **Resources**. Resources represent those elements necessary to supply Assets. They can be explained as internal Service Instances not available for order. For example, the running service instance that provides a dataset (a Data Asset) is a Resource.

Resources and Assets compose Service Offerings, the primary entities that will be listed in Federated Catalogue services. Self-Descriptions describe certifications and other non-functional attributes appropriate to each type of entity, e.g. identifying the entity's owner and location in order to assess the status of the entity relative to EU and Member State legislation, as well as a customer's own requirements. In particular GAIA-X intends to track Resources and Assets at a granular level that will allow evaluation of ownership, location and other attributes needed to ensure legal and regulatory compliance.

Providers are responsible for the creation of their Asset's or Resource's Self-Description. Self-Descriptions must be submitted to GAIA-X for validation before they can be included in Federated Catalogues

All entities in GAIA-X are described with a formally defined "Self-Description" schema or language that is essentially a metadata specification.

- For data assets, Self-Descriptions incorporate the Dublin Core metadata standard[13], as well as incorporating the Open Digital Rights Language[14] as a syntax that can be used to express usage control conditions. Several usage control concepts have been defined so far.
- Services and infrastructure assets are similarly described using Self-Descriptions, but details of the ontology are under development by the Infrastructure Open Work Package group.

All GAIA-X entity Self-Descriptions are tied to a unique Identifier. It is unclear whether Identifiers are assigned by Providers, GAIA-X or a designated external service, whether they are expected to be either persistent, or unique not only within but outside the GAIA-X landscape.

GAIA-X's vision of a trusted ecosystem of resources and assets, with attributes, certifications and qualities on which Consumers can rely, is built on a system of independently issued credentials that can be verified without having to depend on trust either in the Provider or in GAIA-X. In addition to self-declared Claims made by Participants about themselves or about the Service Offering provided by them, a Self-Description can include Credentials issued and signed by trusted parties. Such Credentials include Claims about the Provider or Asset / Resource, which have been asserted by the issuer. The Verifiable Credentials Data Model[15] represents a key foundational element of the GAIA-X approach and architecture.

Self-Descriptions are also expected to provide structured interoperability attributes, allowing customers to select entities (services and data) that should be interoperable with one another; these interoperability attributes could range from technical, syntactic and semantic interoperability, all the way to organizational and legal interoperability. Relevant interoperability standards are being catalogued by an "Architecture of Standards" group within the GAIA-X community, although this group has identified a tremendous range and diversity of interoperability standards that already exist.

GAIA-X is considering the needs of its stakeholder communities (particularly various data ecosystems or digital ecosystems) to shield the visibility of certain entities (data, infrastructure, services) so that they would only be discoverable and accessible within those communities and through their respective Federated Catalogues. GAIA-X also contemplates the possibility of augmenting certain Self-Descriptions for relevant entities (e.g. data assets and related services) using domain-specific metadata ontologies selected by different stakeholder communities, but this has not been specified in detail.

The GAIA-X Architecture does not refer to the FAIR Data Principles, but its architectural approach aligns with those principles:

- Identifiers: GAIA-X indicates the need for all entities (Providers, Resources, Assets) to be uniquely identified, although the mechanisms for this are not discussed.
- Metadata: The GAIA-X Self-Description specification is a form of metadata. The focus of the specification is on defining entities and non-functional attributes of special relevance to support cloud and data sovereignty. For data the Dublin Metadata Core Standard is

[13] https://dublincore.org/
[14] https://www.w3.org/ns/odrl/2/ODRL20.html
[15] https://www.w3.org/TR/vc-data-model/#ecosystem-overview

incorporated by reference into the GAIA-X Self-Description schema, although this is not the only metadata standard that might be used.

- Interoperability: GAIA-X has a clear intent to support and encourage interoperability among and between the entities listed in its Federated Catalogue. The Architecture of Standards Open Work Package aspires to create a harmonized approach to the myriad of interoperability standards available across the many sectors targeted by GAIA-X.
- Use Conditions: GAIA-X has explored the use of the Open Digital Rights Language as a grammar to create robust machine-readable Use Conditions. The Data Sovereignty Open Work Package has defined some of the required concepts within the Self-Description language but more work, as well as validation in the real world, will be needed.

3.3 Roles

3.3.1 EOSC

The EOSC Rules of Participation define roles comparable to, but more general than, those defined in GAIA-X:

- **Participant**: person or legal entity interacting with EOSC resources.
- **User**: participant consuming EOSC resources.
- **Provider**: participant offering EOSC resources for consumption.

3.3.2 GAIA-X

GAIA-X defines a number of key entities that play key roles in the architecture:

- A **Participant** is an entity, as defined in ISO / IEC 24760-1 as an "item relevant for the purpose of operation of a domain that has a recognizably distinct existence", [28] which is onboarded and has a GAIA-X Self-Description. A Participant can take on one or multiple of the following roles: Provider, Consumer, Federator. Provider and Consumer present the core roles that are in a business-to-business relationship while the Federator enables their interaction.
- A **Provider** is a Participant who provides Assets and Resources in the GAIA-X Ecosystem. It defines the Service Offering including terms and conditions as well as technical Policies. Further, it provides the Service Instance that includes a Self-Description and technical Policies. Therefore, the Provider operates different Resources and possesses different Assets.
- **Federators** are in charge of the Federation Services and the Federation which are autonomous of each other. Federators are GAIA-X Participants. There can be one or more Federators per type of Federation Service. Note that in EOSC there is no definition of a role comparable to the "Federator" role in GAIA-X.
- A **Consumer** is a Participant who searches Service Offerings and consumes Service Instances in the GAIA-X Ecosystem to enable digital offerings for End-Users.

All of the above entities are described in the GAIA-X ecosystem with Self-Descriptions (see section 3.2 above). The GAIA-X Conceptual Model defines the role of Federator, and its architecture defines "Data Ecosystem", "Ecosystem" and "Federation" in similar ways. Nevertheless, the role of Ecosystems, as distinct from the "GAIA-X Ecosystem" overall, is not described.

"Asset Owners" and "End Users" are defined in the GAIA-X Conceptual Model but are not associated with Self-Descriptions. Asset Owners, e.g. data owners, describe a natural or legal person, which holds the rights of an Asset that will be provided according to GAIA-X regulations by a Provider and legally enable its provision. End-Users use digital offerings of a GAIA-X Consumer that are enabled by GAIA-X. The End-User uses the Service Instances containing Self-Description and Policies.

Various trust-providing organizations, as well as "Conformity Assessment Bodies," are referred to in the GAIA-X Architecture Document, but it is unclear whether they need to be defined in the Conceptual Model or if their services need to be better defined.

A Federation is defined as a loose set of interacting actors that directly or indirectly consume, produce, or provide Assets and related Resources; however, federations are not defined as entities within the GAIA-X Conceptual Model.

Note that neither Providers or Consumers are distinguished by whether they provide/consumer services or data -- services and data are abstracted as instances of Resources and Assets (discussed above). Within the Conceptual Model, a given Participant can operate as a Provider, a Consumer, and potentially a Federator.

3.4 Services for Federation Participants

Both initiatives have identified a number of core or federation services that enable discovery, access, interoperation, trust and identity, consultancy and support.

3.4.1 Authorization and Access (AAI, Identity & Trust)

3.4.1.1 EOSC

The EOSC Federated AAI is specified in the EOSC Authentication and Authorization Infrastructure [29] and enables Service Providers to deliver services and access to resources to research communities and individual researchers, allowing users to use their institutional, community and eIDAS enabled digital identities. It enables service providers to control access to their services from users holding identities (usernames and passwords) from a very broad set of academic, community or social Identity Providers (IdPs). It federates IdPs, the service providers (SPs) and intermediary proxies into a single, interoperable infrastructure. It supports SAML 2.0, OpenID Connect, OAuth 2.0 and X.509v3 to offer a flexible framework for access management. The EOSC Federated AAI comprises different, compatible proxy solutions.

The EOSC Federated AAI recognizes the role of the different "Communities" that EOSC serves and deliberately seeks to "federate" AAI services across the wide range of existing, productive science communities -- each with their own AAI services and policies. It also follows architectural recommendations found in the AARC blueprint architecture for Authentication and Authorisation Infrastructure (AAI)[16] which reflects the collaborative agreement of these science communities about how to evolve their respective AAI services to balance the need to broaden and simplify access with the need to maintain the security and integrity of the resources being accessed.

[16] https://aarc-project.eu/architecture/

EOSC explicitly adopts several collaboratively developed policies in this area, including REFEDS Research & Scholarship Entity Category (REFEDS R&S)[17], Security Incident Response Trust Framework for Federated Identity (REFEDS-Sirtfi)[18], and Scalable Negotiator for a Community Trust Framework in Federated Infrastructures (IGTF-SNCTFI)[19] [29]. REFEDS in particular defines the minimum attributes required in the research and education community, in particular the profile specifications for single factor authentication and multi-factor authentication

The EOSC Federated AAI also highlights the important role of identity attributes, which are provided in a structured way by the various user communities and used by service providers to make initial decisions about users' rights to access different resources ("authorization").

3.4.1.2 GAIA-X

For GAIA-X, Identity Management will be handled primarily by a self-sovereign identity (SSI) scheme, with distributed identifiers (DIDs). Since this is new technology, it is unclear how ready stakeholders are to adopt SSI/DIDs, and the IAM Federation Service specification [not yet published] describes the possibility for Consumers, e.g., industrial companies, to be authenticated using OpenID Connect (OIDC), although this is not an option for Providers. Since many Participants will need to adopt both Consumer and Provider roles, the requirement to adopt SSI/DID may prove challenging.

For GAIA-X, note that attribute management is not defined, but will be embedded in Self-Descriptions through self-asserted claims, Verifiable Credentials and other metadata attributes. No explicit mechanisms for access control are presented, although the Extensible Access Control Modelling Language (XACML) is referenced in the specification for IAM Federation Services [not yet published].

3.4.2 Compliance

GAIA-X defines special federation services intended to maintain the compliance of Gaia-X entities with requirements defined in the agreed Policy Rules Document, by verifying any claims and credentials presented in the GAIA-X entity self-descriptions. Conceptually these services ensure compliance at initial submission of an entity's self-description, and throughout the life of that entity within GAIA-X. Compliance would be assured through the use of Verifiable Credentials as well as both automatic testing and external auditing performed by Conformity Assessment Bodies.

EOSC approaches to compliance with Rules or Participation are currently under discussion and expected to be clarified later in 2021.

Data protection and GDPR compliance are a common concern of EOSC and GAIA-X that represents an area of potential collaboration. GÉANT have been endeavouring to progress a voluntary Code of Conduct for some time for federated access activities of digital infrastructures for research. On the other hand, CISPE delivered the first pan-European sector-specific Code of Conduct for cloud infrastructure service providers under Article 40 of the European Union's General Data Protection Regulation (GDPR) receiving a green light from the European Data Protection Board (EDPB). The Code helps organisations across Europe accelerate the development of GDPR compliant cloud-based services for consumers, businesses, and institutions[20].

[17] https://refeds.org/category/research-and-scholarship
[18] https://refeds.org/sirtfi
[19] https://www.igtf.net/snctfi/
[20] https://www.codeofconduct.cloud/

3.4.3 Discovery (catalogues and portals)

3.4.3.1 EOSC

The EOSC Portal provides a European-level delivery channel connecting the demand-side (the EOSC customers) and the supply-side (the EOSC providers) to allow researchers to conduct their work in a collaborative, open and cost-efficient way for the benefit of society and the public at large. It enables different kinds of users, with different skills and interests, to discover, access, use and reuse a broad spectrum of EOSC resources (services, datasets, software, support, training, consultancy, etc.) for advanced data-driven research. It supports interdisciplinary research and facilitates resource discovery and access at the institutional and inter-institutional level. It provides access to integrated and composable products and services from the EOSC Catalogue and facilitates the composition of services and products to support multi-disciplinary science, for example with community-specific tools. From the provider's point of view, it helps gaining access to new target groups and offering services under homogeneous terms of use, acceptable use policies, and in different configuration options, so that users are guided in the choice.

3.4.3.2 GAIA-X

Federated Catalogues constitute the repositories for GAIA-X Self-Descriptions to enable the discovery and selection of Providers and their Service Offerings. The Self-Description, as the expression of properties and Claims of Participants and Assets, represents a key element for transparency and trust in GAIA-X.

Although it is possible that there will be a "master" Federated Catalogue for GAIA-X, encompassing all Providers and Service Offerings, it is expected that there will be many community and ecosystem-specific Federated Catalogues, as well as potentially various "private" catalogues listing service offerings intended for use by a single organization.

Gaia-X has specified several characteristics for any given catalogue:

- Depending on the catalogue, "visitors" (unauthenticated users) and authenticated users (representatives of the Participants, known as Principals) may have access to different information from that catalogue. In particular, only Participants and their Principals will be allowed to confirm the Verifiable Credentials that might be available for a given Provider or Service Offering.
- In the context of a given user's search and filtering criteria, search results (typically Providers and Service Offerings) should be presented without any favouritism, ordering results randomly.
- Since entities defined in Gaia-X possess a range of linked attributes, e.g. standards for technical interoperability, which define a "graph" of relationships among entities, users should be given the ability to use these relationships to identify relevant service offerings. For example, the user might describe a sample workflow, with Data A, processed by Service B, producing Data C, and then consumed again by Service D, producing Data E. A "graph" style search would find the combinations of data and services that would fit the requirements of this workflow.

- Each Catalogue should support both a human-machine interface, as well as a "machine readable" API, allowing programmatic searches for qualifying services, allowing automated service selection, and eventually ordering, procurement, assembly and operation.

3.4.4 Selection, Ordering/Procurement/Usage Control, Monitoring, and Accounting/Metering

Although both initiatives envision seamless integration of data and services, we explore this functional area separately for services and for data.

3.4.4.1 Services

3.4.4.1.1 EOSC

Following listing in the EOSC catalogue, providers are offered the opportunity to set up ordering. Functional steps parallel the range of service delivery models that are possible.

- **Service selection.** The EOSC Portal offers a web interface to select resources relevant to the user. Access to each individual service is obtained through APIs and web interfaces depending on the service.
- **Ordering/Procurement/Usage Control**. Where applicable, services can be ordered through the EOSC Portal after authentication of the user. Providers can access a 'Service Order Management Back Office' (SOMBO) that delivers a dashboard to manage each service order. It eases the communication between all parties, facilitates the negotiation between service requesters and service providers, provides facilities to sign SLA/OLA and automates various operations required by the order management process. Authorization requires confirmation that the prospective user has the needed attributes to use the ordered service, as well as determination whether the use is covered by an appropriate "economic model" to compensate the provider for the user's use of the service.
- **Composition/Orchestration/Integration.** Services expose interfaces for programmatic composition of data and services, and systems for workload management systems are provided. Integration in a multi-cloud environment will be possible using documented APIs and orchestration of VM and container execution to deliver distributed applications, running models where data is. In future it is expected that preset 'packages' or integrated and composed services in specific areas will be offered as a way to offer composed resources to users.
- **Monitoring**. Monitoring is the key service needed to gain insights into a group of integrated services. It needs to be continuous and on-demand to quickly detect, correlate, and analyse data for a fast reaction to anomalous behaviour. The challenge of this type of monitoring is how to quickly identify and correlate problems before they affect end-users and ultimately the productivity of their organizations. The features of a monitoring system are monitoring of services, reporting availability and reliability, visualization of the services status, providing dashboard interfaces and sending real-time alerts. Management teams, administrators, service owners within an integrated data ecosystem can monitor the availability and reliability of the services from a high-level view down to individual system metrics and monitor conformance against multiple SLAs.
- **Accounting/Metering.** In the context of an integrated data ecosystem, accounting collects, stores, aggregates, and displays usage information. Usage data is collected from the service

endpoints into a centrally managed Accounting Service. Accounting information is gathered from the service by probes and sensors according to certain data formats. Probes and sensors are deployed locally at the service providers. Data is forwarded from the sensors into a central Accounting Repository where those data are processed to generate various summaries and views for display in the Accounting Portal. Depending on the complexity of the provider, accounting data may go via intermediate repositories that collate accounting data for particular regions, or service groups. EOSC service providers can either directly publish accounting information into the EOSC Accounting Repository or can do so via an intermediate repository that serves for example a specific region or group of providers. It is up to the provider (group) to use the central repository directly, or to apply an intermediary accounting infrastructure and connect it to EOSC.

3.4.4.1.2 GAIA-X

Although contemplated, support for contracts, legal binding, monitoring, metering as well as billing mechanisms, amongst others, is not currently defined in the GAIA-X architecture.

The display of pricing and contract options is described in the specification for the GAIA-X Portal (which provides a web interface for Federated Catalogues) but no mechanism is defined to manage or process this information.

Service orchestration has been described as a desirable Federation Service and might benefit from GAIA-X' detailed and granular Self-Descriptions, but this would require contracting and procurement outside of GAIA-X, either person-to-person, or using a technical solution outside of GAIA-X.

3.4.4.2 Data

Formally, access to and use of data objects in both EOSC and GAIA-X follows the same steps and requires many of the same functions described above for services.

3.4.4.2.1 EOSC

For EOSC, the context of data access and use is one of predominantly open data and open science. The ideal of fully open data has been limited by practicalities, such as protecting the privacy of human subjects, respecting indigenous culture and protecting data about endangered species, giving rise to the concept of FAIR data. The philosophy is that data should be "as open as possible, as closed as necessary".

In reality, if a data object is not fully open, it may rest at many points in the 4-dimensional space of "FAIRness". This may reflect necessity (personal health information must not be disclosed except through informed consent) or practicality (resources are not available to annotate a database of 100,000 records, much less confirm who holds the copyright to each entry). For this and related reasons, EOSC has taken an adaptive approach to data access and use, and each research community needs to establish and uphold its own "fair" policies about FAIR data.

Today most research data is only open within a closed scientific community, and access is managed manually by downloading it from a repository according to the data warehouse model, with individual researchers requesting access from the data provider and having their access rights updated. With the increasing size of data, security policies limiting the access externally to the originating organization, and an increasing volume of data being produced by sensors, the EOSC challenge is to enable data (and related analytical software) discovery, browsing, and access on-the fly.

3.4.4.2.2 GAIA-X

GAIA-X has included a "sovereign data exchange" service in its suite of federation services since its inception. In contrast to EOSC, which works with the research community and largely with open or FAIR data, GAIA-X seeks to work with the industrial community, where participants want to strictly control what data is shared and with whom.

Currently there are no proven and guaranteed, technical solutions for general "secure data sharing" -- fundamentally once a shared piece of data is decrypted and "learned" by the data consumer, it is impossible to "unlearn" it. Mechanisms have been proposed (e.g. distributed ledger technology or fully homomorphic encryption), but they are at early stages of implementation.

GAIA-X' solution to this problem has policy components and technical components. The policy components include the specification, in the Draft GAIA-X Policy Rules Document [16], that Data Providers must annotate any data that can be shared with clear and specific Use Conditions (ideally using a structured schema such as ODRL), and that Data Consumers must respect those Use Conditions. The GAIA-X Architecture Document [12] describes technical components to support Sovereign Data Exchange, namely a Data Contracting Service to record the terms of any data sharing, and a Data Exchange Logging Service to record the actual transfers of data. The GAIA-X Architecture Document mentions "Enforcement", including monitoring of data usage to detect violations and "subsequent (compensating) actions", but the Policy Rules Document does not require data consumers to comply with such an enforcement policy or mention any penalties for violations. More clarity about these components and their operation, as well as additional policy and technical components, may be needed to provide adequate assurance to data providers that shared data will be secured as desired.

Note that the proposed approach primarily addresses bilateral data exchange (between a provider and a consumer) rather than multilateral data sharing (e.g. among the members of a data ecosystem) – which may be a more important use case for GAIA-X participants. The EU's proposed Data Governance Act also considers this question by describing entities that might enable data sharing, namely a Data Intermediary, Data Collective or Data Altruism Organization.

For both bilateral and multilateral data sharing, it will also be important to establish clear policies, and possibly provide supporting technology, to track Use Conditions when multiple data objects are processed together. Such processing can create more sensitive data, requiring stricter Use Conditions, as well as less sensitive data, allowing Use Conditions to be relaxed.

3.4.5 Support and engagement

3.4.5.1 EOSC

The EOSC Helpdesk is the entry point and ticketing system/request tracker for issues concerning EOSC services. New service providers of EOSC can integrate into the Helpdesk and this results in a corresponding support topic listed on the Helpdesk user interface for users to ask questions or raise issues directly to the provider. The provider's support team receives notifications about tickets that are assigned to this topic by the users, or by the ticket handler team.

Besides this, the training and support programme of EOSC is expected to empower various EOSC stakeholders to use data, services, and software. Training activities will address skills development on two fronts: skills for researchers and data practitioners with consolidated cross-provider training modules, and skills for content and resource providers on joining EOSC and aligning services offered.

Data curation services – communicating with appropriate communities about the availability and utility of certain categories of data – are expected to be provided within the different scientific communities operating within the EOSC framework.

3.4.5.2 GAIA-X

GAIA-X has not specified or described training, outreach or advisory services.

3.4.6 Service Coordination, Integration, Operations Management

3.4.6.1 EOSC

The EOSC Service Management System represents the entirety of activities performed by the providers that contribute to the EOSC core to plan, deliver, operate and control the services offered to EOSC. It also covers (to different extent) the activities of those service providers that have been onboarded to EOSC via the EOSC Portal. The activities carried out in the context of the SMS are structured and organised into processes and procedures according to the FitSM IT Management standard[21]. FitSM is a free, pragmatic, lightweight and achievable standard aimed at facilitating service management in IT service provision, including federated scenarios. It was developed via an EC funded project in the research domain and explicitly supports complex federated scenarios such as those in EOSC. By defining requirements, the 14 processes of FitSM help service providers.

3.4.6.2 GAIA-X

Although multi-provider service provision is an explicit objective of GAIA-X, it has not proposed support for Service Coordination, Integration, Operations Management.

3.5 Services for End Users

3.5.1 EOSC

The specific end-user capabilities offered in EOSC depend on the specific objectives of a data ecosystem. The customer-facing services to be offered by EOSC cover a wide range; in the EGI-ACE implementation case they are exemplified by the following services:

- The components that enable federation of distributed cloud and high-throughput computing (HTC) systems and services.
- Compute and storage capacity.
- Federated data management including data transfer services.
- Federated compute management including orchestration and workload management services
- Federated trust and identity management
- AI and ML on demand services
- PaaS services for web-based analytics enabling users to create and manage their own work environments.

[21] https://www.fitsm.eu/downloads

- data ecosystems integrating data and thematic tools that offer community-specific capabilities with low entry barriers.

3.5.2 GAIA-X

GAIA-X aspires to offer a wide range of services to its participants, although this catalogue of possible offerings has not been assembled. It is expected that services will include cloud-based services (IaaS/PaaS/SaaS) as well as data analytics, artificial intelligence and machine learning, high performance computing and even quantum computing. Compute assets could range from traditional Intel architectures, to GPUs, FPGAs and specialized ASIC-based appliances.

4 Recommendations

We conclude the paper highlighting recommendations for future collaboration.

Recommendation 1: Data sharing and protection

Share good practices on cross-domain/organization data sharing policies and FAIR best practices that enable data exploitation and validate these with concrete cross-domain and cross-initiative use cases, and collaborate on data protection and GDPR compliance.

As the science and industry use cases from the EOSC and GAIA-X demonstrate, both initiatives strive to increase the ability to share data across domains and organizational and national boundaries in a secure and trustworthy manner. Depending on the use case, the ability to effectively enable this requires the adoption of policies, common technical tools and security fraeworks that need to be tested against relevant use cases.

EOSC Task Forces: Semantic interoperability, FAIR metrics and data quality

GAIA-X Committees, WGs and Open Work Packages: Data Spaces Business Committee, Users WG, Portfolio WG, OWP Self-Description, OWP Architecture of Standards, OWP Data Sovereignty, OWP Product & Service Board

Recommendation 2: Compliance and federated governance

***Adopt compatible models for data ecosystem business models, federation architectures and related roles, and service integration and delivery models including concerted approaches to develop resource and service registries*.**

EOSC and GAIA-X will succeed to the extent stakeholders can gain benefits from the collaboration. Extracting value from data for scientific and economical purposes in a federated environment will involve multiple suppliers that participate in a common enterprise system with different roles and business models. EOSC and GAIA-X should collaborate to study federation architectures and standards to define and implement the related enterprise systems.

EOSC Task Forces: Rules of Participation compliance monitoring, Defining funding models for EOSC

GAIA-X Committees, WGs and Open Work Packages: Data Spaces Business Committee, Policy & Rules Committee, Users WG, Portfolio WG, Architecture WG, OWP Compliance

Recommendation 3. Technical interoperability

Align the definition of technical and operational interoperability frameworks that make data exchanges between EOSC and GAIA-X possible and share experience on existing technical solutions.

Technical and operational interoperability standards and best practices in EOSC and GAIA-X should support and facilitate the exchange of research and industrial data for relevant use cases. EOSC and GAIA-X should collaborate to adopt an interoperability framework for example through the compliance to a common set of standards. Cross-fertilization opportunities are possible in various areas such as permanent identifier policies and services, federating services, FAIR maturity models and metadata standards.

EOSC Task Forces: PID policy and implementation, Technical interoperability of data and services, AAI architecture

GAIA-X Committees, WGs and Open Work Packages: Technical Committee, Users WG, Portfolio WG, Architecture WG, OWP Self-Description, OWP Architecture of Standards

ACKNOWLEDGEMENTS

We thank Enol Fernandez for input on the EGI-ACE service architecture; we thank colleagues of the EOSC Future technical coordination team for EOSC-hub implementation approaches of the EOSC architecture; we thank colleagues from EOSC Enhance for input on the EOSC profiles work; we finally thank Owen Appleton (EGI Foundation), Bob Jones (CERN and EOSC Board of Directors) and Ronan BYRNE (HEAnet and EOSC Board of Directors) for revisions and advice.

REFERENCES

[1] Turning FAIR into reality, Final report and action plan from the EC expert group on FAIR data (2018), DOI: 10.2777/1524 (https://op.europa.eu/s/o8fU).

[2] Towards a European strategy on business-to-government data sharing for the public interest. Final report prepared by the High-Level Expert Group on Business -to-Government Data Sharing (2020), DOI: 10.2759/731415 (https://op.europa.eu/s/o8fV).

[3] Sustainable and Smart Mobility Strategy – putting European transport on track for the future. EC Communication COM (2020) 789 Final (https://op.europa.eu/s/o8Oe)

[4] Powering a climate-neutral economy: An EU Strategy for Energy System Integration. EC Communication COM (2020) 299 Final (https://op.europa.eu/s/o8Og).

[5] A Farm to Fork Strategy for a fair, healthy and environmentally-friendly food system. EC Communication COM (2020) 381 Final (https://op.europa.eu/s/o8Oc)

[6] Stepping up Europe's 2030 climate ambition Investing in a climate-neutral future for the benefit of our people. EC Communication COM (2020) 562 Final (https://op.europa.eu/s/o8Oi).

[7] The European data strategy: shaping Europe's digital future, 2020, doi:10.2775/987881 (https://op.europa.eu/s/o8fW).

[8] Strategic Research and Innovation Agenda of the European Open Science Cloud, v. 1.0, 2021 (https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf).

[9] Solutions for a sustainable EOSC, (https://op.europa.eu/s/o8gd)

[10] EOSC strategic implementation plan (2020), DOI: 10.2777/87077 (https://op.europa.eu/s/o8ga).

[11] ESFRI Data, Computing and Digital Research Infrastructures, Roadmap 2018, (http://roadmap2018.esfri.eu/media/1051/rm2018-part2-digit-20.pdf).

[12] GAIA-X Architecture Document, Release 21.03 (2021) (https://gaia-x.eu/pdf/Gaia-X_Architecture_Document_2103.pdf).

[13] The Potential of Cloud Federation, Horizon Cloud Project (2020), DOI: 10.5281/zenodo.4733565 (https://zenodo.org/record/4733565).

[14] EOSC AISBL Statutes (https://www.eosc.eu/sites/default/files/EOSC_Statutes.pdf).

[15] EOSC European Partnership (proposal draft), May 2020 (https://ec.europa.eu/info/sites/default/files/research_and_innovation/funding/documents/ec_rtd_he-partnership-open-science-cloud-eosc.pdf).

[16] GAIA-X: Policy Rules and Architecture of Standards, 2020 (https://www.bmwi.de/Redaktion/EN/Publikationen/gaia-x-policy-rules-and-architecture-of-standards.pdf?__blob=publicationFile&v=6).

[17] Policy Rules Document, PRD 21.04 (https://gaia-x.eu/pdf/Gaia-X_Policy%20Rules_Document_2104.pdf).

[18] EOSC interoperability framework (2021), DOI: 10.2777/620649 (https://op.europa.eu/s/o8hi).

[19] European charter of access for research infrastructures (2016), DOI: 10.2777/524573 (https://op.europa.eu/s/o8ht).

[20] EOSC – A tool for enabling Open Science in Europe; Appleton, O., Asmi, A.; Bird, I. et alt. (2020), DOI: 10.5281/zenodo.4044010 (https://zenodo.org/record/4044010).

[21] EOSC Rules of Participation (2021), DOI: 10.2777/305411 (https://op.europa.eu/s/o8gV).

[22] Six Recommendations for Implementation of FAIR Practice, 2020, DOI: 10.2777/986252 (https://op.europa.eu/s/o8Fk).

[23] A Persistent Identifier (PID) policy for the European Open Science Cloud (2020), DOI: 10.2777/926037 (https://op.europa.eu/s/o8Fm).

[24] AARC Blueprint Architecture (https://aarc-project.eu/architecture/).

[25] Destination Earth: Ecosystem Architecture Description, JRC Technical Report; Nativi, S., Craglia, M. (2020), DOI:10.2760/08093 (https://op.europa.eu/s/o8mC).

[26] Data Use Ontology (https://bit.ly/2Sr2tsh).

[27] Specifications of the EOSC Provider Profile (https://eosc-portal.eu/providers-documentation/eosc-provider-portal-provider-profile).

[28] ISO / IEC. IT Security and Privacy – A framework for identity management: Part 1: Terminology and concepts (24760-1:2019(en)). ISO / IEC. https://www.iso.org/obp/ui/#iso:std:iso-iec:24760:-1:ed-2:v1:en

[29] EOSC Authentication and Authorization Infrastructure (AAI): report from the EOSC Executive Board Working Group (WG) Architecture AAI Task Force (TF); 2021, DOI: 10.2777/8702 (https://op.europa.eu/s/o8Fo).