Grant Agreement N°: 871920
Topic: ICT-13-2016
Type of action: CSA


**Horizon Cloud – The Forum for Strategy Focused Cloud Stakeholders**

# The Potential of Cloud Federation

## Appendix 15 of Green Paper v0.9

Revision: v0.9

| | |
|---|---|
| **Work package** | WP 3 |
| **Task** | T3.1 |
| **Submission date** | |
| **Deliverable lead** | EGI Foundation |
| **Version** | 0.9 |

# Abstract

Federations are multi-organizational alliances designed to achieve mutually agreed objectives. Federations can be an effective way to deliver coordinated IT services particularly in support of public good objectives. Organizational principles and best practices of federations are described, as well as mechanisms for technical and service organization and coordination. The use of federated IT service structures to enable effective cloud-to-edge integration as well as secure access, sharing and analysis of distributed data is also described.

**Keywords:** cloud, cloud computing, federation, federated cloud, federated edge, federated data, technical alliance, marketplace

### Document Revision History

| Version | Date | Description of change | List of contributor(s) |
|---|---|---|---|
| V0.82 | 2020/09/25 | Complete Revision from earlier versions | Mark Dietrich, Tiziana Ferrari Gianni dalla Torre, Massimiliano Claps, Federico Michele Facca |
| V0.85 | 2020/10/30 | Aligning Service Management Functions with Federation Business Models, Reference to Additional Privacy Preservation Projects | Mark Dietrich |
| V0.9 | Release with Green Paper V1.0 | Final formatting | Mark Dietrich |
| | | | |

### Disclaimer



| Project co-funded by the European Commission in the Horizon Cloud Programme | | |
|---|---|---|
| **Nature of the deliverable:** | R | |
| **Dissemination Level** | | |
| PU | Public, fully open, e.g. web | √ |
| CI | Classified, information as referred to in Commission Decision 2001/844/EC | |
| CO | Confidential to H-CLOUD project and Commission Services | |

** R: Document, report (excluding the periodic and final reports)*

*DEM: Demonstrator, pilot, prototype, plan designs*

*DEC: Websites, patents filing, press & media actions, videos, etc.*

*OTHER: Software, technical diagram, etc*

# EXECUTIVE SUMMARY

Federations are currently receiving extra attention as mechanisms to increase service capacity and capabilities in a multi-supply environment to augment each individual federation member's ability to serve a wider user base. In October 2019 the governments of Germany and France announced the Gaia-X federated cloud initiative[1], with a strong focus on creating a federated cloud and data capability. The EU discusses both cloud federation and data spaces (related to federated data) in its communication "A European Strategy for Data"[2] (EUSD). There is also ongoing research on "federated cloud technology", much of it EC-funded and adopted by digital infrastructures to address challenging data processing requirements of research communities.

## The Structure of This Document

Despite this increased level of interest, the process for creating an effective federation is poorly understood. This document addresses this gap in the following ways:

- The essential characteristics of federation are described in section 2. Within that organizational model several possible federation business models are described along with their benefits for both federation members and customers. In addition, organizational dimensions of standards compliance and coordinated service management are described. Examples of federations are presented, mostly in the research domain, and mostly federations of infrastructure.
- Section 3 examines research on multi-organizational alliances and business relationships to identify best practices, and valuable management and governance functions, that should be incorporated into business/operating plans of federations seeking sustainability and effectiveness. There is very little research into best practices for federations per se, so we explore organizational research on similar organizational forms to gain applicable insights.
- Section 4 presents a roadmap for constructing a federation. It translates the learnings from organizational research into concrete, detailed steps that should be followed in establishing any new federation.
- Section 5 explores the services that might be provided through a federation to customers, as well as the services required for the federation to operate effectively. Example federation architectures are presented, as well as frameworks for those architectures, that can organize those services and underpin the technical governance of the federation.
- Section 6 considers how federation can support the expansion and adoption of edge computing.
- Section 7 discusses how data sharing and "data spaces" can be supported by federated data services, and specifically explores how federated technology approaches can address security and privacy issues associated with data spaces.

## Key Conclusions

The EUSD proposes the creation of a European Cloud Federation (EUCF) as well as a cloud services marketplace. With this premise, several challenges are identified along with responsive recommendations:

**S-F Challenge 1: Coordinated/federated approaches must be structured around the objectives of their stakeholders, balancing community focussed initiatives with pan-**

[1] Project GAIA-X
[2] EC. Communication: A European strategy for data. 2020.

**European solutions.** The needs of different stakeholder communities must be balanced against the need for common or aligned solutions. The effectiveness of any federation will depend strongly on the clarity of its value proposition and how it is constituted to realize that value proposition.

Recommendations prescribe a number of steps required for success:

- Develop detailed business cases for identified use cases in each of the nine sectoral data spaces described in the EUSD that quantify the societal gains and costs to achieve the desired benefits and ascertain feasibility and related ICT innovation needs.
- For each business case, select the most appropriate federation business model that fulfils the requirements while providing the best value with the least effort.
- Create an open infrastructure and testing capability that could flexibly support demonstrations, proofs of concept and pilots of how federated cloud and federated data solutions could be assembled, operated, managed and governed, including collection of data that would validate the business cases developed earlier.
- Support the creation of multiple EUCF-affiliated initiatives and their cross-domain collaboration, which will specify domain-specific use cases, objectives and beneficiaries, federation partners and stakeholders, governance and decision-making mechanisms, scope of possible federated activities, and applicable business models.
- Develop a lightweight model for the EUCF as an umbrella coordinating body of sector- or use-case-focussed European federated cloud initiatives, supporting coordination of their research and innovation activities, cross-sector collaboration on interoperability, facilitating best practice operations, and providing relevant shared services such as certification activities.
- Implement a EUCF with a phased approach that flexibly aligns activities across multiple domains, and that allows achievement of “quick wins”. Pilot projects and demonstrators will clarify requirements and identify applications and use cases where an early version of the EUCF can achieve success, which in turn will build credibility and support.
- Set up the EUCF following known organizational recommendations.
- Evolve existing best practices and standards (e.g. ITIL, FitSM, etc.) for federated service management to ensure federated cloud initiatives have reference requirements, processes, procedures and policies that ensure the compatibility of service delivery and planning across different initiatives. Develop “starter kits” to assist with implementation of each federated business model, with sample templates for required governance and service management processes (definitions, roles, process maps, etc.).

**S-F Challenge 2: Defining, Evolving, Selecting, Agreeing on and Managing the Architecture, Technical Standards and Tools for Federated Clouds and Distributed Data Access and Exchange.** Creating a distributed yet federated, technically effective data-processing system is an active subject of research -- many technical approaches are being studied, and many technical approaches are in use in the community, but they are not converging into “standards” because the underlying technologies are rapidly evolving and because the scope of integration is expanding from the data centre out to the more heterogeneous edge computing environment. Where distributed capabilities need to work together, there must nevertheless be an agreement on the framework of the system (the architecture) and the standards to be adopted within that framework, even in the face of this rapid change. Establishing a platform architecture enables technical discussions to be modularized and compartmentalized and facilitates agreement on standards that enable specific services to interoperate.

Responsive recommendations include:

- Develop and evolve a Federated Cloud Reference Architecture (FCRA), to the extent possible incorporating the NIST CFRA, EGI and Gaia-X's technical architectures, and evolve it to ensure conformance of emerging federated cloud initiatives. This should specifically characterize how practical compliance frameworks and service catalogues would align with the EUSD's contemplated "Cloud Rulebook" and "Service Marketplace" concepts.
- Create and maintain a federated cloud interoperability framework as an evolving suite of technology, standards and tools that are consistent with the FCRA allowing interoperation within a given federation and across multiple federations, compliance with European values and identify components that are interoperable. This suite of components would help EU customers navigate the many options for cloud-based solutions and would help formalize how they are described and the possibilities for integration.

Coordinate research and innovation activities for funding in Horizon Europe by aligning cross-domain cross-use case research and innovation activities of common interest for different federation stakeholders to increase synergies, innovation potential and avoid duplication across the industry, research and public administration sectors.

**S-F Challenge 3: Federated data has great potential to support secure, private sharing of data held by many different organizations.** Best practice roadmaps are urgently needed to ensure federated data sharing initiatives are established and operated efficiently while preserving and ensuring the highest level of trust that affected sensitive data will be kept private and secure.

Recommendations include:

- Create guidelines for implementing different data sharing approaches using federated data platforms.
- Support efforts to increase semantic interoperability for data within and across sectors and include harmonization of data usage models to enable automated, yet secure and appropriate, data sharing.
- Develop regulatory sandboxes that allow experimentation and scaled-up testing of privacy-preserving technologies.
- Continue support for research, innovation and deployment of existing privacy-preserving technologies in practical application domains. These technologies are at various technological readiness levels (TRL) and would benefit from continued investment and support for early stage adoption and deployment.
- Support creation of technical standards for preserving privacy. Such standards would provide risk assessment tools, test suites for validation of performance, as well as evaluation of data for sensitive content.
- Continue support for research, innovation and deployment of distributed data analytics tools, as well as data placement tools, that minimize security privacy risks and maximize speed, computational and network efficiency as well as energy efficiency.

Continue support for research, innovation and deployment of distributed data analytics tools, as well as data placement tools, that minimize security privacy risks and maximize speed, computational and network efficiency as well as energy efficiency.

# TABLE OF CONTENTS

**1 INTRODUCTION ..... 13**

**2 UNDERSTANDING FEDERATION AND THE FEDERATED ORGANIZATIONAL MODEL 15**

2.1 The Role of Federations in Information Technology ..... 15

2.2 Essential Characteristics of Federations ..... 15

2.3 Key Dimensions of Federation Business Models ..... 17

2.4 Standards Compliance ..... 17

2.5 Federation Business Models and Associated Benefits ..... 18

2.5.1 Open Marketplace ..... 19

2.5.2 Structured Marketplace ..... 20

2.5.3 Reseller ..... 21

2.5.4 One Stop Shop ..... 22

2.5.5 Full Integrator ..... 23

2.6 Service Management Coordination ..... 23

2.7 The Benefits of Federation ..... 25

2.7.1 Results of Survey on Benefits of Cloud Federation ..... 27

2.8 Examples of Federation ..... 29

**3 BEST PRACTICES FOR STRUCTURE AND GOVERNANCE FROM OTHER ORGANIZATIONAL MODELS ..... 37**

3.1 Multi-Organization Alliances ..... 37

3.2 Social Enterprises ..... 39

3.3 Technical Alliances or “Platforms” ..... 40

**4 PROCESSES FOR CREATING A SUSTAINABLE MODEL FOR A EUROPEAN CLOUD FEDERATION (EUCF) ..... 45**

4.1 Alignment of Mission and Vision ..... 45

4.2 Governance Model and Stakeholders ..... 45

4.3 Business Case Development ..... 46

4.4 Validating the Business Cases ..... 48

4.5 Determining the Right Scope for Each Initiative ..... 48

4.6 Value of the EUCF as Umbrella Coordinator ..... 49

4.7 Phased Implementation Approach ..... 50

4.8 Operating Structures ..... 50

**5 FEDERATED CLOUD ARCHITECTURE AND SERVICES ..... 53**

5.1 The Role of a Federation Architecture ..... 53

5.2 Federation Services ..... 55

5.2.1 Master List of Possible Services ....55
**5.2.1.1 Support Services Available to Customers, Providers and Externally ....55**
**5.2.1.2 Technical Services Available to Customers ....55**
**5.2.1.3 Federated Services Available to Customers ....56**
**5.2.1.4 Federated Services Available to Service Providers and Integrators ....57**
5.2.2 Desired Federation Services: Survey Results....58
5.3 Architectures that Encompass Federation Services ....60
**6 FEDERATED EDGE SOLUTIONS....66**
**7 FEDERATED DATA SOLUTIONS....69**
7.1 Federated Data as a Flexible Approach to Data Sharing ....69
7.1.1 Contractual Data Sharing Structures....70
7.1.2 Collective Data Sharing Approaches ....70
7.1.3 Application Programming Interfaces (APIs) ....71
7.1.4 The IDSA Reference Architecture ....71
7.1.5 Data Usage Models ....72
7.2 Technical Challenges to Data Sharing ....74
7.2.1 Increased Granularity and Detail in Data ....74
7.2.2 Embedded Data Protection Solutions Needed....75
7.3 Tight Coupling of Data and Data Processing....76
7.4 Technical Services Required to Support Federated Data and Data Spaces ....77
7.4.1 Services Enabling Data Spaces ....77
7.4.2 Services to Support Privacy Preservation....78
**8 CHALLENGES AND RECOMMENDATIONS....81**

# LIST OF FIGURES

**No table of contents entries found.**

# LIST OF TABLES

**No table of contents entries found.**

# ABBREVIATIONS

# 1 INTRODUCTION

Federations are currently receiving renewed attention as mechanisms to increase service capacity and capabilities in a multi-supply environment to augment each individual federation member's ability to serve a wider user base. As the H-CLOUD Green Paper[3] concludes, inherently distributed systems can benefit from federation. Important examples exist in public administration, healthcare and transport/mobility and research, as they need to enable the secure access, sharing and analysis of sensitive data already being stored and managed by multiple players in a community -- often residing in private cloud infrastructure. On the supply side, in October 2019 the governments of Germany and France announced the Gaia-X federated cloud initiative[4], with a strong focus on creating a federated cloud and data capability. The EU discusses both cloud federation and data spaces (related to federated data) in its communication "A European Strategy for Data"[5] (EUSD). There is also ongoing research on "federated cloud technology", much of it EC-funded and adopted by digital infrastructures to address challenging data processing requirements of research communities.

Federated organizational models can be used in many domains, not just information technology. Federation is often discussed in the context of federated cloud (multi-cloud integration) and federated data (data sharing)[6]. The analysis below refers to both federated cloud and federated data more generally as federated IT service structures, or "federations" for short.

This document examines federation as an organizational model and provides recommendations on the best approach to creating effective, sustainable federations. In the context of federated IT service structures, technical issues related to architectures and the services to be provided are also critical, and best practices and guidance are presented. Finally, we explore the application of federation concepts to enable seamless "cloud-to-edge" computing and data sharing initiatives.

[3] Add reference to v1.0 of Green Paper when available

[4] Project GAIA-X

[5] EC. Communication: A European strategy for data. 2020.

[6] Data sharing, for reasons of privacy, security, and both energy and technical efficiency, is increasingly likely to involve controlled access to distributed data sets held by different data stewards (federated data), rather than gathering data into a single database or data lake (data pooling).

# 2 UNDERSTANDING FEDERATION AND THE FEDERATED ORGANIZATIONAL MODEL

Federation is one form of multi-organizational alliance in which some processes and related policies and activities are governed and coordinated in a collaborative way, and sometimes delegated to a central "federating entity" by the federation members, while other processes, policies and activities remain the responsibility of the members of the federation. Federation can be applied to many endeavours, including research, e-infrastructure services, data sharing capabilities among public administration/public safety organizations, coordinated delivery of health and social services in the community, and business partnerships. Agricultural cooperatives are very similar to federations. Ideally there should be some type of asset or resource, common to all the partners, which can be shared across the federation to better serve customers and users.

## 2.1 The Role of Federations in Information Technology

As discussed in section 4.3 of the H-CLOUD Green Paper, many cloud customers need to integrate services from multiple public cloud providers and with their own private cloud capabilities depending on their use case. This integration may be triggered when customers want to combine best-in-class services from different providers, to combine service territories across national borders, or when groups of organizations (e.g. in healthcare) want to share or combine data or data-processing resources funded by multiple independent national funding agencies.

When this service integration is performed by a single customer, it is called "multi-cloud". Customers sometimes hire outside consultants to perform the desired integration.

When this integration is performed collectively by multiple partners, this is called "community cloud", "industrial cloud/B2B platform" and "federated cloud", depending on the circumstances. All of these require cooperation and coordination of the participating service providers. Federations encompass the governance, processes, policies and technical solutions used by multiple service providers to cooperate and coordinate their services, focussed particularly on coordinating service planning, delivery and management. The providers themselves are the members of the federation.

If services from multiple providers do not need to be integrated, customers' needs might be addressed by marketplaces: central platforms providing discovery and access services (discussed in detail in sections 2.5.1 and 2.5.2 below). In this case, coordination among providers is achieved through adherence to a single business model (set by the marketplace's operator) and its rules of participation.

## 2.2 Essential Characteristics of Federations

Federations exhibit several essential characteristics:

- **A federation is an alliance of multiple organizations.** This is true by definition, but this also means that a federation is a collective entity that is not necessarily "owned" by any single organization.
- **Participating organizations are "members" of the federation and collaborate for common goals**. Federation members typically "join" the federation by acknowledging the common goals of the federation, 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 -- which 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 resources 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.

These essential characteristics are reflected in the NIST Cloud Federation Reference Architecture[7] (NIST CFRA, section 2.1)

While these are the "essential" characteristics of federations, some federations lack one or more of them (see further discussion in Section 2.8: Examples of Federation). Such gaps reflect the maturity and/or scope of these federations, rather than casting doubt on the "essence" of these characteristics and can create organizational and operational challenges. In the most extreme case, research into "federated cloud" technologies typically does not contemplate the organizational steps required to get multiple federation participants to agree on the standards and processes implied by the technology under study. Nevertheless, those steps must occur in order to realize any of the promised technical benefits. (Section 4 discusses this process in detail.)

From a strategic point of view, all federations need to maintain and evolve their mission and vision to retain alignment across the federation members. Even mature federations displaying all of the above characteristics face organizational challenges. The need to evolve a federation's agreed goals and objectives is a challenge that can affect the relationship between a federation and its members -- both positively and negatively[8].

"Collaboration for common goals" is an essential characteristic of federations. For different federations, those agreed goals can range from self-interest (e.g. "improving the profitability of federation members") to the public good (e.g. "advancing scientific knowledge", "improving health care"). A federation's commitment to public good objectives can create benefits for both providers and customers:

- For service providers (members), a higher purpose can make it easier to agree on and conform to more stringent technical standards and operational processes. For example, providers might be more willing to accept independent certification of technical compliance in pursuit of improved health care, even though such certification might increase their costs more than could be justified by possible increases in revenue.
- For customers, a federation's commitment to a higher purpose can increase trust in the federation and its services and motivate greater take-up of those services.

Section 3 includes research into how adopting public good objectives can enable superior

[7] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500-332.pdf

[8] Hummel J.T. (2019). Collaboration and innovation between heterogeneous actors. PhD dissertation. Amsterdam NL: Rozenberg Publisher. ISBN 978–90–3610–557–6. https://research.vu.nl/en/publications/collaboration-and-innovation-between-heterogeneous-actors

performance of multi-organization alliances, social enterprises as well as technical alliances.

## 2.3 Key Dimensions of Federation Business Models

Section 5 below explores details of federated architectures and services, but it is important to note that, even among federated research clouds, there is wide variation in the range of technologies, tools and/or standards supported, and wide variation in the services provided. The variation depends on the objectives of the different federations, and the processes, policies and activities that members have agreed to federate in pursuit of those objectives.

How a federation delivers services to customers fundamentally defines its value proposition for those customers, as well as the benefits experienced by federation members. In the most sophisticated case, all of a federation's services are fully interoperable and integrated and make it easy for a customer to discover, select, access and use the services it needs. However full integration may not be possible or even desired by all customers, or even all providers, so a range of federation business models are possible[9]. These models vary along three primary dimensions:

- the degree to which providers conform to technical, organizational and legal standards,
- the degree to which services are integrated for the customer, and
- the degree to which service planning, delivery and management are coordinated among members.

These dimensions are correlated to one another. For example, for the most lightweight federation (an "Open Marketplace", described below), where the different services offered by service providers are not integrated or interoperable, service delivery and management are also not coordinated, and there may be little conformance to standards. By contrast, tightly integrated federations (operating as a "Full Integrator"), offering fully integrated services, strongly coordinate service providers and enforce compliance with a number of relevant standards and policies. Between these two extremes, several other models are possible, allowing federations to adapt to the needs of their community and the use cases they fulfill.

## 2.4 Standards Compliance

One essential characteristic of federations is the agreement of their members to conform with technical standards and operating procedures that enable the interoperation, collaboration and sharing required by its customers. Well-integrated services depend on compliance with appropriate organizational, technical and legal/regulatory standards.

- Complying with agreed service management standards and procedures enables the coordination of service management, which in turn enables the most complete integration of services. The scope of these standards and procedures is determined by the business model of the federation, discussed in section 2.5.
- Complying with agreed "quality" standards, primarily non-functional characteristics of their services, such as use of standard forms of contract, energy-efficiency, and eligibility for certain payment or procurement arrangements.
- Complying with certain technical standards (e.g. HL7[10] standards for the exchange, integration, sharing, and retrieval of electronic health information), or agreeing to use

[9] This presentation of business models reflects analyses found in "FedSM D3.1: Business models for Federated e-Infrastructures" (https://zenodo.org/record/3982794#.XzVSrxNKj_Q) and "EOSC Pilot D5.1: Initial EOSC Service Architecture" (https://eoscpilot.eu/sites/default/files/eoscpilot-d5.1.pdf).
[10] http://www.hl7.org/implement/standards/index.cfm?ref=nav

common software tools or components (e.g. OpenStack[11] or Kubernetes[12]) may be required for services to interoperate or be integrated.

- Legal and regulatory compliance is obviously required, but transparent and trustworthy disclosure is critical, since the legal and regulatory obligations that apply to each provider, and potentially to different services offered by that provider, may differ from the obligations expected by a given customer. This is a particular issue when services are provided across borders within Europe and from outside Europe.

Compliance in each of these domains is handled differently:

- The extent of a service provider's compliance with service management standards and procedures is assessed and monitored by the federating entity and typically affects the degree to which the provider can participate in various federation offerings to customers. For example, a provider would not be allowed to list "fully integrated services" in the federation service catalogue unless the relevant service management standards were met.
- Compliance with some quality standards (such as contract terms) can be asserted by the provider itself, while other factors, such as energy efficiency, might need to be evaluated by external services.
- Compliance with technical standards might also be asserted by the provider itself or evaluated by external services or by the federating entity as agreed within the federation. (Such "internal" compliance assessment might be seen as a benefit of federation membership.) The use of agreed software tools or components could be handled through provider assertions, rather than any formal validation process. More complex assertions about technical standards compliance, interoperability and integration could be tested by the provider using test suites provided by the federation, or directly tested and validated by the federation as part of the integration process. In the end, either the services can be integrated, or they cannot.
- Assertions of legal and regulatory compliance would primarily be the responsibility of providers themselves, although the federation could maintain a registry of proven compliance and make this information available in service selection tools such as an online service catalogue. Some aspects of legal and regulatory compliance are matters of fact, such as the location of infrastructure assets and the jurisdictional home of the owner of those assets; some compliance issues are more complex, e.g. assessing compliance with ISO 27000 cybersecurity standards. The federation could also provide advisory services to assist providers in achieving needed levels of compliance (for example guiding providers in establishing security credentials or GDPR compliance).

Interactions among these different forms of compliance can also affect other dimensions. Even if several services are tightly integrated, with fully coordinated service management, differences in legal or regulatory status among the different services will affect the compliance of the integrated service. For example, lack of GDPR compliance by one component service might make the integrated service non-compliant. Similarly, separate compliant services performed in different jurisdictions, might become non-compliant if they are integrated across a border.

## 2.5 Federation Business Models and Associated Benefits

The degree to which services need to be integrated for a federation's customers drives the choice of federation business model, which can range from lightweight (*Open Marketplace*) to tightly integrated (*Full Integrator*). Each type of business model enables a range of benefits for

[11] https://www.openstack.org/
[12] https://kubernetes.io/

both customers and service providers (federation members).

Different business models may be used for different sets of services, depending on the ability of those services and service providers to participate in different levels of integration. Different business models imply different economic models for the federation and pricing structures for customers, including recovery of investments associated with the initial and ongoing operation of the federating entity. This could range from surcharges on marketplace transactions to customized pricing for *One Stop Shop* or *Full Integrator* service offerings. (An economic analysis of federation business models is outside the scope of this document.)

### 2.5.1 Open Marketplace

The most lightweight form of federation is an *Open Marketplace*, which provides tools for discovery, comparison and selection of services available through the federation. The federation architecture defines categories for different services, as well as appropriate technical and operational standards against which service providers can disclose their own compliance. The customer must confirm standards compliance as well as service interoperability, contract with each service provider, integrate and use the services and interact with providers for ongoing service management. Clearly, integration and/or interoperability are not required by all customers, as evidenced by the success of marketplaces such as Apple's App Store.

Key features of an *Open Marketplace*:

- Architecture and standards are defined by the federation, allowing providers to consistently and reliably list and promote their services. The federation provides unbiased technical and standards expertise in order to manage the architecture and ensure coherent information is provided about listed services.
- The federation operates the marketplace itself and conducts joint outreach and promotion on behalf of the members.
- From a governance perspective, in addition to facilitating agreement on the architecture and standards that will form the framework for the marketplace, the federation provides guidance on anti-trust issues to insure transparent rules for marketplace access and "truth in advertising".

For customers, an *Open Marketplace* offers the following benefits:

- Awareness: Searching for services in a single marketplace is easier than having to contact multiple vendor sources.
- Discovery: An *Open Marketplace* should be able to offer a wider variety of service offerings.
- Selection: The *Open Marketplace's* consistent information makes it easier and cheaper for customers to identify, compare and select the services they need. The customer has greater trust in the information provided, since it is both consistent and "managed" by an entity other than the provider.

For a service provider, an *Open Marketplace* increases awareness of its services and provides access to a larger pool of potential customers, who are attracted to the ease-of-use of the marketplace.

An *Open Marketplace* could set minimum quality and technical (interoperability) standards as a condition for participation and listing of service, but without requiring tests of compliance. Such standards might be similar to:

- The EUSD's proposed "cloud rulebook", codifying the various non-functional requirements that apply to European cloud service providers[13],
- Gaia-X's compendium of cloud services requirements (part of its "Policy Rules and Architecture of Standards"[14]).

The original Helix Nebula project[15] set out to create a federated *Open Marketplace* for research services with commercial cloud providers, but found that buyers from the research community did not have a clear understanding of their own cost structures for cloud-style computing, and they also lacked confidence that the services available through the marketplace were appropriate for their own purchase.

The UK government's G-Cloud[16] service is an *Open Marketplace*, restricted to cloud services that claim compliance with certain privacy and security standards and eligible for purchase by UK public administrations through standard agreements. Recognizing the needs of publicly funded customers (public administration, healthcare, research), an *Open Marketplace* should also support controlled identification of services and service providers that have been qualified for procurement from various customers. This might be accomplished by providing additional filters for selection, or by "white labelling" the marketplace capability for certain groups of customers, so that they only see eligible services and providers in the marketplace.

Other examples of *Open Marketplaces* include Cloud28[17], CloudWatt (now defunct), Numergy (also defunct), and Deutsche Borse[18].

### 2.5.2 Structured Marketplace

A *Structured Marketplace* extends the *Open Marketplace* model by offering certification or validation of providers' claims about compliance with quality and/or technical standards, particularly interoperability. Testing and compliance would be required by the federation, potentially through the use of test suites maintained by the federation, or by external providers.

The *Structured Marketplace* provides specific details about service interoperability in order to allow customers to select services of interest to them in a structured way, and the *Structured Marketplace* might even provide automated tools to filter eligible and interoperable services based on a customer's requirements. As in the *Open Marketplace*, the customer still contracts directly with the selected providers and is still responsible for the assembly and integration of the chosen services, but this integration should be simplified by choosing services that have already been confirmed to be interoperable. Nevertheless interoperability may not ensure successful integration, and integration may be limited to the services of one vendor or may only apply to selected combinations of services from different vendors.

The inclusive governance of a *Structured Marketplace* helps address the difficulty marketplaces can have building trust with customers in order to be successful. Inclusive governance makes it easier for members (providers) to agree on the quality and technical standards to which they will conform and the certification and testing procedures that will be used. This increases the transparency, consistency and comparability of marketplace offers and makes them more trustworthy, in turn giving customers greater confidence in selecting from the listed services. Commitment by the federation to a higher purpose (such as improving healthcare) might inspire federation members to agree to more stringent standards than otherwise, further increasing trust by customers.

---

[13] EUSD, p21-22.

[14] https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-policy-rules-and-architecture-of-standards.html

[15] http://www.helix-nebula.eu/

[16] https://www.digitalmarketplace.service.gov.uk/buyers/direct-award/g-cloud/choose-lot

[17] http://www.cloud28plus.com/

[18] https://www.deutsche-boerse.com/dbg-en/

Other benefits of a *Structured Marketplace* include:

- Contracting: Known interoperability limits vendor "lock in". This is further enhanced by providers agreeing to use standard service agreements with balanced contract terms and agreeing to reversibility/portability standards[19].
- Assembly: Known interoperability characteristics simplify the integration of services.
- Use: Validated compliance with legal and regulatory requirements improves customer's confidence in the services they have selected for use.

Clear interoperability also enables two kinds of service expansion, which in turn make the *Structured Marketplace* more attractive to customers:

- Sharing of infrastructure, such as compute, among providers to support scale-out and bursting of computationally intensive workloads.
- Expansion of service territories and service offerings. One service provider can serve a new geographic market by collaborating with another federation member in the new market. That service provider is similarly expanding the range of services it can offer its own customers through such collaboration.

Customers benefit from access to a wider range of services, while service providers benefit through increased revenues and better return on investment (ROI) on existing assets and capabilities. Such service expansion requires improved service management coordination, particularly for support functions, as well as additional governance activities to fairly and objectively manage competitive/collaborative relationships among service providers and avoid anti-trust/cartel concerns.

A common capability of a *Structured Marketplace* is federated access, allowing customers to access services across the federation using single sign on credentials from one provider[20]. This requires associated coordination of information security services by providers.

Finally, clear interoperability and standards compliance create opportunities for service providers to more broadly coordinate their service management processes, potentially reducing costs by establishing common functions such as Level 1 support services or security response teams. Customers benefit through improved service levels and simpler management of their own integrated services.

Several pre-commercial procurement projects in the research domain (Helix Nebula Science Cloud[21], Archiver[22]) have created limited instances of *Structured Marketplaces* by defining a set of functional requirements and then working with commercial providers to meet those requirements, first for a defined set of research community "buyers" and then opening up the resulting services to other research users. Both of these projects established test suites to allow providers to test their own services against the agreed standards.

Gaia-X also seems to be creating a *Structured Marketplace*, including precise "self-descriptions" of services that could address both quality and technical interoperability standards, as well as contemplating independent certification of those compliance statements.

### 2.5.3 Reseller

The federated *Reseller* extends the *Structured Marketplace* business model by enabling a

[19] E.g. the SWIPO (Switching Cloud Providers and Porting Data) Code of Conduct (www.swipo.eu)

[20] For research infrastructure, federated identity mechanisms control access to services that are mostly structured to be "free at the point of use", so federated AAI may function as a "paywall" and a proxy for usage accounting and billing services. Federations targeting commercial customers prioritize usage accounting and billing services in the same way.

[21] https://www.hnscicloud.eu/

[22] https://www.archiver-project.eu/

customer to contract directly with the federating entity, or another entity contracted by the federated entity, acting as *Reseller* for the services it has selected. The *Reseller* still does not take responsibility for integration or ongoing service delivery or management, however the *Reseller's* contractual relationships with service providers enable it to work with those providers to help resolve problems, either at the integration stage or ongoing service.

The *Reseller* business model requires coordination of billing, usage accounting and tax accounting services between the federation and participating providers, although these services may optionally be coordinated in the *Structured Marketplace* business model. It builds on the coordinated support services required in the *Structure Marketplace,* particularly federated access mechanisms.

The *Reseller* business model allows customers to access an expanded range of interoperable services that otherwise would not be available (e.g. in the country/region, in the customer's preferred language) and to implement more complex (or more geographically distributed) use cases.

The *Reseller* business model allows service providers to further expand their service territory, extending the service expansion benefit of the *Structured Marketplace.* The federating entity would need to maintain appropriate legal "residence" and tax status in different countries in order to enable this.

Federation governance supervises these activities to ensure fair and transparent treatment of all federation members. Since the federating entity itself takes a role in contracting, the governance function must be clearly independent of the federating entity to prevent self-dealing.

### 2.5.4 One Stop Shop

The *One Stop Shop* extends the *Reseller* business model with the federating entity taking responsibility for actually integrating services selected by the customer. This requires the federating entity to have the technical expertise and capacity to perform this integration, or resources to support the providers themselves integrating their services on behalf of the customer.

After integration, responsibility for service management may be split in various ways between the federating entity and actual service providers. For example, the federating entity may provide level 1 technical support, but it might direct higher level concerns to individual service providers, leaving some possibility of "finger pointing" between two or more providers in the event of incidents or problems. If problems persist, direct contracting with the federated entity might make it possible to substitute service providers.

The *One Stop Shop* requires broader coordination of service management, in turn allowing common services to be provided by the federating entity: e.g. security management and incident response, Level 1 helpdesk services, authentication and authorization, usage accounting and service monitoring, new service development and roll-out, etc. Service providers' in-house capacities/capabilities benefit from access to specialized support, expanded language coverage, and specialized data services from other member providers. A formal focus on support procedures, transparency and accountability for customer support and satisfaction, and access to a broader portfolio of experience and expertise all contribute to an enhanced customer experience.

Customers benefit from:

- Assembly: better integration of the set of services required, ability to implement complex use cases, easier/cheaper integration of desired services,
- Use: ability to support a wider range of specific use cases coming from a broader user base, improved service and support, common support channels.

Service providers benefit from:

- Assembly: easier to execute complex, multi-cloud integrations,
- Use: improved service, reduced need to hire specialized personnel, reduced need to invest in specialized service management processes, reduced costs.

Note that Gaia-X has demonstrated something like a *One Stop Shop,* piloting an automated system that combines a selection process, allowing filtering of sample service descriptions to find appropriate services for a requirement, with automated integration of the selected services with orchestration tools that respect functional and non-functional use case requirements. However, scaling this kind of approach could be challenging. A full analysis of requirements would confirm whether there is a strong need for automated, moment-to-moment provisioning of interoperable services from different vendors. For many public administration and publicly-funded customers (healthcare, research[23]), procurement practices mean that customers prefer to qualify vendors for specific services or sets of services, and then select the most-cost effective provider for a predictable amount of services on a regular (e.g. monthly) basis, rather than making those choices on a moment-to-moment basis. This avoids vendor lock-in (from both a technical and financial perspective), significantly shortens normal procurement timelines, and avoids the complexities of moment-to-moment provisioning.

### 2.5.5 Full Integrator

The *Full Integrator* business model extends the *One Stop Shop* model and is similar to that of a commercial system integrator, which takes responsibility for all aspects of technical and service integration. The *Full Integrator* works to satisfy a clearly defined target use case, integrating the required service components, testing their operation and effectiveness, and providing required documentation and training. The customer contracts with the *Full Integrator* to access and use the service, and the *Full Integrator* is responsible for managing the service and working with integrated service providers as needed (performance is monitored, support services are clearly identified and accessible and prepared to support the customer, changes to the service are managed either transparently to the customer or with appropriate notice and support, etc.). The *Full Integrator* may have the ability to substitute providers of functionally equivalent services in order to meet non-functional customer specifications such as price, performance, or security requirements. (Such substitution requires services to be interoperable, but interoperability alone does not make substitution possible.)

Benefits of the *Full Integrator* business model flow to both providers and customers:

- Customers benefit from consistent and efficient management of multiple service providers during service use,
- Providers benefit from enhanced service delivery capabilities, integrating with specialized services, resources and expertise from other providers.

## 2.6 Service Management Coordination

Coordinated service management is a critical aspect of federated operations. Well-integrated services depend on well-coordinated service management. As discussed above, the different federation business models require or enable different levels of coordinated service management. For each business model, providers agree to coordinate a number of service management processes, which in turn creates benefits for both customers and the providers themselves.

Two frameworks offer best practices for service management that can be employed in any IT service organization and that are especially valuable when services from multiple suppliers

[23] Identified by the Helix Nebula Science Cloud project.

need to be delivered, effectively, to a given customer:

- FitSM[24] and
- Service Integration and Management[25] (SIAM).

The need for this kind of service management coordination has been identified for some time -- for example by the original Helix Nebula project in 2013[26].

FitSM identifies fourteen (14) distinct service management processes that should be coordinated among multiple service providers in a multi-supply environment. Table 1 maps each federation business model to the service management processes that should be coordinated in order to deliver that model of federation[27]. The table indicates where process coordination is mandatory (M), where a minimal level of coordination is needed (m), or where coordination is optional (o).

| **Federation Business Model:** / **FitSM Process** | **Open Marketplace** | **Structured Marketplace** | **Reseller** | **One Stop Shop** | **Full Integrator** |
|---|---|---|---|---|---|
| Service Level Mgmt (SLM) | m | m | M | M | M |
| Customer Relationship Mgmt (CRM) | m | m | M | M | M |
| Information Security Mgmt (ISM) | o | o | M | M | M |
| Service Availability & Continuity Mgmt (SACM) | o | o | m | M | M |
| Continuous Service Improvement (CSI) | | | m | M | M |
| Project Mgmt (PM) | | | m | M | M |
| Supplier Mgmt (SUPPM) | | | m | M | M |
| Service Portfolio Mgmt (SPM) | | | m | M | M |
| Incident & Service Request Mgmt (ISRM) | | | m | m | M |
| Service Reporting Mgmt (SRM) | | | m | m | M |
| Capacity Mgmt (CAPM) | | | o | m | M |
| Configuration Mgmt (CONFM) | | | | m | M |
| Change Mgmt (CHM) | | | | m | M |
| Release & Deployment Mgmt (RDM) | | | | m | M |

*Table 1: Service Management Coordination Required in Different Federation Business Models*

The *Full Integrator* model for service integration requires full coordination of all of these service management processes to be coordinated, which in turn requires service providers to agree on related standards and procedures to execute each process. The *One Stop Shop* model for service integration usually implies that providers of the services being integrated have agreed

[24] https://www.fitsm.eu/
[25] https://www.scopism.com/free-downloads/
[26] http://www.helix-nebula.eu/sites/default/files/HelixNebula-NOTE-2013-002.pdf
[27] This mapping is updated from D5.3: EOSC Federated Service Management Framework. https://eoscpilot.eu/sites/default/files/eoscpilot-d5.3.pdf

to fully coordinate 8 processes, with a minimal level of coordination on the remaining 6 processes. Less-integrated business models can accommodate coordination of fewer processes.

Even for federation business models with lower levels of service integration (described in sections 2.5.1-2.5.3) formal service coordination can create benefits for federation members. For example, managing support requests that potentially involve multiple suppliers would benefit not only from coordinated "Incident and Service Request Management" (ISRM above), but also consistent Customer Relationship Management (to make it easier to share information about who requested support) as well as consistent Supplier Relationship Management (to make it easier for suppliers to talk to one another). Coordinating these three processes makes it possible for the providers involved to reduce their costs on related support functions. For example, they could establish a common support ticket management system and then consolidate front line support capability to handle simpler support requests ("level 1" requests, using the ITIL methodology) and forward more complicated requests (levels 2 and 3) to the appropriate provider.

Information Security Management (ISM) can also be coordinated to varying degrees. A common feature of many research federations is the coordination and integration of Authentication, Authorization and Identity (AAI) services, which allow customers to use access credentials established with one provider to be used to access services (as appropriate and authorized) from other providers.

Finally, Service Level Management (SLM) includes management of marketplaces, service catalogues and online service catalogues. This is a function included in all of the models above and is provided by most federations. This function is usually performed by the federating entity, and federation members only need to "comply" with this process by understanding rules for participation in the marketplace and submitting and maintaining the required information.

## 2.7 The Benefits of Federation

In general federation members optimize the utilization of their resources and increase the satisfaction of their customers, who in turn benefit from access to a wider variety and/or greater quantity of services. Federation partners achieve these results without losing autonomy and with limited or no need for financial transfers/payments to other partners.

As a validation of this analysis, the H-CLOUD project conducted a survey of European IT professionals to gather informed opinions about key benefits expected from cloud federation, as well as technical services expected to be offered by such a cloud federation. Section 2.7.1 describes the survey in more detail and analyses the results related to the benefits of federation. Section 5.2.2 below presents the results in connection with expected services.

Table 2 summarizes the specific benefits, for both providers and customers, enabled by the different federation business models listed above. Where relevant the proportion of survey respondents who expect each kind of benefit are shown in parentheses beside each benefit.

Note that the expected benefits of cloud federation captured in the survey map to all of the potential federation business models described above (except for Full Integrator). This suggests that, subject to considering the specific use cases to be addressed by federation, a more integrated business model may be needed to enable all the benefits expected by stakeholders.

| Federation Business Model | Benefits for Service Providers | Benefits for Customers |
|---|---|---|

| | | |
|---|---|---|
| *Open Marketplace* | • Increased visibility<br>• Access to a larger pool of customers, greater take-up of services. | • Increased variety in service offerings (28%)<br>• Easier/ cheaper identification, comparison and selection of different service offers in the federation |
| *Structured Marketplace* | • Reduced need to invest in peak capacity (22%)<br>• Local investments encouraged by higher ROI.<br>• Ability to serve a broader market, while limiting operating costs<br>• Ability to offer expanded services (43%)<br>• Operating costs can be reduced. (45%) | • Easier access to more/better resources (14%)<br>• Better/wider services available to customers in underserved territories or market segments (33%)<br>• Get support in the local language<br>• Greater confidence in identification, comparison and selection of interoperable service offers from the federation<br>• Reduce "vendor lock-in" (18%)<br>• Confidence knowing that standard, balanced contracts are being used.<br>• Easier to implement robust access controls (e.g. with single credential)<br>• Easier/cheaper assembly of services (25%)<br>• Confidence they are complying with all EU and relevant national laws and regulations<br>• Know that the data and information that they have placed in the federation will be handled in accordance with EU values<br>• Increased trust in federation partners. |
| *Reseller* | • Services can be resold by federation to customers outside normal service territory | • Access to expanded range of interoperable services that otherwise would not be available (e.g. in the country/region, in the customer's preferred language) (33%) |
| *One Stop Shop* | • are easier to achieve. (49%)<br>• Reduced need to hire specialized personnel<br>• Reduced need to invest in specialized service management processes<br>• Reduced costs (45%) | • Better integration of the cloud services portfolio used (32%)<br>• Ability to implement complex use cases<br>• Easier/cheaper integration of desired services. (32%)<br>• Ability to support a wider range of specific use cases coming from a broader user base.<br>• Improved service and support<br>• Common support channels |
| *Full Integrator* | • Provider's service delivery capabilities are enhanced by integrating with specialized services, resources and expertise from other providers | • Consistent and efficient management of multiple service providers |

*Table 2: Benefits Flowing from Different Federation Business Models*

Regardless of the business model, federation enables several other benefits for both customers and providers:

**Improved Security:** Federation can reduce risks from security and privacy breaches through providers adopting coordinated information security processes, and the ability to share key security functions and resources (e.g. threat intelligence services, a security operation centre, security incident response teams, vulnerability testing, training systems) gives providers access to more complete security. Even in federated structures, training needs to be consistent and complete, processes must be consistently applied, and providers must remain transparent

and accountable for their role in maintaining security. Whenever services are integrated, inside or outside a federation, the cybersecurity threat surface and access points increase, so attention to security must increase when services are integrated.

Among the IT professionals surveyed, increased security was the most selected benefit expected from cloud federation. This suggests that careful planning will be required to make sure cloud federation can fulfil these expectations.

**Improved Performance:** Federation offers providers a number of opportunities to share functions and resources with other providers, thereby reducing costs. This allows federated service providers to compete with larger cloud service providers on service, while maintaining a competitive price structure for commodity infrastructure such as compute and storage.

**Improved Innovation:** Federation providers are free to innovate and develop new services and could be more inclined to do so in the context of a more complete service management foundation. Once new services prove themselves, a federation offers several mechanisms for service expansion, easier access to the wider, federated base of customers, potentially faster adoption by virtue of the federation's community of customers, and potentially support from other providers that see how integration with the new service might allow them to expand their own business.

### 2.7.1 Results of Survey on Benefits of Cloud Federation

From 16 July to 5 August 2020, H-CLOUD conducted an online survey completed by 100 IT professionals based in one of Europe's five biggest countries[28]. Among other questions, respondents were asked whether they saw themselves as users of (40) or providers to (12) a European cloud federation. They also had the option of saying they could be both a user and provider (39), as well as neither (7). There were therefore a total of 79 responses from a user's perspective, and 51 responses from a provider's perspective (including 39 respondents that were included in both categories).

Respondents were asked to highlight the benefits and disadvantages they would expect from a European cloud federation. Figure 1 lists the expected benefits, from most selected to least selected, and includes selections from all respondents (N=100), as well as users (N=79) and providers (N=51).

In general, the expectations of users and providers were well aligned based on the ranking of expected benefits. At the same time, users had somewhat lower expectations from cloud federation than providers: on average, users selected any given benefit 14% less than providers. Users selected "increased security" 26% less than providers, but this benefit was still the most likely benefit expected.

[28] <Reference to Survey Report when available>. A total of 100 IT professionals comprised of 20 respondents from each of Spain, Italy, France, United Kingdom and Germany, completed the questionnaire. A total of 70% of respondents are employed at an organisation headquartered in the EU, which is of large (43%; 250-200 employees, up to 500 million EUR revenue or total assets) or medium size (27%; 51-250 employees, 10-50 million EUR revenue, 10-43 million EUR total assets). Most of the organisations respondents work or operate in the Information and Communication Technology (ICT) sector (23%), green-related sector (14%), or healthcare services (13%). Of those working in the ICT sector, a majority work on software (16%), cloud services (12%) and IT security services (11%).

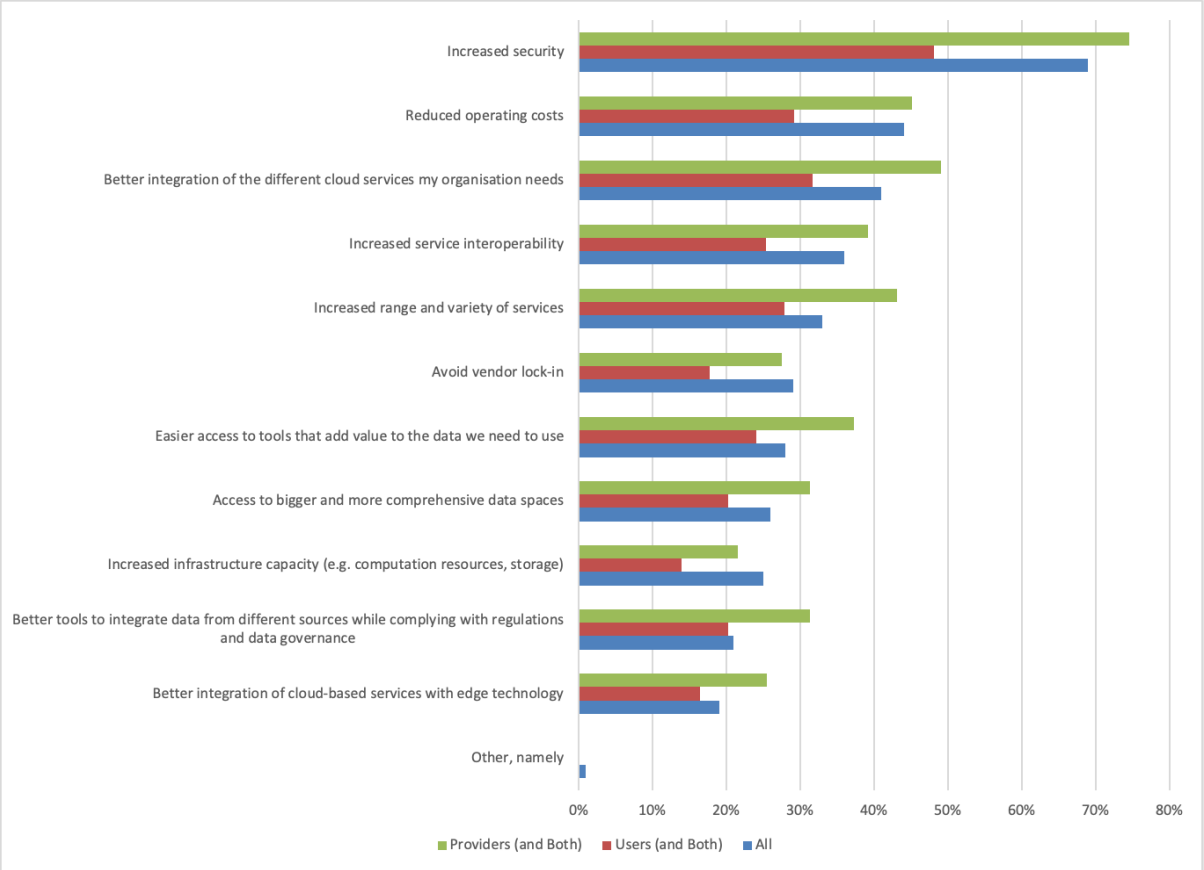


*Figure 1: Expected Benefits of Cloud Federation Reported in a Survey of IT Professionals*

Figure 2 lists the expected disadvantages of cloud federation, from most selected, to least selected, and includes selections from all respondents (N=100), as well as users (N=79) and providers (N=51).

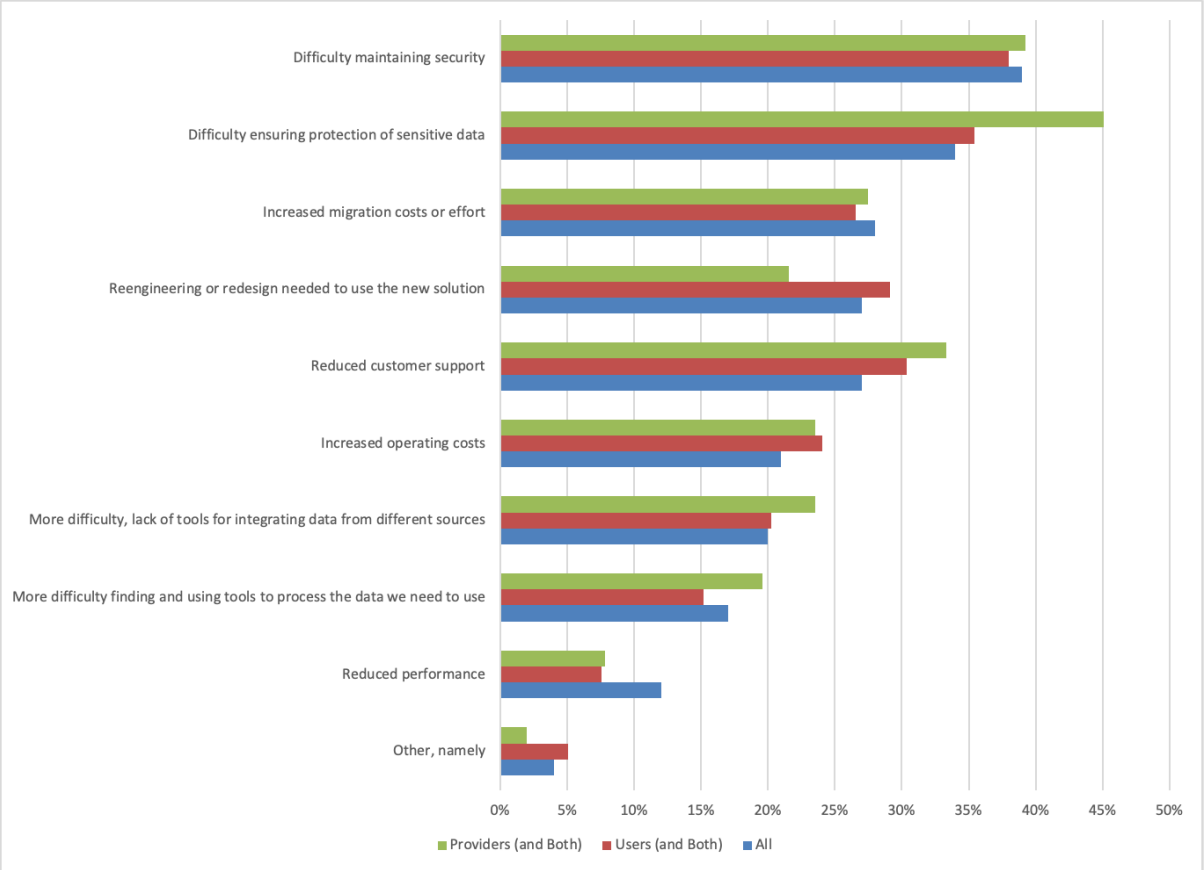


*Figure 2: Expected Disadvantages of Cloud Federation Reported in a Survey of IT Professionals*

Again, the expectations of users and providers were well aligned based on the ranking of expected disadvantages. Both groups expected specific disadvantages to the same degree.

Comparing the expected benefits and disadvantages, the survey included several "pairs" of opposing benefits and disadvantages, for example, "Increased security" as well as "Difficulty maintaining security" and "Difficulty ensuring protection of sensitive data". The H-CLOUD team wanted to identify clusters of respondents, some of whom, e.g., might expect the benefit, while others might expect the disadvantage. This was found to some extent, but cross tabulations of respondent data also exposed respondents that expected both of the opposing results:

- This was very pronounced for security, where 28% of all respondents selected both "Increased security" and "Difficulty maintaining security", and 24% of all respondents selected both "Increased security" and "Difficulty ensuring protection of sensitive data".
- Similarly, for data-related benefits and disadvantages, between 2% and 9% of all respondents selected different combinations of both benefit and disadvantage.
- Finally, 6% of all respondents expected both increased and reduced operating costs.

These data highlight the complexity of federation and its expected benefits.

## 2.8 Examples of Federation

The federated approach to cloud provisioning has been adopted by a number of research infrastructures to implement data-centric exabyte-scale computing facilities, pooling local, regional and/or national investments to implement extreme scale infrastructures offering secure access to computing resources, storage, data and applications to a widely distributed community of users (customers). Research cloud federations typically address the need of sharing access to big research data repositories by multi-national communities that

collaboratively participate in international research projects.

The EGI Federation[29] is a successful example of an international cross-border federation (delivering more than 1.1 million CPU cores and 1 Exabyte of storage across EU and non-EU countries) based on a decentralized operational model according to which management of service delivery and access policies are governed at national or regional level, with a central coordinating body responsible for defining and enforcing federation-wide policies and providing central services that enable the federation to function. The success of the EGI Federation is based on the adoption of a governance structure that clearly defines the separation of roles and responsibilities between the central coordinating body and the federation members, and on a business model which ensures the sustainability of the central services and coordination functions.

Other examples of federated cloud and federated data include:

- The Open Science Grid[30], operating primarily in the US, but also incorporating resources globally
- The Worldwide LHC Computing Grid (WLCG) is a tightly integrated global federation of compute and storage infrastructure, which uses well-coordinated service management processes to ensure the productivity of scientists working with the Large Hadron Collider (LHC) at CERN.
- the Cooperating Network of the Earth Observation Data Centre (EODC)[31]
- ELIXIR[32] in Europe
- Aristotle Cloud Federation[33] in the United States
- National scale computing capacity infrastructure, e.g. the de.NBI Cloud – the German Network for Bioinformatics Infrastructure[34], and NECTAR[35] -- the Australian research facility providing computing infrastructure, software and services that allow Australia's research community to store, access, and analyse data.
- The creation of added value through simulation and data exploitation platforms otherwise not accessible to users outside organizational and national borders (e.g. the ESA Thematic Exploitation platforms[36] and the WeNMR structural biology simulation tools[37]).

Table 3 lists 25 examples of federations worldwide, categorizing them according to their domain of activity (e.g. healthcare) as well as the federation business model that appears to be in use.

- Research federations represent the majority of the examples (14 out of 25). This is probably because the federated model is well suited to the situation where nationally funded research organisations in different countries want to support international research but also want to spend money within their own borders. The higher purpose of research federations (support for the advancement of knowledge) also makes it easier to organize a research federation.

[29] https://www.egi.eu/
[30] https://opensciencegrid.org/
[31] https://www.eodc.eu/cooperation-network/
[32] https://elixir-europe.org/
[33] https://federatedcloud.org/
[34] https://www.denbi.de/cloud
[35] https://nectar.org.au/ . NECTAR is now part of the Australian Research Data Commons (ARDC).
[36] https://eo4society.esa.int/platform-services/
[37] https://wenmr.science.uu.nl/

- The *Open Marketplace* business model also represents the majority of examples (13 out of 25, 9 of them operating in the research domain), probably because this model is the simplest to organize. Table 3 also lists 4 *Structured Marketplace* and 3 *Reseller* business models, although 2 commercial *Resellers* actually shut down operations.
- EGI and the Worldwide LHC Computing Grid (WLCG) are listed as *Full Integrator* research federations, reflecting their strongly integrated service catalogues, as well as strongly coordinated management processes.
- Twelve of the 25 examples federate compute infrastructure (IaaS) using the open source OpenStack virtualization software. Section 5 discusses technical issues in more detail, but this reflects OpenStack's broad appeal across the IT world, regardless of domain.

| Domain | Business Model | Name | Services | Common Federated Services | AAI | IaaS | PaaS | SaaS | Data as a Service (esp. API access to data) |
|---|---|---|---|---|---|---|---|---|---|
| 1. General Market | A. Open Marketplace | Cloud28plus | App Center, customers can search for solution providers with functional and regulatory requirements, SLA, certification level, security requirements, geographic location | | | OpenStack, Azure | VMware, Docker | | |
| 1. General Market | C. Reseller | CloudWatt (defunct) | | | | OpenStack | | | online data storage services |
| 1. General Market | C. Reseller | Numergy (defunct) | | | | yes | | | |
| 1. General Market | (Not an operating organization at this point, developing standards that would support at least a Structured Marketplace, potentially more integrated business models) | Gaia-X | data sovereignty, data governance, federated catalogue, compliance | Identity, Trust Management, Monitoring, Metering | yes | contemplated | contemplated | contemplated | Sovereign Data Exchange |
| 1. General Market | (not an operating organization, but provides standards that support Structured Marketplace for data) | International Data Spaces Association | certification of compliance to reference architecture | gateway IDS-compliant and certified | yes | | | | |
| 2. Public Administration | B. Structured Marketplace | UK G-Cloud | framework agreements, Digital Marketplace | Cloud Support call | | Cloud Hosting | Cloud Hosting call | Cloud Software | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | call | | call | |
| 3. Healthcare | A. Open Marketplace | ACT (i2b2 + SHRINE) | Feasibility of clinical study, platform for research patient registries | | | | | | yes |
| 3. Healthcare | A. Open Marketplace | SPHN | Network of restricted infrastructures, API services. Driver & infrastructure projects. | | yes(?) | | Containers | | |
| 4. Earth Observation | A. Open Marketplace | ESA Thematic Exploitation platforms | Access algorithms and data remotely. Move the user to the data and tools. Documentation. | usage accounting, reporting | | yes | | Application repositories or stores | Access to earth observation (EO) and non-EO data |
| 4. Earth Observation | C. Reseller | EODC | Single API, Metadata Catalogue & Explorer, Data Status | usage accounting, reporting | | OpenStack | | preconfigured VM | yes |
| 4. Earth Observation | (not an operating organization, provides interoperable software that support at least Structured Marketplace, and potentially Reseller models) | WEkEO | Support Training | usage accounting, reporting | | yes | yes | yes | yes |
| 5. Research | A. Open Marketplace | WeNMR structural biology simulation tools | Data analysis and modelling services for structural biology and life science | | | | | Data analysis Web portals | Planning to connect with EUDAT data repository |
| 5. Research | A. Open Marketplace | Cambridge Research Computing | HPC, Storage, Cloud, Training | | | OpenStack | | | yes |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5. Research | A. Open Marketplace | de.NBI | SimpleVM | | Via Elixir | OpenStack | by projects Galaxy Phenomenal | | |
| 5. Research | A. Open Marketplace | Elixir | Compute Data Resources Software Tools Support Training Governance<br>Standards activities: data-management, repositories, integration, tools and training | "Responsibility for maintaining ELIXIR-wide services is devolved to the nodes but guided by the ELIXIR Platforms and their leaders (“Ex-Cos”)"<br>Allocations on Embassy Cloud for 12mo VMs, building ability to bill for microservices | Common Identity Mgmt ("single trusted identity") to be extended multi-cloud | Working to establish multi-cloud system (presumably based on containers) | | Remote Access governed by AAI | GA4GH API-style access to standard data sources; standardized pipelines/workflows available |
| 5. Research | A. Open Marketplace | ESCAPE | DIOS Federated data lake, Extend FAIR standards, scientific analysis platforms & workflow | | | OpenStack | | | yes |
| 5. Research | A. Open Marketplace | Fenix-ri | Elastic compute, interactive computing, HPC, Virtual Machines | | yes (in development) | OpenStack | | | Federated data, archive |
| 5. Research | A. Open Marketplace | Jetstream | Jetstream: NSF HPC Cyberinfrastructure, Galaxy, Unidata, OpenMRS, SEAGrid, NAMDRunner, ChemCompute | | XSEDE AAI | OpenStack | OpenStack Magnum | | OpenStack Manila |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5. Research | A. Open Marketplace | SAIL | Data-store/lake & VDI (virtual desktop), Safe Havens with Serp Software | | | VMware, OpenStack | | | |
| 5. Research | A. Open Marketplace | SLATEci | Federation of Kubernetes clusters, cloud-nativefication, CI-CD of data-science, Edge cluster registration | usage accounting | | | Kubernetes | Jupyter Notebook (in development) | MinIO, MariaDB |
| 5. Research | B. Structured Marketplace | ARDC | | usage accounting | AAF | OpenStack by Nectar | Kubernetes | | yes |
| 5. Research | B. Structured Marketplace | Aristotle Cloud Federation | Share resources, Transparent instance movement & Allocations between institution<br>VM snapshot of complex software systems | XDMoD cloud accounting and metrics of federed resources with Cloud Support<br>DrAFTS: AWS spot instance price predictor | yes (?) | Eucalyptus than OpenStack. | Launch containers in public clouds or NSF clouds, Apache Hadoop | SLURM clusters | |
| 5. Research | B. Structured Marketplace | EOSC | Storage (FAIR and long term preservation), management, analysis and re-use of research data, Marketplace portal for users, Onboarding portal for providers, Information | monitoring, usage accounting, | yes | yes | TORQUE (optionally with MAUI), SLURM, SGE, HTCondor, Mesos, Nomad, Kubernetes | yes | yes |
| 5. Research | E. Full Integrator | EGI | Compute (Cloud, container, HTC, Workload Manager), Storage and Data (on line, archive, data transfer), Security, Training (FitSM, ISO 27001, computing and storage), Applications (on | monitoring, usage accounting, community coordination, attribute management | yes | OpenStack, Tosca | Kubernetes, Marathon, Hadoop | yes, Jupyter Notebook | yes |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | demand and notebooks) | | | | | | |
| 5. Research | E. Full Integrator | WLCG | Middleware | monitor, usage accounting, fair-share | CERN Single Sign-On + x509 auth. | OSG/HTC ondor | middleware | | yes |

*Table 3: Federation Examples*

# 3 BEST PRACTICES FOR STRUCTURE AND GOVERNANCE FROM OTHER ORGANIZATIONAL MODELS

Unfortunately, federation structures and governance models have received little attention from organizational researchers, so there is little guidance from experts on how federations should be organized. However, related organizational research offers insight into the design choices that can best support the objectives of a federation. Different bodies of research focus on three forms of organization similar to a federation:

- Multi-organization Alliances
- Social Enterprises
- Technical Alliances.

This literature review is illustrative, not systematic, but nevertheless exposes significant "better practice" guidance for federation organizational and governance choices, supported by peer-reviewed research. A more comprehensive analysis could yield additional insights and recommendations.

## 3.1 Multi-Organization Alliances

Most research on multi-organization alliances has been conducted within the last 15 years. The most notable examples of alliances are cooperatives: in 2017 there were 1,420 co-operatives across 52 countries with a turnover of more than US$100 million, and the largest 300 cooperatives had an average combined turnover of over US$3 billion each (Birchall 2017[38]).

Other examples of alliances are associations -- typically alliances of individuals, often to further the professional stature and development of their members. Many professional associations are very large in their own right: The Institute of Electrical and Electronic Engineers (IEEE) is a US non-profit corporation with over 400,000 members worldwide and annual turnover of over US$500 million. Industry associations are also significant, although there has been limited research into their structure and governance.

van den Broek and van Veenstra (2015)[39] compile four archetypical modes of governance among the different organizations within a given alliance:

- Market: governed by formal contracts, with little focus on trust. This might be thought of as the "base case" for alliances -- instrumented through specific bilateral agreements, without any overarching governance structure.
- Bazaar: focussed on collaborative activity, based on and building on reputation. Today this might be best illustrated by the "gig" economy, or by the project-oriented alliances created during film production.
- Hierarchy: marked by administrative focus and bureaucratic approach. Supply chains, centred around the dominant manufacturer, are a good example of a hierarchical alliance. While contracts are in place, the full range of relationships between each party is not governed by contracts, but instead by the administrative requirements of the dominant partner.

[38] Birchall, Johnston. (2017). The Governance of Large Co-operative Businesses. https://www.uk.coop/sites/default/files/uploads/attachments/governance-report_2017_final_web.pdf

[39] van den Broek, Tijs and van Veenstra, Anne Fleur, "Modes of Governance in Inter-Organizational Data Collaborations" (2015). ECIS. 2015 Completed Research Papers. Paper 188. ISBN 978-3-00-050284-2. http://aisel.aisnet.org/ecis2015_cr/188

- Network: relying on common goals, social contract and reciprocity. Cooperatives and associations are common networked forms of organizations.

Within the network mode of governance, additional organizational options for alliances are[40]:

- Choice of organizational form for the central entity: dictated by jurisdictional scope and its legal ability to accommodate the governance structure desired. For example, the Canada Not-for-Profit Corporations Act limits the ability of Canadian non-profit organizations to establish different classes of members or, correspondingly, different "representative" roles within the governance structure. Comparable legal restrictions are found in other organizational forms in many countries -- founders need to consider which features are desired before selecting an organizational form or even the legal jurisdiction for constitution.
- Different membership structures: including other types of stakeholder in the governance structure as formal alliance members. The research literature does not agree on whether a "single stakeholder" model is more or less effective than a "multi-stakeholder" model of governance. For cooperatives in particular, proponents of a single stakeholder governance model (Birchall 2017) highlight the importance of creating benefits for the primary members. Others (see section 3.2 below) argue that multi-stakeholder governance encourages the building of trust, knowledge and learning among the broader group of stakeholders, which in turn contribute to the success of the alliance itself.

Different governance structures: some alliances include additional governance organs and mechanisms in their structure (see Birchall 2017).

- Organizations with large numbers of "ultimate" members sometimes create an intermediate member council, to which members are elected, and which then elect a governing board or council.
- Organizations requiring geographic or constituent balance can create multiple corresponding intermediate councils (e.g. regional councils, user councils) which can then elect their own representatives to the governing board. IEEE's governance structure uses a rich set of intermediate councils to ensure representative balance among a very large membership.
- Even when intermediate councils do not formally elect representatives to the board, they can act as advisory bodies, whose advice the board should consider as part of the governance framework.

Regardless of structure, an overarching factor is the emphasis on deliberative decision-making, supporting open communication and consensus building.

- Governance competencies: Many organizations (including for profit organizations) require directors or governors to bring specific expertise and competency to board activities. Some organizations require some number or proportion of independent directors on the board or governing council, to ensure expertise is included, to encourage the use of best practices, and to avoid "control" of governance processes by dominant members.

Operational structures: Alliances employ the following operational structures or characteristics:

- Specialized resources (e.g. personnel assigned to the work of the alliance, formal secondment, creation of a separate entity) focussed on making the alliance work,

[40] These options are compiled from a number of sources, including Albers, Sascha & Wohlgezogen, Franz & Zajac, Edward. (2013). Strategic Alliance Structures: An Organization Design Perspective. Journal of Management. 42. https://doi.org/10.1177/0149206313488209, as well as specific conclusions from Birchall 2017.

- Formalized procedures (documented, standardized) to improve its effectiveness,
- Communications mechanisms that support a broad set of interfaces between and among the central entity and alliance members, typically through working groups and committees,
- Monitoring and evaluation mechanisms to track performance of the alliance in meeting its objectives.

Macdonald, et al. (2019)[41] found that alliances with robust mechanisms both for communications and for monitoring and evaluation, experience improved resiliency and organizational capacity, implicitly contributing to greater effectiveness and impact. At the same time the number of partners in the alliance, as well as their diversity, detracted somewhat from that effectiveness, suggesting that alliances need to find the right balance between stakeholder inclusiveness and relevance in order to optimize engagement and effectiveness.

## 3.2 Social Enterprises

Social enterprises (SEs) harness entrepreneurial dynamics to create public goods and serve the public or general interest. Some SEs are multi-organizational or multi-stakeholder structures.

Sacchetti et al. (2019)[42] summarize several key organizational features of SEs:

- Agreement on the "social aim" that the SE seeks to achieve.
- Inclusive and participatory approach embodied in governance and decision-making processes. This principle applies to inclusion of stakeholders in the governance process, rather than just owners, shareholders or founders.
- The "redistribution" (accumulation and reinvestment) of surplus resources. This can refer simply to building a reserve fund to ensure the sustainability of the organization, or to active collection and redistribution of surplus resources[43].
- Fulfilling non-monetary motivations of participants through the first three features: participants include both individual employees (and volunteers) as well as the organizations participating in the enterprise.

The commitment to serve a higher purpose and to create benefits for more than just the primary stakeholders of an organization (called the "cooperative pact" by Sacchetti et al.) can motivate stakeholders to subsume their direct interests to that larger purpose.

Defourny & Nyssens (2016)[44] characterize several types of SEs emerging from mutual interest associations and cooperatives (as discussed in section 3.1), as well from the public and private sectors. Types include social cooperatives and social businesses, as well as entrepreneurial nonprofits and public sector social enterprises (spin outs from the public sector).

Several studies examine the relationship of stakeholding and governance to the effectiveness of a SE in fulfilling its mission, and all identify advantages for multi-stakeholder structures:

[41] Macdonald, Adriane & Clarke, Amelia & Huang, Lei. (2019). Multi-stakeholder Partnerships for Sustainability: Designing Decision-Making Processes for Partnership Capacity. Journal of Business Ethics. 160. https://doi.org/10.1007/s10551-018-3885-3

[42] Sacchetti, S., Borzaga, C., Tortia, E. (2019) "The institutions of livelihood and social enterprise systems", Euricse Working Paper Series 109 | 19. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3519810

[43] For example, for a cloud federation, the sharing of infrastructure resources might be regarded as a type of redistribution in support of the federation's mission.

[44] Defourny, J. and Nyssens, M. (2016) "Fundamentals for an International Typology of Social Enterprise Models", ICSEM Working Papers, No. 33, Liege: The International Comparative Social Enterprise Models (ICSEM) Project. http://www.iap-socent.be/sites/default/files/Typology%20-%20Defourny%20%26%20Nyssens_0.pdf

- Borzaga & Mittone (1997)[45] conclude that organizations with multiple stakeholder groups, including the beneficiaries of the organization, participating in governance can reduce information asymmetries in their operations (increase trust). The efficiency of their operations can also approach that of comparable for-profit organizations through the commitment and stewardship of their employees, who are motivated by the higher purpose of the organization, participation of beneficiaries in governance, as well as possibly their own participation in governance.
- A survey of social enterprises with both multi-stakeholder and single-stakeholder governance (Fazzi 2012[46]) finds that multi-stakeholder organizations demonstrate superior financial performance, innovation, engagement with beneficiaries and transparency of communication, compared to single-stakeholder counterparts.
- Sacchetti & Borzaga (2017)[47] determine that involving additional stakeholder types in governance and decision-making creates specific benefits: 1) stakeholders are substantively involved through deliberation, which goes beyond the formal engagement entailed in the right to vote in organisational assemblies (which is typical of the ownership relation), as well as beyond the contractual obligation to deliver a service, and 2) shared control enables the activities of the organisation to be run cooperatively and in the public interest. Sacchetti et al. specifically refer to the additional stakeholder's participation as "membership rather than ownership".

Despite these advantages, Sepulveda et al. (2020)[48] recognize that effective engagement of multiple types of stakeholder in both stewardship and decision-making requires purposeful action, including adoption of suitable legal forms, inclusive organisational cultures, visionary leadership and concrete actions that align with the organisation's social mission: "it is neither structure nor culture, but rather a synergistic interplay of the two that matters".

## 3.3 Technical Alliances or "Platforms"

Technical alliances lead the development and operation of technological platforms or ecosystems. Compared to multi-organization alliances and social enterprises, technological platforms have been more extensively studied. Most technical alliances arise in the for-profit sphere (e.g. Google's Android ecosystem, Apple's AppStore supplier ecosystem) so research often focuses on relations between the dominant organization (e.g. Google or Apple) and its platform community or "ecosystem". The organizational, governance and operational structures of technical alliances have mostly been characterized, rather than analysed for effectiveness, much less for best practices. (Some exceptions are discussed below.)

Technological platforms can range from internal platforms, supply chain platforms to industry platforms (Gawer 2014[49]) -- with the latter being generalized into technologically based innovation ecosystems (e.g. Thomas & Autio 2020[50]). Thomas & Autio in particular note that

[45] Borzaga, Carlo & Mittone, Luigi. (1997). "The Multi-Stakeholders Versus the Nonprofit Organisation". Discussion Paper from Universita degli Studi, Trento, Italy. https://www.researchgate.net/publication/24136644_The_Multi-Stakeholders_Versus_the_Nonprofit_Organisation

[46] Fazzi, Luca. (2012). "Social Enterprises, Models of Governance and the Production of Welfare Services". Public Management Review 14. 359-376. https://doi.org/10.1080/14719037.2011.637409.

[47] Sacchetti, S. & Borzaga, C. (2017), The Foundations of the "Public" Organisation: Strategic Control and the Problem of the Costs of Exclusion, *Euricse Working Papers,* 98|17.

[48] Sepulveda, Leandro & Lyon, Fergus & Vickers, Ian. (2020). Implementing Democratic Governance and Ownership: The Interplay of Structure and Culture in Public Service Social Enterprises. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations. https://doi.org/10.1007/s11266-020-00201-0.

[49] Annabelle Gawer, "Bridging differing perspectives on technological platforms: Toward an integrative framework", Research Policy, Volume 43, Issue 7, 2014, 1239-1249, ISSN 0048-7333, https://doi.org/10.1016/j.respol.2014.03.006.

[50] Thomas, L. D. W., and E. Autio (2020), "Innovation ecosystems in management: An organizing typology", In Oxford Encyclopedia of Business and Management. Oxford University Press. Available from:

"digitalization [is] opening up new ways for companies to share ideas and knowledge and flexibly combine their outputs without contract-based coordination".

Features of technological platforms include (Tiwana, Konsynski, & Bush, 2010[51]):

- Platform architecture (including decomposition, modularity, design rules)
- Platform governance (including decision rights, control mechanisms, proprietary vs. shared).

Tiwana, Konsynski, & Bush highlight important synergies between platform architecture and the governance of that platform.

Schreieck et al. (2016)[52] extend the granularity of these two platform features:

- **Roles** cover the number of sides the platform connects, the ownership regimes, the distribution of power which can be centralized or decentralized and the relationship to stakeholders of the platform ecosystem.
- **Pricing and revenue sharing** play an important role as a governance mechanism in for-profit platform ecosystems. For example, Microsoft paid software developers to create the first apps on the Windows phone platform in order to attract more users. Later on, the developers had to generate revenues by selling their apps to the end-users or displaying advertisements.
- **Boundary resources** are tools, regulations or other resources that govern co-creation of value in platform ecosystems. Research has focused on APIs and software development kits (SDKs) but also includes "rules for participation" and standards (formal or *de facto*). Note that this parallels the "formalization" aspect of multi-organization entities.
- **Openness** refers to "the easing of restrictions on the use, development and commercialization of a technology". Specifying the use of open source software represents a specific commitment to openness.
- **Control**, in platform ecosystems, refers to how the platform owner governs the processes within the platform ecosystem and can be divided into formal and informal control mechanisms.
- **Technical design** comprises the modular architecture of the platform, the definition of its interfaces and the compatibility to relevant systems.
- **Competitive strategy** recognizes that platforms do not exist in a vacuum, and that customers have choices.
- **Trust** is relevant for the relationship between platform owner (in this context, the federating entity) and complementors (in this context, service and software providers) as well as for the relationship between customers and the platform ecosystem as a whole.

https://www.researchgate.net/publication/336603231_Innovation_ecosystems_in_management_An_organizing_typology [accessed Sep 02 2020].

[51] Tiwana, Amrit & Konsynski, Benn & Bush, Ashley. (2010). "Research Commentary —Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics". Information Systems Research. 21. 675-687. https://doi.org/10.1287/isre.1100.0323

[52] Schreieck, Maximilian & Wiesche, Manuel & Krcmar, Helmut. (2016). "Design and Governance of Platform Ecosystems – Key Concepts and Issues for Future Research". Presentation to Twenty-Fourth European Conference on Information Systems (ECIS), İstanbul, Turkey

Bianco et al. (2014)[53] further differentiate boundary resources into:

- Application Boundary Resources: technical resources that enable services or components to interact with the other services or components in the federation architecture. Application boundary resources include APIs, technical and data standards.
- Development Boundary Resources: technical resources or tools that enable the services or components to be developed and maintained within the federation architecture. Development boundary resources include software development kits (SDKs), development environments (IDEs), debuggers and test suites.
- Social Boundary Resources: resources that enable transfer the knowledge of creating software and to act as a means of interaction between the keystone player and developers. Examples include incentives, copyright and IPR policies, platform guidelines and documentation, but also training material, promotion and training events, and online community forums.

Tura et al. (2017)[54] consider the many factors that must be addressed in connection with the creation of a new technical platform that has the potential for sustainability and success. Most of these factors centre around the platform's competitive strategy (one of the items identified by Schreieck et al. (2016), above) in particular:

- Launch and diffusion, attracting participants, and ensuring access to the platform. Platforms tend to face the 'chicken-and-egg' problem, where users on each side of the platform are motivated to join only after the other side is sufficiently populated. Strategies include reaching critical user mass and launch timing of platform technologies.
- Competitiveness against incumbents. Platforms are sometimes subject to 'winner-take-all' dynamics, where it is difficult to occupy markets where dominant platforms exist. However, a unique niche for value creation may still enable new platforms to enter the markets. Designing unique competitive positioning to the platform provides a feasible starting point.
- Defining a framework for innovation, learning and growth. The former refers to choices of how the platform is renewed; who takes part in the development (e.g. the users or the platform owner) and how to ensure innovation possibilities for different sides of the market (e.g. incentivizing and enabling third party content developers' innovation). Platform growth relates to the scalability of the platform. Some platforms are naturally more scalable, but design choices can nevertheless influence growth. These include openness of the platform and thus relationships to other extant platforms.

These factors also play into ongoing governance of the platform, as examined by Schreieck et al. (2018)[55].

Building on these "ingredients" of a technical platform (such as a federated IT service structure), Hermes et al., 2020[56] analysed recent literature to determine "how digital platform

[53] V. D. Bianco, V. Myllärniemi, M. Komssi and M. Raatikainen, "The Role of Platform Boundary Resources in Software Ecosystems: A Case Study," *2014 IEEE/IFIP Conference on Software Architecture*, Sydney, NSW, 2014, pp. 11-20, https://doi.org/10.1109/WICSA.2014.41.

[54] Nina Tura, Antero Kutvonen & Paavo Ritala (2018) "Platform design framework: conceptualisation and application", Technology Analysis & Strategic Management, 30:8, 881-894, https://doi.org/10.1080/09537325.2017.1390220

[55] Schreieck, Maximilian & Hein, Andreas & Wiesche, Manuel & Krcmar, Helmut. (2018). "The Challenge of Governing Digital Platform Ecosystems" in C. Linnhoff-Popien et al. (eds.), *Digital Marketplaces Unleashed*, Springer-Verlag, 2017. https://doi.org/10.1007/978-3-662-49275-8_47.

[56] Hermes, Sebastian & Pfab, Simon & Hein, Andreas & Weking, Jörg & Böhm, Markus & Krcmar, Helmut. (2020). "Digital Platforms and Market Dominance: Insights from a Systematic Literature Review and Avenues for

owners attain market dominance". The analysis explores how possible platform strategies interact with relevant environmental contexts to influence potential market dominance. As one example, the analysis highlights the trade-off between: 1) high modularity of a platform architecture, increasing a provider's incentive to innovate and better enabling coping with rapid technological change and/or increasing demand for open standards; and 2) low platform modularity, reducing competition and preventing platform imitation. The analysis does not offer definitive guidance on best practices, since they depend strongly on context, but it does offer a filter for assessing future success of platforms such as are contemplated by the EUSD. In particular the authors speculate on methods to "dethrone" dominant platforms through the use of industry consortia[57] (such as Gaia-X) or new forms of platform cooperatives (such as Stocksy[58], Partago[59], Fairmondo[60]) which implement technical platforms with cooperative organizational forms.

Future Research", Presentation at Twenty-Fourth Pacific Asia Conference on Information Systems, Dubai, UAE, 2020.

[57] Hermes, Sebastian & Töller, Nadja & Hein, Andreas & Weking, Jörg. (2020). "Gaining Control over Critical Platforms: A Comparative Case Study of European Consortia". Presentation at 28th European Conference on Information Systems, Marrakesh, Morocco. The authors examine the European Processor Initiative (EPI) and efforts related to the European Payment Services Directive and intend to examine the Gaia-X initiative.

[58] www.stocksy.com

[59] www.partago.be

[60] www.fairmondo.de

# 4 PROCESSES FOR CREATING A SUSTAINABLE MODEL FOR A EUROPEAN CLOUD FEDERATION (EUCF)

As discussed in Section 2.2, an essential characteristic of federation is that federation members cooperate and coordinate their activities. Even before the federation is created, this requires deliberation among affected stakeholders, considering questions of mission and vision, objectives, governance processes, governance stakeholders, development of business cases, validation of business cases, scope of and form of business model, the activities to be performed, and assignment of roles and responsibilities across stakeholders. This section details the steps in this deliberative process and the questions that must be addressed at each stage, in order to arrive at a collective agreement about the objectives, scope and form of desired organization.

## 4.1 Alignment of Mission and Vision

Developing consensus on objectives should be informed by identification of the key requirements and use cases that the federation intends to support. The adoption of a large-scale federation in the commercial sector requires alignment of business objectives so that all federation members can achieve innovation, development and competitiveness (see Section 2.5 for specifics on this issue).

- For example, the European Open Science Cloud (EOSC) has engaged in an extensive analysis of the requirements and use cases that EOSC's stakeholders need to address. This analysis found a high level of overlap across these requirements, which has been translated into clear functional objectives for EOSC, including five main types of services for European researchers[61]. (These services are listed in section 5.2 below, and they focus on support for the sharing of research data.)
- Gaia-X stakeholders have so far submitted roughly 40 use cases, covering a wide range of sectors, involving many different stakeholders, and probably requiring very different kinds of services and solutions. It will be important to distil these requirements into clear objectives that will guide Gaia-X's next stage of development.

**Consider a Public Good Objective:** Section 2.2 above highlights how "commitment to a public good objective" can increase customers' trust in federation partners and improve its effectiveness. Sections 3.1-3.2 also describe how the beneficiaries of multi-organizational "alliances" as well as social enterprises are often included as stakeholders, potential partners, and possibly participants in the governance of such organizations.

## 4.2 Governance Model and Stakeholders

Another key step is defining what organizations should be members of the federation, the possible creation of a separate federating entity, how key stakeholders can have a voice in the governance of the federation, and the roles and responsibilities of the federation members. This analysis is often linked to the proposed objectives, depending on the circumstances:

- Federated research clouds typically have research performing organizations as the partners, often operating in both provider and customer roles, dedicated to the public good objective of advancing science in a given domain, as well as supporting specific research communities. Researchers are the ultimate "users" of research clouds, and while they may not have a partnership role, their voice is carried by their research organization, as well as through dedicated advisory boards.

[61] European Open Science Cloud (EOSC) strategic implementation plan., p. 23

- Existing industry associations are dedicated to helping their members' businesses succeed. For example, CISPE, an association of EU cloud providers, has developed a framework to comply to EU regulations (the CISPE Cloud Code of Conduct -- see section 4.1 of the H-CLOUD Green Paper) and could explore other standardization efforts, certification mechanisms, marketing efforts, as well as establishing shared services that could expand the range of services and coverage available to customers. However, such associations are clearly focussed on their commercial interests.
- Industrial clouds and B2B platforms have been created by diverse sets of organizations with a diverse set of common objectives -- which mostly benefit the ecosystem participants themselves and sometimes primarily the dominant player in the ecosystem (for example Google in the Google Android ecosystem)[62].
- Data alliances, such as health data hubs and humanitarian data sharing, typically involve a broader range of stakeholders, as well as dominant players such as national health authorities or international humanitarian organizations. The objectives often involve the public good, but they are also sometimes less specific, since the objectives frequently include exploration of the data in hopes of discovering new insights. Including representatives of key "beneficiary" stakeholders (e.g. patients for health data hubs, refugee representatives for certain humanitarian data sharing initiatives) in governance processes helps these organizations stay focused on their objectives.

Sections 4.1 and 4.3 in the H-CLOUD Green Paper highlight the role, not only of cloud service providers, but also systems integrators, consulting firms, cloud brokerage firms, standards organizations, certification and audit firms, the open source software community and software vendors as potential stakeholders in supporting the objectives of the proposed EUCF. Section 2 above touches on many of these activities (such as integration, certification and advisory), so these kinds of organizations might play a useful role in the EUCF. Similarly, the ultimate beneficiaries of a EUCF would be current and potential cloud customers, including public sector organizations, as well as small and medium sized enterprises (SMEs).

Coordinated/federated approaches must be structured around the objectives of their stakeholders, balancing community focussed initiatives with pan-European solutions. The European landscape for cloud- and data-driven innovation is complex and fragmented, with many potential use cases, customers, providers, innovators and stakeholders. The EUSD itself addresses a range of requirements and opportunities across nine sectors of the economy. The needs of different stakeholder communities must be balanced against the need for common or aligned solutions. The effectiveness of a EUCF will depend strongly on the clarity of its value proposition and how it is constituted to realize that value proposition.

## 4.3 Business Case Development

Detailed business cases should be developed for identified use cases in each of the nine sectoral data spaces described in the EUSD. These business cases should quantify the societal gains and costs associated with achieving the desired benefits and should determine feasibility and related ICT innovation needs. These business cases would identify existing initiatives (e.g. health data hubs) and high-impact use cases, elaborate specific data sharing use cases building on existing good practices, identify data and cloud resources that might be good candidates for sharing and re-use, and quantify specific gains and costs for businesses, research organizations and Public Administration to use federated cloud and data services as a platform for cross-sector data sharing involving private data, public data and governmental data. Ideally this would follow a process similar to that described by the High-Level Expert Group on Business-to-Government Data Sharing[63] (page 48) which starts by identifying the problem to be solved, the conditions for data re-use, possible compensation structures, and

[62] <Link to come: Policy recommendations to further develop B2B Industrial Digital Platforms in Europe>

[63] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64954, p. 48

then considers the optimum model for data access. Each business case should include a requirement and gap analysis to identify services needed from underlying technology platforms, as well as the stakeholders that would need to be involved in their delivery and supervision. This analysis may identify clusters of requirements where solutions at the correct technological readiness level are available and could be deployed, for example to ensure secure and interoperable access to data and cloud services. This analysis will also identify gaps in existing solutions that should be prioritized for additional research and/or development.

Tura et al. 2018 [54] offer a framework of questions that should be addressed in the recommended business cases in order to identify promising roadmaps for the development of initiatives in each sector:

- **Platform** architecture: Describes the structure of the actors, market structure and fundamental setup of the platform
  - Core interaction: What is the main purpose of the platform? What is the (core) interaction that takes place in the platform? What are the value creating assets being shared or transacted?
  - Market structure: Which markets are involved in the platform (two-sided, multi-sided markets)?
  - Key actors: Who are the actors representing different market structures and providing main functions?
  - Platform openness: How open is the platform and what is the strategy to manage openness? How openly, or under what terms, is the data and information shared?
- **Value creation logic:** Defines the benefits of the platform and how each user contributes to value creation i.e. describes value functions for platform users.
  - Actor roles: Who benefits from the platform and how? What are the roles of the stakeholders and how will they change? How to achieve commitment of the stakeholders?
  - Value proposition: What are the different value propositions for different participants?
  - Network effects: How do network effects work? Who is needed to enable network effects?
  - Revenue model: What is the price for participating for each stakeholder?
  - Governance: Describes the common rules, laws, practices and managerial level of the platform.
  - Leadership: Who manages the platform and how?
  - Ownership: Who owns the platform?
  - Platform rules: What types of rules are enforced? How are the services/content regulated? What consumers, producers, providers, and competitors are allowed to do?
- **Platform competition:** Describes the setup for launching the platform and competitive selections.
  - Launch: How to ensure access to the platform? How to attract users from different sides of the platform? How to solve-the chicken-and-egg problem?
  - Competitiveness: How does the platform compete against -- or enhance -- incumbent solutions? What is the unique competitive advantage of the platform?
  - Innovation & learning: Are platform innovations needed and how are innovation targets set and approved? Who is involved with platform development and how?

- Platform growth: How will the platform grow? How big and scalable is the platform intended to be?

For each business case, a federation business model should be selected that best fulfils the community's requirements while providing the best value with the least effort. Different models may be appropriate for different business cases. Consideration should be given to whether the proposed action is a response to a market failure, which can be corrected with a temporary action, or whether continued support is needed to correct a systemic problem.

## 4.4 Validating the Business Cases

Before establishing a formal EUCF, there should be investments to encourage participation in federated cloud and data sharing pilots and create a dedicated virtual support and training centre. Creating open infrastructure and testing capabilities could flexibly support demonstrations, proofs of concept and pilots of how federated cloud and federated data solutions could be assembled, operated, managed and governed, including collection of data, which would validate the business cases proposed above. This would bring together a number of customer organizations (e.g. public administrations, industries and research organizations with a specific data sharing use case) to define requirements. These requirements would then be addressed by providers integrating existing tools (i.e. no research and minimal software development), and solutions would be supervised using best-practice federated governance models. As skills are a major asset for a successful repurposing and re-architecting of applications by potential cloud customers, a support centre will be necessary to provide the technical expertise needed for an effective use of the pilot infrastructure.

Particular attention should be paid to assessing provider costs and value received, in order to identify sustainable business cases for continuing operation. Based on the analysis presented in the H-CLOUD Green Paper, promising use cases can be found in each of the public administration, transport/mobility and health care sectors, for which sectoral data spaces are all proposed in the EUSD.

## 4.5 Determining the Right Scope for Each Initiative

Building on these pilots and demonstrations, multiple initiatives can be created around domain-specific use cases, objectives and beneficiaries, partners and stakeholders, governance and decision-making mechanisms, scope of possible federated activities, and applicable business models. Supporting multiple initiatives reflects the multiplicity of cloud federation situations, with different objectives, where different sets of existing players should play a role:

- Despite generally low rates of cloud adoption across the EU, there is wide variation in those adoption rates across different countries and sectors. This indicates that a EUCF should be structured to accommodate different solutions for different countries and sectors.
- In some sectors (notably public administration and healthcare), there are structural concerns about adoption linked to the EU's own data protection policies. The EU has established a high hurdle to protect European data -- and European rights and values. A EUCF could now provide stakeholders with the tools they need to clear this hurdle.
- Many sectors already have vibrant ecosystems of large and small players trying to build platforms, networks and data-driven businesses (e.g. mobility, agriculture, manufacturing). While these share common underpinning technological requirements, they do not share common communities, history or even terminology, so a common technical platform is not the primary requirement.
- For SMEs across the EU as well as publicly funded sectors (public administration and health care in particular), internal resource constraints are limiting adoption, so

initiatives responding to their needs may offer a different set of supports than might be appropriate for specific sectors.

- Digital Innovation Hubs[64] have a growing role as public-private partnerships dedicated to stimulating innovation. Even here, fragmentation is occurring across geographic regions, as well as within sectors. The DIHnet project[65] is the first attempt at implementing a coordination function between DIHs but will soon end with other initiatives looking at federating DIHs (i.e. in Big Data starting with the BDVA iSpaces[66]). The EOSC DIH[67] could be an opportunity but is becoming more inclusive of other types of services than just cloud. In addition, it is still difficult to be able to fully extract and understand the DIHs already working with cloud technologies and those that might require them[68].

Clear objectives, agreed to by all partners, enable cooperation and coordinated action in support of those objectives, sometimes at the expense of individual partners' priorities. Participating in the standard-setting process and then complying with the resulting standards is an example where partners trade their own flexibility and freedom of action for the benefits of coordination/federation. Higher goals, such as improved public health or the advancement of science, can motivate higher levels of cooperation, such as sharing of data or excess computational capacity.

## 4.6 Value of the EUCF as Umbrella Coordinator

Given the possibility of multiple federated cloud initiatives, focussing on specific sectors or related use cases, it would be valuable to establish the EUCF as a lightweight coordinating initiative, identifying cross-sector research and innovation activities, facilitating collaboration on interoperability and adoption of best practices, and providing relevant shared services such as certification and testing activities.

The EUCF should be positioned as a higher-level collaboration of federated initiatives in multiple domains -- a federation of federations -- encouraging coordination, demand aggregation, service and technology synergies, best practices, evaluation of new standards and technologies, codes of conduct, certification, security, research and innovation activities. The EUCF could support activities such as:

- Technical initiatives to align EU cloud providers around standards, security, data protection, certification, etc. and then to assist them in improving their capabilities in each area. (This work does not start with a blank page -- there are useful technical architectures and standards to consider -- these are discussed in section 5 below.)
- Creating a customer-focussed community, giving customer organizations a well-defined voice in guiding the work of a EUCF and prioritizing actions with the greatest potential. A user community could share best practices and success stories, increase awareness of what is possible and build trust in the EUCF and in these technologies in general.
- Creating formal roles for the consultants, integrators and other intermediary organizations whose business it is to help organizations adopt these new technologies. These same organizations can work to build capacity for cloud adoption in key sectors -- e.g., assisting SMEs to create effective security processes, helping healthcare providers safely share data to improve patient care and public health, etc.

[64] https://s3platform.jrc.ec.europa.eu/digital-innovation-hubs
[65] https://dihnet.eu/
[66] http://www.bdva.eu/I-Spaces
[67] https://eosc-dih.eu/
[68] https://s3platform.jrc.ec.europa.eu/digital-innovation-hubs-tool

- Support for formal testing, evaluation, certification and compliance functions, facilitating innovation by smaller suppliers, rather than their being held back by lack of resources. This activity aligns with the test and evaluation initiatives being considered in the proposed Digital Europe Programme.

Supporting translation of better practice supports from one region or sector to other areas.

## 4.7 Phased Implementation Approach

The EUCF should be implemented with a phased approach that flexibly aligns activities across multiple domains and that allows achievement of "quick wins". The discussion above highlights the many topics that must be addressed before a EUCF can be organized and functional. Pilot projects and demonstrators will clarify requirements and identify applications and use cases where an early version of the EUCF can achieve success, which in turn will build credibility and support.

The following aspects of implementation should be addressed:

- What mechanisms are in place to build the two sides of the market contemplated by EUCF? The implementation plan should make sure initial providers and customers have success with a EUCF.
- What mechanisms will build trust and confidence in the EUCF? How can customers have confidence in the "marketplace" offers presented by providers?
- How will different communities and sectors be prioritized by the EUCF. As noted in the Green Paper, there are clear challenges for SMEs, public administration and the healthcare sectors.
- How will the EUCF align with the procurement structures of both targeted customers and initial providers? What mechanisms can be used to encourage use of the EUCF's services? What role will vouchers or virtual access mechanisms play?

A similar phased approach was employed in the creation of European Open Science Cloud[69], as well as the early phases of the creation of the International Data Spaces Association (IDSA)[70]. Both examples provide insights into the deployment process for the EUCF.

## 4.8 Operating Structures

Federation requires technical, operational and organizational integration, and often requires the creation of a central organization to manage collaborative activities and perform common functions. (The primary exception is in most industrial clouds and B2B platforms, where these common functions are usually performed by a dominant industrial partner, such as a major manufacturer at the centre of its supply chain.)

Sections 4.1-4.7 above provide a roadmap for defining and establishing an effective multi-organizational alliance such as the EUCF. While there has been limited research into structural options, much less best practices for organizing and governing these activities effectively, Section 3 identifies several preferred operating structures:

- Form a central organization focussed on the coordination efforts required to make a EUCF work,
- Establish and follow well-documented procedures (for example, governance protocols) to improve its effectiveness,

[69] European Open Science Cloud (EOSC) strategic implementation plan

[70] Otto, B., Jarke, M. Designing a multi-sided data platform: findings from the International Data Spaces case. Electron Markets 29, 561–580 (2019). https://doi.org/10.1007/s12525-019-00362-x

- Establish communications mechanisms that support a broad set of interfaces among the EUCF, its primary and affiliated members, typically through working groups and committees,
- Establish monitoring and evaluation mechanisms to track EUCF performance against its objectives,
- Identify funding mechanisms and a business plan to ensure sustainability of the EUCF's activities and particularly its central coordinating body.
    - A EUCF would provide services to both for-profit cloud service providers and customers and should be able to develop a value proposition that would ensure sustainability without continuing funding support.
    - However, in the near term, as a EUCF is established and works to "upskill" and "upcertify" both customers and providers, these activities will need to be supported until those customers and providers themselves achieve a level of self-sufficiency and sustainability in their own cloud journeys.

The clear intent of the EUCF is to enable coordinated provision of multi-supplier services to EU stakeholders, so effective service management is a critical success factor for the EUCF. Proven methods for effective service management can be found in the FitSM[71] and Service Integration and Management[72] (SIAM) frameworks.

This critical dimension is not addressed by either the Gaia-X or NIST Cloud Federation Reference Architecture (CFRA). Gaia-X's current documentation does not describe how a Gaia-X compliant federated cloud capability might be configured, operated or be supported by various providers. Gaia-X is intended to support service integration from multiple vendors, and this is described technically but not operationally. Similarly, the NIST CFRA outlines the need to manage the services being integrated but does not provide guidance on the tools or approaches needed to do this.

Existing best practices and standards (e.g. ITIL, FitSM, etc.) for federated service management should be refined and evolved to ensure federated cloud initiatives have reference requirements, processes, procedures and policies that ensure the compatibility of service delivery and planning across different initiatives. "Starter kits" should be developed to assist with implementation of each federated business model, with sample templates for required governance and service management processes (definitions, roles, process maps, etc.).

[71] https://www.fitsm.eu/
[72] https://www.scopism.com/free-downloads/

# 5 FEDERATED CLOUD ARCHITECTURE AND SERVICES

Technically, a cloud federation requires the management of a common set of policies and procedures, shared, scalable and secure cloud-based solutions for data and infrastructure access across the federation, and a layer of federation services to manage resource allocation and other central processes. The federating entity governs the reference standards, policies and requirements that allow infrastructure, data and services to work together, shared and possibly pooled for use by federation customers.

## 5.1 The Role of a Federation Architecture

Tiwana, Konsynski, & Bush, 2010 (referenced above) define “platform architecture” as a conceptual blueprint that describes how a platform ecosystem is partitioned into a relatively stable platform and a complementary set of modules that are encouraged to vary, and the design rules binding on both. For a cloud federation, the cloud federation architecture is a framework for different components that can effectively work together, be shared or pooled. Those underlying components might be characterized in several ways:

- Infrastructure components: the computational, storage, networking and other assets that can be accessed by federation customers.
- Data: the data objects (files, database records, etc.) that can be accessed.
- Information services: discrete information services that perform functions on input data and return output data, possibly updating stored data as part of the process.

These are not mutually exclusive approaches but rather different ways of describing an integrated information system. For example, a specific storage asset may contain data of interest to the customer, so authorizing access to that storage effectively allows access to the data itself. Similarly, particular data might be encapsulated in a “data service” to which a customer submits a request, along with authorization information, and receives the data, or some version of that data, that the customer is authorized to receive.

Discussions of how to federate both infrastructure and data often reflect the “state of the art”. For example, sharing of compute infrastructure has involved different approaches over time:

- Submission of compute job requests in “batches” to a scheduler, which reserves physical servers, networking and storage resources to each job, configures the required software on those servers, launches and monitors the program specified by the customer, and then releases the physical resources for use by the next job(s). This approach is still used today by most high performance computing (HPC) systems, as well as by major distributed computational resources such as the Worldwide LHC Grid Consortium (WLCG) and the Open Science Grid (OSG).
- Distribution of a package of software as an “image” to be executed by virtual machines (VMs) managed by virtualization systems like OpenStack and VMware. The software might run for a while and then finish, after which the virtualization system would assign new work to the virtual machines, or it might run continuously in order to process requests being sent to it by authorized customers (or by other software components). As described in section 2.8, federations based on OpenStack compute resources are very common today.
- Distribution of a package of software in a standardized software “container” (such as Docker) that can be executed by many physical servers. Sharing the basic operating system (usually Linux) of the host server makes this approach more efficient than virtual machines and simplifies updates by replacing the container image (or its parts). The availability of this method triggered a large increase in the use of containers, even within a single application, which in turn required development and adoption of

orchestrators (such as Kubernetes). This reflects the current “state of the art”, and several existing OpenStack federations are considering moving to this way of sharing compute infrastructure.

- The software itself is restructured so that common operations are performed by dedicated “function as a service” services, therefore completely eliminating the need to manage server infrastructure. This technique is gaining in popularity but is still considered an advanced approach.

Similarly, sharing of data has evolved over time:

- Customers access remote data by “logging in” to remote computers that hold this data and accessing the data directly on the remote system.
- Customers access remote data with data-specific local software that manages authorization, networking and interface with the remote data source.
- Data-specific software is replaced by more generic remote access software that uses a flexible data framework to access different data sources.
- Remote access software is replaced by an application programming interface (API) to allow the customer’s software to access the remote data by making program calls to the API.
- A programmable API is replaced by a web-based interface (typically a RESTful API). APIs can be used to transfer data to the customer, or they can be used to trigger remote calculations and analyses on the remote data, avoiding the need to transfer or expose sensitive data[73].

The “state of the art” varies from community to community, so it is important to define a federated architecture in a flexible way, addressing common functional requirements, without being limited by the evolution of solutions to those requirements. At the same time, the evolution of “shared compute” and “shared data” described above are both converging toward “compute as a service” and “data as a service”, suggesting that a service-oriented architecture is a robust approach.

A service-oriented architecture also accommodates the need to define services that are visible to and accessible by customers, as well as services that are internal to the federation and help federation members work within the federation operate more effectively.

In addition to defining platform architecture and its role, Tiwana, Konsynski, & Bush, 2010 (referenced above) highlight the important role played by architecture in framing governance of standards, technology and compliance.

Creating a distributed yet federated, technically effective data-processing system is an active subject of research -- many technical approaches are being studied, and many technical approaches are in use in the community, but they are not converging into “standards” because the underlying technologies are rapidly evolving and because the scope of integration is expanding from the data centre out to the more heterogeneous edge computing environment. Where distributed capabilities need to work together, there must nevertheless be an agreement on the framework of the system (the architecture) and the standards to be adopted within that framework, even in the face of this rapid change. Establishing a platform architecture enables technical discussions to be modularized and compartmentalized and facilitates agreement on standards that enable specific services to interoperate.

At the same time, a platform architecture is not permanent -- cloud technologies in particular are evolving rapidly, and the federation’s technical experts should regularly review both the

[73] This technique is used by both the International Data Spaces Association (IDSA) and the Global Alliance for Genomics and Health (GA4GH). Mechanisms are still needed to certify the remote calculations/analyses to make sure they do not inappropriately expose sensitive data.

architectural framework and incorporated technical standards and solutions and propose updates or changes as needed, at least for a defined period and/or in the context of serving a particular market (e.g. research or public administration).

Architectural choices should only be made in the context of the “business” role and “purpose” of the federated cloud. For example, EGI and EOSC prioritize service management (i.e. the “use case” or “business case”) over technical standardization and accommodate different technical approaches, as needed, in favour of the objectives of their users and clients according to their use case requirements.

As recommended in Section 4, specific business cases should be analysed to determine specific technical requirements and the timeframes in which those requirements apply. Specific services can be evaluated against those requirements to build an effective solution that will be fit-for-purpose for the planned lifetime of the business case.

A consequence of this recommendation is that it is likely that different services, different technologies, and possibly multiple standards will be needed to meet the requirements associated with use cases that are priorities for any federated cloud initiative, particularly one as broad as the EUCF. While different tools and processes might be used in different use cases, a consistent architecture can and should be created to organize those tools and processes.

## 5.2 Federation Services

This section presents a master list of possible federation services, as well as the results of a survey of cloud computing specialists regarding services they would like to see in a possible European cloud federation.

### 5.2.1 Master List of Possible Services

Services are listed primarily according to their visibility and availability to different groups.

#### 5.2.1.1 Support Services Available to Customers, Providers and Externally

- Marketing, Outreach, Advisory, Education & Training, New user orientation
- Support Services, including helpdesk, competence centres, knowledge bases
- Consulting, Extended implementation services, co-development of new service

#### 5.2.1.2 Technical Services Available to Customers

- **Infrastructure-based services,** including:
  - Computing, including IaaS, PaaS, Containers, Serverless, Batch Computing, SaaS/Marketplace and spot pricing/preemptible computing.
  - High-bandwidth network, provided in the form of Virtual Networks, Load Balancing services, DNS services and Content Delivery Network
  - Storage coupled to the local computing resources or remotely accessible through a secure data layer. Storage can be provided in form of block, object and file storage.
- **Resource Management Services**: allowing infrastructure-based services to be controlled and managed. Tools should be as technology agnostic as possible, but may also need to be unique for each federation. Examples include:
  - **Resource schedulers** and reservation systems

- **Deployment** and **provisioning** tools
- **Logging** and **monitoring*** functions for both hardware and services

- **Customer Interfaces: Console & APIs**. To support the usage of the federation service and infrastructure a set of components can be defined, it is quite common to the provide them through Direct APIs, which can be easily integrated into external software, and often on top of that API as command line interface (CLI) made available for the major platforms (Linux, Windows, MacOS), in the same way Web Applications or OS specific graphical clients can be provided.
- **Advanced and Value-Added Technology Services Available to Customers.** New services are continually being developed, and the federation should adopt, or implement, services that help customers meet their objectives. Examples include:
  - **Community-specific tools** (such as scientific applications, data acquisition and processing tools, shared notebook services such as Jupyter)
  - **Database Services**, for example relational DB, in-memory, SQL - NoSQL, and time series.
  - **Big Data & Analytics,** tools for processing large and varied datasets, including batch and stream processing, stream data ingest, analytics, data extract, transform and load (ETL). This category would also include data analysis tools optimized for particular kinds of data, such as Ophidia[74].
  - **Workflow Management and Orchestration**
  - **Machine Learning**, including widely used AI models for audio, voice, image and video recognition, text and translation, as well as general purpose algorithms (both training and inference).
  - **IoT - Edge**, enabling registration of, secure connection with, and control and configuration of a variety of devices located remotely (at the “edge”). Services could include data quality and consistency, device management and analytics services. Edge-based computing is evolving rapidly, so support for multiple standards, formats and regimes, such as provided by the OpenFog consortium, will probably be needed.
- **Secure processing of distributed data.** An important class of value-added services, many of which are still in the development stage:
  - One example of this is federated machine learning, which allows a common machine learning framework to be set up in multiple locations (data centres, jurisdictions, etc., with choices to be managed by the customer), which can then process locally-stored data and train a common machine learning model (for example, training an image recognition model on sensitive health images, without having to bring the sensitive data to one location)[75].
  - Other privacy-preserving technologies include homomorphic encryption, secure multi-party computation, differential privacy, distributed ledgers (blockchain), and automated risk management and compliance tools. See section 7.2 for further discussion.

#### 5.2.1.3 Federated Services Available to Customers

- **Federated authentication**: A single entry point that enables the authentication of an user based on a trusted identity, the trust is delegated to the organization providing the

[74] http://ophidia.cmcc.it/
[75] See discussion in section 7.2.2.

identity and which has to be part of the federation or to be recognized as legitimate Identity provider.

- **Federation dashboard**: The central place to visualise and manage the resources available for a given role. When the federation includes multiple service providers for the same type of service, all these resources should be seen presented in a unified view.
- **Basic Data Transfer*** between the different service providers
- **Service Catalogue** or **Marketplace*** where the user can search for approved federation services based on her/his requirements, select the ones of interest, and arrange for access and use.
- **Usage Accounting***. Track a customer's usage of different resources by different users
- **Financial Management and Billing***. Arrange for financial reconciliation among providers and customers, track use of payment mechanisms like vouchers and virtual access, and manage Billing* and collection from customers for services provided*.

* Service elements defined in the NIST Cloud Federation Reference Architecture. Many service elements required by a cloud federation are not addressed by the NIST CFRA.

#### 5.2.1.4 Federated Services Available to Service Providers and Integrators

- Administrative & Management tools
- Shared Identity services
- Centralized reporting and accounting
- Internal communication
- Collective bargaining, Internal mediation
- Promotion
- Validation
- Facilitate technical integration
- Authorization, Identity Management*
- Configuration Management Database
- Service Management Activities (14 processes from FitSM)
- Service Management Collaboration Tools
- Documentation services
- Operations Portal
- Membership Management*
- Policy management* (examples from EGI/EOSC listed)
  - Compliance Framework
  - Rules of Participation (membership criteria and requirements)
  - Service Management Procedures (14 processes, see Section 2.6)
  - Service Portfolio Management Tool

  - Interoperability Guidelines (service specific)
  - Security policies, federated incident response processes
  - Certification of data resources
  - Minimum requirements for catalogues
  - Policy monitors
- Portability & Interoperability*
- Auditor*, auditing internal processes, usage accounting, as well as accuracy of marketplace listings.
- Security and Security Coordination*, including federated security response teams, security operations centres, threat intelligence capabilities, software vulnerability testing, etc.
- Management of compliance testing and certification, including creation and maintenance of the federation's own test suites, qualification of external certification services, management of certification processes, advisory services to support providers in their own certification efforts (such as EU Cloud CoC, GDPR, ISO 27000 cybersecurity, energy efficiency).
- Data Interoperability Services, resources such as test suites to support use of data standards, including industry specific data exchange/interoperability standards, such as HL7 for healthcare and MDS for micromobility.
- Integration services, either supervising efforts of affected providers, managing internal DevOps teams to achieve service integrations, or managing external system integrators.
- Facilitation of new service development, evaluating market requirements, working with relevant service providers, guiding development, testing, release and support processes.
- Technology assessment services, scanning new technology developments, as well as emerging standards, in the marketplace, guiding strategies for evaluation, piloting, demonstration and wider adoption of each new technology.
- Advisory services for providers, pointing out opportunities for improvement, opportunities for expansion, as well as synergies with other providers.
- Strategic planning for the federation.

### 5.2.2 Desired Federation Services: Survey Results

H-CLOUD's survey of IT professionals gathered information on the type of services that should be prioritized for a European cloud federation.

Figure 3 compares respondents' preferences for general categories of service -- categorized by the traditional IaaS, PaaS and SaaS, as well as "data as a service" (DaaS). The H-CLOUD survey asked respondents about their preferences in general, as well as to support data spaces and cross-border services. Similar to the analysis in section 2.7, preferences from all respondents (N=96), users (N=78) and providers (N=50) are compared.

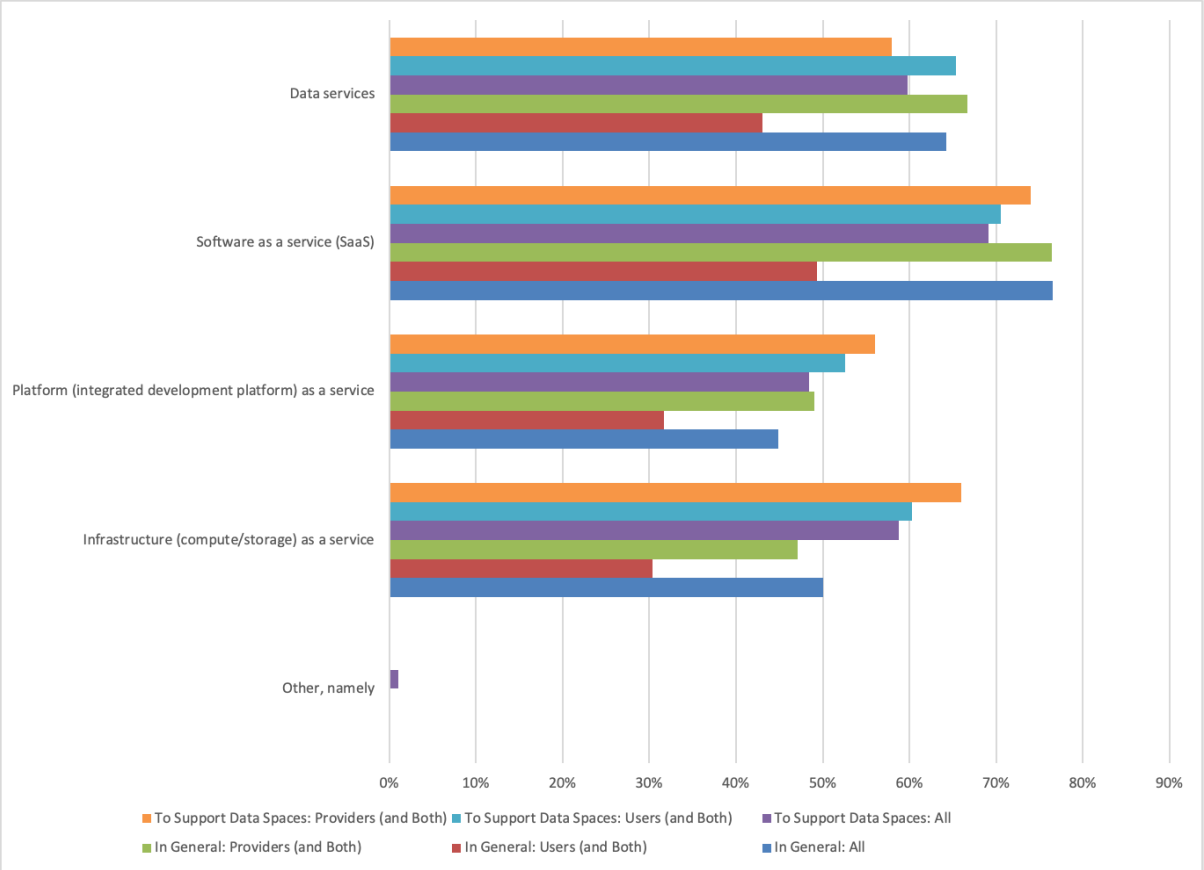


*Figure 3: Categories of Federated Cloud Services Preferred by IT Professionals*

In general, all respondent groups expressed broad interest in all four categories of cloud service (IaaS, PaaS, SaaS, DaaS), with slight preference for the platform-as-a-service (PaaS) category, followed by data-as-a-service. This pattern continues in the context of supporting data spaces and cross-border services, although among providers infrastructure-as-a-service (IaaS) edges out data-as-a-service to follow SaaS.

H-CLOUD's survey also asked respondents to highlight the five most important specific service classes, both in general as well as in the context of supporting data spaces and cross-border cloud services. Figure 4 compares respondents' responses to these questions.

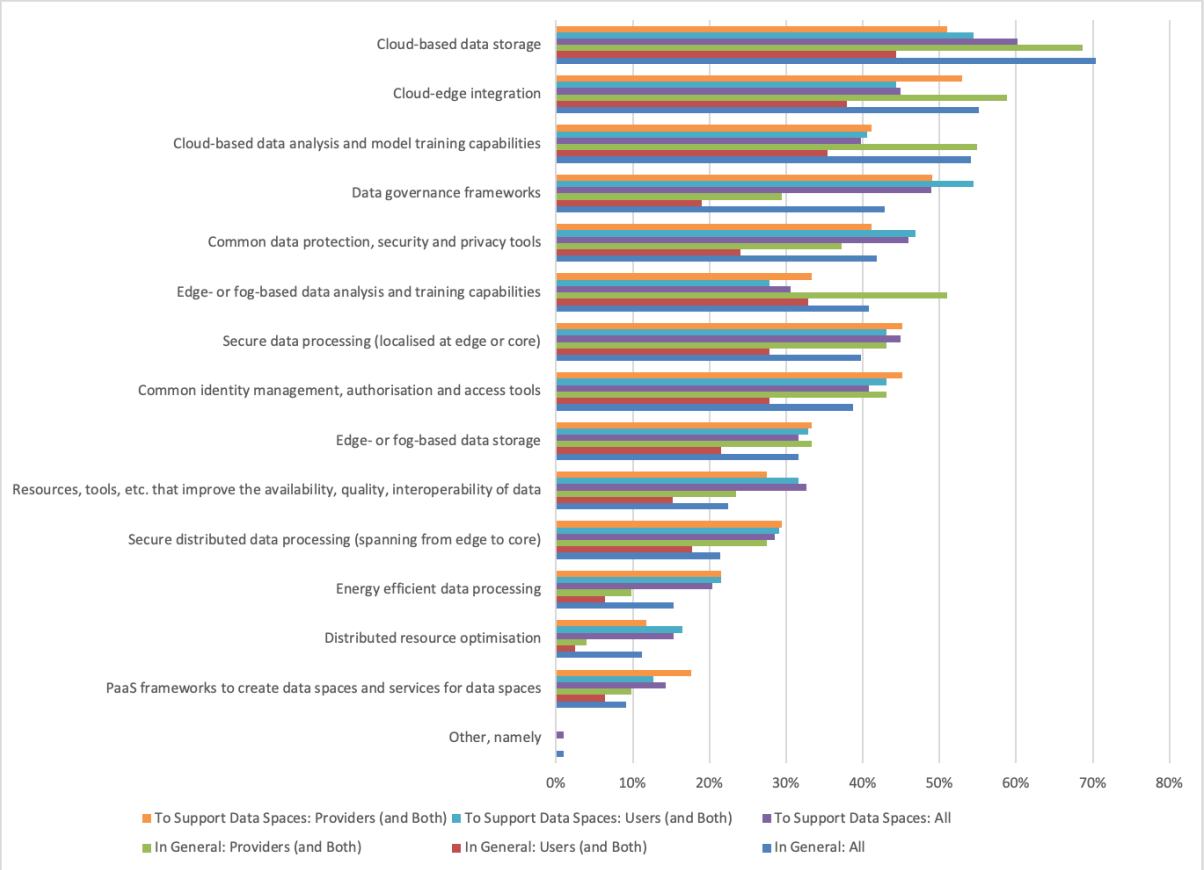


*Figure 4: The Most Important Classes of Specific Federated Cloud Services Identified by IT Professionals*

Figure 4 ranks service classes by their importance to all respondents in the general case, and there is broad alignment of the importance of the different service classes. However, in the general case some data-space-related services appear to have lower priority for users and providers than for all respondents, including for example “Data governance frameworks,” “Common data protection, security and privacy tools”. These shortfalls in importance disappear when the question is asked in the context of data spaces and cross-border services.

Two service classes related to edge are judged to be relatively important by these respondents -- “cloud-edge integration” and “edge- or fog-based data analysis and training capabilities” -- even though these services are still being developed. By contrast “energy efficient data processing” was not perceived as an important class of service by this set of respondents.

## 5.3 Architectures that Encompass Federation Services

A number of federated cloud initiatives have published their architectures, making it possible to compare their approaches and make recommendations relevant to the EUCF:

- EGI
- The European Open Science Cloud (EOSC)
- NIST Cloud Federation Reference Architecture
- Gaia-X.

EOSC is charged with supporting a wide range of scientific requirements, set by a diverse range of scientific communities. EOSC has defined a three-part architecture for its “federating core” -- the fundamental asset of the EOSC, composed of the technical, human, policy and

resource elements required to facilitate, monitor and regulate as appropriate day-to-day transactions across the federation[76]. This architecture includes a number of services highlighted in the "Implementation Roadmap for the European Open Science Cloud"[77] as well as other services required to enable proper functioning of the federation. (See Table 4.)

Like EOSC's architecture for its "federating core", the NIST CFRA describes a framework for defining and structuring the activities of a cloud federation, without specifying the technologies, standards or tools that will be used to execute each activity. The NIST CFRA was explicitly created to align and streamline the efforts of many different federated cloud initiatives, as well as offering guidance to federated initiatives in non-technical domains. Relevant standards are called out in Section 8 of the CFRA report, but their inclusion does not constitute endorsement.

The components of the EOSC architecture are listed in Table 4, along with a comparison with services/components that are included in the Gaia-X Technical Architecture, the EGI Federated Cloud and/or the NIST Cloud Federation Reference Architecture.

| Major Components of EOSC Federating Core, and Services included | Included in… | | | |
|---|---|---|---|---|
| | EOSC Implementation Roadmap | Gaia-X | EGI | NIST CFRA |
| **Hub Portfolio:** The activities and tools necessary to provide coordinated access to and management of resources provided in the EOSC Shared Resources or the Service Portfolio. EOSC resources are expected to be delivered at national and European level, together with the support and expertise necessary to address complex digital needs of the EOSC user communities: | | | | |
| ***Portal (service catalogue, marketplace, discovery services) **** | Y | Y | Y | |
| Support Services, including training, competence centres and knowledge bases | Y | | Y | |
| ***AAI (central -- when not provided by a user community) **** | Y | Y | Y | Y |
| Data Transfer services | | Y[78] | Y | Y |
| Monitoring | | Y | Y | Y |
| ***Usage Accounting**** | | Y | Y | Y |
| Configuration Management Database (CMDB) | Y | Y | Y | |
| Collaboration Software | | | Y | |
| Operations Portal | Y | | Y | |
| Security policies and security coordination functions | | | Y | Y |
| **Compliance Framework:** The policies and processes for suppliers and users to engage with the EOSC. | | | | |
| Rules of Participation | Y | Y | | Y |

[76] EOSC Federating Core Community Position Paper v1.01

[77] EUROPEAN COMMISSION Brussels, 14.3.2018 SWD(2018) 83 final COMMISSION STAFF WORKING DOCUMENT Implementation Roadmap for the European Science Cloud

[78] Although not specifically included in the Gaia-X architecture, the IDS "Data Collector" component provides mechanisms for the secure transfer of data between participating organizations.

| | | | | |
|---|---|---|---|---|
| Service Portfolio Management Tool | Y | | Y | |
| Interoperability Framework (including agreement to use common data standards or community APIs) | Y | Y[79] | Y | Y |
| Service Management System | | | Y | |
| **Shared Resources:** Resources including scientific artefacts and the storage and compute hosting platforms needed to deposit, share and process them. These resources are "federated" among the participants and might be shared or pooled: | | | | |
| High bandwidth networking | Y | | | Y |
| AAI services | Y | ?[80] | Y | Y |
| High-performance federated cloud storage | Y | | Y | Y |
| High-performance and high-throughput federated compute capabilities | Y | | Y | Y |
| Federated data search capabilities | Y | | Y | |
| Data repositories and data management tools | Y | | Y | |
| A catalogue of training materials and competence reference materials | Y | | Y | |
| Open Science policy and practice recommendations for institutions and other EOSC stakeholders | Y | | | |
| Software code repository | | | Y | |
| Persistent artefact ID services | Y | | | |
| Personalised workspaces for researchers, based on federated AAI | Y | | Y | |

*Table 4. Federated Architecture Components*

EOSC's extensive architecture highlights how other technical architectures can overlook key components of a complete federated cloud capability. Gaia-X in particular focuses on compliance and documentation of non-functional characteristics, as well as the mechanisms for secure data transfer, but provides limited guidance for operational activities (for example, who do customers call when there is a problem using Gaia-X?). Although the NIST CFRA mentions many important "human-to-human" functions, such as customer support, the CFRA focuses on the technical, machine-to-machine aspects of cloud federation.

Three components are highlighted in Table 4 in bold: portal, AAI and accounting capabilities. These three services might be regarded as the "minimum viable product" of a cloud federation. Many cloud federations only offer these capabilities, allowing users to access services on different federated facilities without requiring their interoperability, integration or standardization. Individual federated partners might separately decide to make those services interoperable, integrated or standardized -- but this is not a requirement of these federations.

Within EOSC's architecture, different technical components can be accommodated in a coherent way, allowing interfaces and standards between components to be managed, and identifying opportunities for rationalization and evolution. For example, Table 5 presents the

[79] Gaia-X intends to align with a wide range of data standards that are agreed or being finalized.

[80] Gaia-X's discussion of identity management, authentication and authorization contemplates federated AAI but favours centralized AAI mechanisms to support security and trust across the Gaia-X network.

current catalogue of supported components across the EGI Federated Cloud offering:

| | Functional group/Applicable standards and implemented specifications |
|---|---|
| **Federation Services** | **Monitoring/**ARGO REST API, Nagios API |
| | **Accounting**/APEL Grid Job Usage Record, APEL Grid Summary Job Record, Cloud VM Usage Record, OGF StaR |
| | **Messaging**/ARGO Messaging System (Google PubSub protocol) |
| | **Security**/Security Incident Response Trust Framework for Federated Identity (SIRTFI), Scalable Negotiator for a Community Trust Framework in Federated Infrastructures (SNCTFI) |
| | **Information Discovery**/GlueSchema |
| | **Software quality assurance**/IEEE 730-2014, ISO/IEC/IEEE 12207:2017, Semantic Versioning |
| **Compute and Storage (Cloud IaaS/HTC)** | **Compute Access/**OpenStack, Kubernetes, Open Container Initiative (OCI) |
| | **Storage Access**/SRM, GridFTP, WebDAV, S3, CDMI, POSIX |
| | **Application packaging**/Open Virtualization Format (OVF), Open Container Initiative (OCI) |
| **Federated Access** | **AAI**/SAML2.0, OAuth 2.0, OpenID Connect 1.0, X.509, LDAP, OAuth 2.0 Token Introspection, OAuth 2.0 Token Exchange, OAuth 2.0 Device Authorization Grant, SCIM 2.0 |
| | **Metadata schema and harvesting**/DataCite Metadata Schema, OAI-PMH |
| | **Compute Orchestration/**OASIS TOSCA |
| **Platform Services** | **Compute Orchestration/**OASIS TOSCA, DIRAC API |
| | **Notebooks/**Jupyter Notebook Format |

*Table 5: EGI Federated Cloud Architecture*

This comparison allows us to draw several conclusions:

- As much as possible, a “best practice” Federated Cloud Reference Architecture (FCRA) should reflect the NIST CFRA, EGI and Gaia-X’s technical architectures, as well as evolving it to ensure conformance of emerging federated cloud initiatives. This should specifically characterize how practical compliance frameworks and service catalogues would align with the EUSD’s contemplated “Cloud Rulebook” and “Service Marketplace” concepts.
- A federated cloud interoperability framework should be created and maintained as an evolving suite of technology, standards and tools that are consistent with the FCRA, allowing interoperation within a given federation and across multiple federations, compliance with European values and interoperability of identified components. This

suite of components would help EU customers navigate the many options for cloud-based solutions and would help formalize how they are described and the possibilities for integration. This recommendation is similar to “product labelling” recommendations in other industries, making it clear what is included in a given component, how it works, how it works with other components, and any limitations on performance.

- Funding in the coming Horizon Europe programme should ensure that research and innovation activities are aligned to support cross-domain applications to increase synergies, innovation potential and avoid duplication across the industry, research and public administration sectors. This recommendation is meant to increase innovation capacity of the EU programmes by ensuring that EU funded research and innovation has a large potential of adoption by multiple stakeholders across different sectors.

# 6 FEDERATED EDGE SOLUTIONS

The EUSD promotes the realisation of data spaces and federated clouds across the entire cloud-edge continuum. Edge computing, as discussed in Section 4.2 of the H-CLOUD Green Paper, provides a number of key advantages for certain scenarios, including: i) lower latency; ii) faster real time processing; iii) increased data security and privacy control. Such elements are clearly relevant for both data spaces and federated clouds.

Currently hyperscalers offer edge solutions to move data from customer's edge devices to hyperscalers' clouds: thus, creating a lock-in effect across the cloud-edge continuum. Individual edge "users" often invest in their own edge infrastructure -- for example, in smart city or smart grid applications -- but the scale of these investments limits both penetration and adoption. Potential edge users might be prepared to invest in remote sensors and Internet-of-Things (IoT) hardware, but they also need an efficient infrastructure to connect these devices to their own central or cloud data centres. Linking these remote devices directly to their data centre incurs high network costs (e.g. consider the bandwidth required for surveillance cameras), while intermediate buffering and storage devices trigger both increased hardware costs as well as increased costs to install and maintain these devices in the field. Investments in intermediate edge devices also create their own inertia or "lock in" effect, slowing adoption of newer edge technologies in the medium term.

A "public edge infrastructure" has been suggested as a solution to this problem, providing geographically distributed intermediate hardware to capture data from remote devices and potentially process that data in the field to minimize both communication costs and delays in processing (ensuring low latency). However, this kind of service, similar to IaaS for the cloud, has yet to appear: mobile network operators, the players with highest potential in this respect, have not yet moved ahead, and small players trying this path do not have the capacity to achieve geographical scale.

An alternative pathway could involve the sharing of existing networks of intermediate edge processing devices. For example, a power utility and municipality could both agree to share access to their respective collections of edge processing devices, giving each player increased geographic coverage across their service areas. If only two players are involved, bilateral contracts are the best approach, but as the number of participants increases, with each addressing a different service area, the beneficial "network effects" of a federation become clear.

Many of the federation business models described in Section 2.5 could be used, but integration of services is a definite requirement: that is, we need to confirm that a given intermediate edge processing device can connect and interoperate with one or more specific remote sensing or IoT devices. At the same time, we need to do this many, many times, rather than once at the beginning of a project, so the granularity of both services (the specific devices, located in specific places, with specific communications capabilities) and service integration requirements (are the connections permanent or active temporarily, how much data needs to be stored, what processing is needed) increases exponentially, compared, for example, to current use cases for federated cloud.

These quantitative changes drive qualitative changes in the solution, making automated discovery, selection, and provisioning of services and/or resources a requirement for federated edge computing to work. Contracting and payment could still be managed in a more predictable fashion, limiting dynamic provisioning to those resources offered by eligible edge service providers, or alternatively in a manner similar to the growing trend toward microservices, where small financial charges are made for many small service executions (i.e. individual instances of downloading and processing data from a particular remote device).

Federated edge computing faces a number of technical challenges, but the benefits are clear for both providers and customers:

- For providers: Edge service providers, even those well positioned in the market (e.g. major mobile operators), can improve the value of their services when they are combined with others, and they can potentially reach new customers.
- For customers:
  - Deployment of cloud-edge solutions can be scaled across Europe.
  - Edge-based services can be accessed without requiring customers to own their own intermediate edge equipment.
  - Data can be processed close to the source with edge-based services, rather than requiring all data to be "pushed" to the cloud.
  - Edge-based services can be adopted without being locked into the use of one cloud solution (multi-homing).

This federated approach could also accelerate adoption despite a lack of clear standards in edge technologies, as long as each edge device is registered with the federation as complying with specific edge communication standards. A dynamic provisioning process could filter feasible devices by the standards they support, as well as the physical location and hardware criteria described above. This process could also help identify gaps in service that federation members could potentially fill, allowing providers to increase revenues and better serve customers.

# 7 FEDERATED DATA SOLUTIONS

The EUSD and other initiatives highlight the growing benefits and importance of data sharing and the possibility of using federated data to enable this sharing. As documented elsewhere in the H-CLOUD Green Paper v1.0, important ecosystems in public administration, healthcare and transport need to enable secure access, sharing and analysis of sensitive data already being stored and managed by ecosystem players -- often on private cloud infrastructure. There is great potential for federated data clouds to create “data as a service” capabilities that can be accessed more broadly, yet still on a controlled basis, to enable the projected societal benefits of both “big” and “open” data.

Despite its potential to support secure, private sharing of data held by many different organizations, best practice roadmaps are urgently needed to ensure federated data sharing initiatives are established and operated efficiently while preserving and ensuring the highest level of confidence that affected sensitive data will be kept private and secure.

## 7.1 Federated Data as a Flexible Approach to Data Sharing

Federated data is one operational approach to the problem of how to “share data,” and specific instances of data federation need to align with the overarching governance context of the corresponding data sharing activity. The flowchart presented by the High-Level Expert Group on Business-to-Government Data Sharing[81] addresses governance issues in the following sequence:

1. What is the problem?
2. What data is needed to potentially solve the problem?
3. What are the benefits and harms of using this data to try to solve this problem (likelihood and intensity of possible benefits and harms; immediacy/urgency of the situation; potential harm of non-use of data)?
4. What conditions are needed to maximize benefits and minimize harms and to produce the desired benefits at an acceptable social cost?
5. Who needs to be involved in this activity (data holders, data owners, etc.)?
6. What operational and technical approach should be used to generate the needed solution or search for the desired insight?
7. How will the solution/insight be communicated and acted upon?
8. What steps are needed to insure transparency and accountability with all relevant stakeholders?

Federated data is a possible answer to the 6th question in this list -- rather than the first question that needs to be addressed. Whatever operational approach is chosen, it must align with the purpose and stakeholder relationships defined in the steps listed above. Best practices in data governance are outside the scope of this report, but it is clear that all data governance regimes -- and their chosen operational and technical solutions -- must keep answers to questions of “why”, “for whom” and “who is affected” clearly in view.

Various operational and technical approaches have been identified by the High-Level Expert

[81] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64954, p. 48

Group (referenced above), as well as Huyer et al. (2019)[82], Klein and Verhulst (2017)[83] and Verhulst et al. (no date)[84], including:

- Contractual data sharing,
- Data pooling, trusted intermediaries, data partnerships,
- Use of APIs.

Federated data approaches can be used to flexibly implement all of these different data sharing approaches:

### 7.1.1 Contractual Data Sharing Structures

- Data access would be "packaged" as remotely accessible data services provided by each party acting as a data provider through contractual agreement. These data services would not be exposed through a marketplace or service catalogue.
- Federated identity management would ensure that access to each service is limited to authorized individuals (or requesting services).
- Semantic interoperability of the data would be addressed directly between the contract parties.
- Governance requirements would be implemented implicitly by the kinds of data pre-agreed to be provided to different requesters and explicitly through contractual agreements between the parties. The federated data solution would not be able to check whether or not shared data was being used for the pre-agreed purpose.

### 7.1.2 Collective Data Sharing Approaches

These include data pooling, trusted intermediaries and data partnerships. They all share similar characteristics:

- Whether the various individual organizations agree to a partnership or a separate intermediary organization is created, all of these parties could join a federation and interact as members through the federation. If a separate intermediary organization is not created, a virtual version would still need to take responsibility for the initiative's responsibilities and would probably join the federation as a separate member. Each collective data sharing initiative would need to determine how to govern data access and use and how to ensure accountability through the intermediary organization back to individual data providers and ultimately to data owners.
- Technically speaking, the "pool", "intermediary" or "partnership," real or virtual, would create and operate one or more new data services accessible through the federation, possibly visible in the federation's service catalogue.
- The collective organization would perform one time, continuing or "as needed" data ingest from partner data providers, process the data as required, implement security and privacy-preserving functions consistent with the agreed governance model, and

[82] Huyer, E., and Cecconi, G., and Capgemini Invent, *Analytical Report 12: Business-to-Government Data Sharing*, European Data Portal, March 2019, pp. 15-18, (https://www.europeandataportal.eu/sites/default/files/analytical_report_12_business_government_data_sharing.pdf)

[83] Klein, T., and Verhulst, S., 'Access to new data sources for statistics: business models and incentives for the corporate sector', OECD Statistics Working Papers, No 6, 5 May 2017, OECD Publishing, Paris, 2017, (https://doi.org/10.1787/9a1fa77f-en)

[84] S. Verhulst, A. Young and P. Srinivasan, (no date), An introduction to data collaboratives, The GovLab, (http://datacollaboratives. org/static/files/data-collaboratives-intro.pdf).

fulfil requests to the collective data services as appropriate (based on the identity of the requestor and the nature of the request).

- Governance requires that each data request be evaluated against agreed criteria, e.g. purpose of data sharing, extent of controls in place at the requester to ensure continued privacy/security of any sensitive data being requested, consent of data owners to the disclosed purpose of the sharing, etc. For example, where a requester cannot provide adequate protection for sensitive data, that data might need to be automatically sanitized, aggregated and/or encrypted to comply with the agreed governance rules.
- In the particular case of research data partnerships, or other partnerships formed with the promise of mutual benefit, evaluation of data requests against questions of "purpose" might be relaxed in favour of the mutual benefit, but security/privacy of sensitive data would still need to be maintained.

These collective approaches are similar to but should not be confused with "data lakes"[85], which in and of themselves lack mechanisms (technical and operational) to implement required governance.

### 7.1.3 Application Programming Interfaces (APIs)

APIs are a common technical mechanism by which data services are accessed. Requests, formatted according to the API specification, are sent to the data service (computer servers) designated for handling requests. The data service determines whether the request can be granted and, if so, arranges for the requested information to be assembled, and possibly processed, before being sent back as a response to the requestor. APIs (custom or standard) would be needed for each data service.

For data sharing, generic data access APIs must be enhanced to capture and process related governance factors: is the data consumer authorized to request the kind of data being requested, what certifications does the data consumer have, what modifications (sanitization, encryption, aggregation, etc.) are required to allow some version of the requested data to be shared, is the proposed data usage consistent with the consents obtained from the data owners, etc.

APIs can be expanded conceptually to "bring compute to the data" by allowing data processing activities to be packaged so that they can be executed by the data provider. This minimizes data movement as well as the potential exposure of sensitive data in transit or by the data requester, but this also requires assessing the portable processing software to ensure the safety of its operation. This approach is reflected in the Open Algorithms (OPAL) project[86] as well as the trusted applications offered through International Data Spaces' AppStore.

The roadmap for federated data sharing described above should include specific guidelines for using federated platforms to enable different forms of operational approaches to be easily implemented by data sharing communities.

### 7.1.4 The IDSA Reference Architecture

International Data Spaces Association (IDSA) has developed a useful reference architecture[87] that can accommodate some of the approaches described above. The IDSA springs from the

[85] "While a data warehouse is a repository for structured, filtered data that has already been processed for a specific purpose, a data lake is a storage repository for large amount[s] of structured, semi-structured and unstructured data the purpose for which is not defined yet. In the data lake, all data is kept irrespective of the source and its structure and it is transformed when it is ready to be used. " IDC, 2020, The Secondary Use of Health Data and Data-driven Innovation in the European Healthcare Industry.

[86] https://www.opalproject.org

[87] IDSA Reference Architecture Model (RAM) 3.0 https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf

manufacturing sector, with a focus on industrial Internet of Things (IoT) and corporate data sovereignty, rather than addressing data sovereignty for individuals or other types of organizations (e.g. municipalities, agriculture).

The IDSA's reference architecture centres on "point to point" data exchanges between two participants in the IDSA ecosystem, the Data Provider and Data Consumer. Exchanges use "IDS Connectors", which enhance the capabilities of APIs by integrating automated negotiation of governance issues, for example, determining whether the Data Consumer's proposed use of data is acceptable to the Data Provider. Initial discovery of Data from Data Providers is managed by IDS Brokers, and transactions between Providers and Consumers are logged by the IDS Clearinghouse. IDS Core Participants, as well as entities with other roles, are each certified against strict criteria, including, e.g. compliance with ISO 27000 security standards, ensuring that the full IDSA ecosystem operates as a community of highly trusted parties. This is an important feature of the architecture, since the architecture itself cannot technically guarantee that Data Consumers will respect the limitations on use set by Data Providers -- Data Providers have to "trust" the Data Consumers.

The IDSA architecture is based on the premise that bilateral data sharing is the only form that can ensure adequate data sovereignty -- in contrast with other forms (such as data pooling, trusted intermediaries, data partnerships)[88]. Integration of data from multiple sources, e.g. by a trusted intermediary, would require that intermediary to be certified to become an IDS Core Participant, request data from multiple Data Providers and then manipulate the received data in order to generate new knowledge and insights. IDSA's position, while apparently simplistic, actually reflects the difficulty for data pooling, trusted intermediaries, and data partnership approaches to sustainably implement the controls required for proper governance. As IDC notes (addressing data lakes more generally): "skepticism and challenges in adopting data lakes solutions are grounded in data quality, governance, security and privacy"[89].

### 7.1.5 Data Usage Models

Most data sharing initiatives establish "data usage models" as a common framework across the initiative to describe different ways that shared data might be used and to give data providers a language to reliably describe what can and cannot be done with the data they provide. Unfortunately, differences between usage models, and usage model philosophies, in different domains and sectors make it difficult to automate negotiation of data sharing between particular data providers and consumers and stand in the way of effective data sharing through collective efforts such as trusted intermediaries.

The IDSA Reference Architecture's usage control model[90] gives IDS Data Providers a way to describe "policies" for allowable uses of their data, as well as the scope of application of those policies. The goal is to provide enough detail to allow automated negotiation between Data Providers and Consumers regarding acceptable data use. Policies can cover allowed and prohibited uses, limiting use to specific users, timeframes, in connection with specific events, not more than a certain number of times, and requiring a variety of actions after use (deletion, notification and logging). Although structurally detailed, the "contents" of these policies are not as detailed, so the nature of uses that are allowed or prohibited has not been defined in the reference architecture, suggesting that fully automated use negotiation may be difficult to achieve without a detailed taxonomy of data use that will reliably capture the broad range of situations that might be encountered.

Several other usage models are worth considering for comparison:

[88] "Decentralized data exchange by means of Connectors, in contrast to other architectures of data networks (e.g., data lakes or cloud services), ensures full data sovereignty." p. 103 of the IDSA RAM v3.0.
[89] IDC 2020, ibid, p. 19
[90] https://www.internationaldataspaces.org/wp-content/uploads/2020/06/IDSA-Position-Paper-Usage-Control-in-IDS-2.0.pdf

- While IDSA's usage model provides tools for a corporation to control the sharing of its corporate data, the FAIR principles have emerged to promote open access to data collected and/or generated by research. "FAIR" stands for Findable, Accessible, Interoperable and Re-usable. The intent of the FAIR principles is summarized in this 2013 statement from the G8 Science Ministers: "[o]pen scientific research data should be easily discoverable, accessible, assessable, intelligible, usable, and wherever possible interoperable to specific quality standards."[91] FAIR makes no distinctions among possible uses and strives to enable all uses that are consistent with open access. For example, works derived from research data are encouraged, as long as they themselves can be openly used and re-used, perhaps with attribution.
- "Open access" to research data has been increasingly challenged by the growth of personally identifiable information (PII) collected in the course of life sciences research. The Global Alliance for Genomics and Health (GA4GH)[92] is responding to this challenge and working to accelerate progress in genomic research and human health, by cultivating a common framework of standards and harmonized approaches for effective and responsible genomic and health-related data sharing. An ongoing project at GA4GH, working with partners around the world, including EMBL in Europe, is the Data Use Ontology (DUO) project[93], which has built and continues to refine a detailed ontology describing possible uses for specific data (e.g. general research, research on paediatric lymphomas, etc.), each combined with specific data use requirements (not for profit only, collaboration required, publication embargoes, IP claim terms, payment terms, etc.), to be matched by comparable details on the proposed data use (e.g. specific categories of research, for profit research, etc.). GA4GH's DUO system will support increasingly automatic evaluation of appropriate re-use of genomic data, mostly intended to support scientifically appropriate re-use, but also enabling/controlling non-scientific uses, or uses that have intellectual property or compensation implications. Even with this framework, data sharing remains a significant challenge[94].
- Moving from health research to health care, Health Level Seven International (HL7)[95] is a not-for-profit, ANSI-accredited standards developing organization dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services. HL7 is supported by more than 1,600 members from over 50 countries, including 500+ corporate members representing healthcare providers, government stakeholders, payers, pharmaceutical companies, vendors/suppliers, and consulting firms. The HL7 data standard provides an extensive list of "reason codes" covering all activities in healthcare from treatment of individuals and public health activities, to billing and reimbursement and marketing. Detailed activities include emergency room treatment, insurance eligibility verification, military discharge, and testing of healthcare IT systems. Reason codes are primarily used for activity-based coding of information, but they are also used to manage inter-organizational requests for healthcare-related data, for example verification of insurance coverage.

These few examples, drawn opportunistically rather than systematically from adjacent domains (research, genomics research, health care), quickly illustrate significantly different data usage models as well as philosophies. These different usage philosophies can motivate different

[91] G8 Science Ministers Statement, 13 June 2013 https://www.gov.uk/government/news/g8-science-ministers-statement
[92] www.ga4gh.org
[93] https://github.com/EBISPOT/DUO
[94] Phillips, M, et al. Nature 578, 31-33 (2020). https://doi.org/10.1038/d41586-020-00082-9
[95] www.HL7.org

technical architectures, which in turn can stand in the way of data interoperability. These differences can be particularly hard to harmonize when sharing data through collective approaches such as data pools, partnerships or intermediaries.

Efforts to increase semantic interoperability for data within and across sectors are critical and must also include harmonization of data usage models to enable automated, yet secure and appropriate, data sharing. Existing data usage models should be systematically analysed in hopes of finding common philosophies and features that can form the basis for a common, cross-sector usage model that could facilitate greater levels of properly governed data sharing.

## 7.2 Technical Challenges to Data Sharing

The H-CLOUD Green Paper identified the following Major Challenge (M3): "Secure and trusted data access, sharing and processing across different organisations". The EUSD itself notes that "[t]he analytical tools come to the data, not the other way around. This makes it easier to keep the data secure and to ensure control over who accesses what data for what purposes."[96] The EUSD also highlights the need for continuing research on "technologies that are crucial for the next stages of the data economy, such as privacy preserving technologies and technologies underpinning industrial and personal data spaces."[97] This section explores these issues in greater detail.

### 7.2.1 Increased Granularity and Detail in Data

Section 7.1.2 above illustrated the need to record and manage the purpose of every instance of data sharing. This information must be retained after the request for data sharing -- in general it becomes part of the "provenance" record of any data produced as a result of data exchange. Provenance data is a special type of metadata, tracking the sources, purpose, and processing associated with the production of each element of new or processed data. In the IDSA Reference Architecture, provenance is captured by the IDS Clearinghouse, but similar mechanisms are needed in all data sharing initiatives.

The EU's General Data Protection Regulation mandates similar record keeping and logging, allowing audits by both competent authorities and data subjects of the uses of personal data. In addition to these transactional records, mechanisms are needed to collect and track data subject consent to different forms of use. One analysis of the technical implications of GDPR has recommended the use of "sticky policies": "a security policy, which is associated with (stuck to) a piece of data such that access to and use of that data is only possible if the policy has been complied with. Attempts to un-stick the policy or modify or replace it should render the data inaccessible."[98] (This is distinct from the separate and significant challenge of notifying data subjects about requests for use of their data, collecting and managing their responses, and possibly generalizing those consents for future uses.)

Common access control mechanisms, such as role-based access control (RBAC), are also being supplanted by more flexible access models, such as attribute-based access control (ABAC). ABAC decides on data access based on a broader variety of factors including requester attributes, context of the request, and can be tuned to support appropriate access depending on the sensitivity of the data. As with the data usage models above, ABAC requires detailed modelling of the access and use environment in order to benefit from ABAC's added expressiveness[99].

[96] EUSD, p.13
[97] EUSD, p.22
[98] RestAssured 2018. Methodology for Decentralized Data Lifecycle Management. Deliverable D6.1, https://restassuredh2020.eu/wp-content/uploads/2018/07/D6.1.pdf
[99] Verginadis, Y., Michalas, A., Gouvas, P. et al. PaaSword: A Holistic Data Privacy and Security by Design Framework for Cloud Services. J Grid Computing 15, 219–234 (2017). https://doi.org/10.1007/s10723-017-9394-2

These requirements add to already growing requirements to manage metadata associated with data, but exponentially increase the challenges, since data provenance and sticky policies might be attached to individual data fields and records, rather than to files or datasets. They also challenge the growing paradigms of object-based storage and key value stores, which disaggregate the contextually rich environment of a traditional relational database management system (RDBMS).

### 7.2.2 Embedded Data Protection Solutions Needed

The analysis of sensitive data has traditionally employed a "walled garden" approach, with comparatively unconstrained data access within the walled garden (for example, within an enterprise IT establishment), and controls applied at the wall for both access into, and potential export of data outside, the walled garden. This paradigm is reflected in the IDSA Reference Architecture, where the IDS Connectors and the broader IDSA data ecosystem provide the conditions and controls to limit the release of sensitive data only to trusted parties for agreed purposes. It is also reflected in the "health data lake" initiatives underway across the EU. Both of these approaches have limited scalability and sustainability.

Sustaining and supporting the collective data sharing approaches described above, as well as, more generally, the broad idea of "data spaces" contemplated by the EU, requires a new approach, built on "privacy [as well as security] by design". This new approach combines:

- **Threat analysis of potential data sharing and exchange scenarios**. Kairouz et al (2019)[100] illustrate the threat analysis process that is required, applying it to federated machine learning, a promising technique for secure analysis of distributed data that nevertheless still suffers from vulnerabilities that require careful mitigation.
- **Design and implementation of a data protection solution, incorporating multiple technologies needed to address possible risks.** Several reviews of available security and privacy preserving technologies[101] identify multiple promising approaches, while also emphasizing the need to employ more than one in order to respond to the range of possible threats. Key technology categories include homomorphic encryption, secure multi-party computation, differential privacy, federated machine learning, distributed ledgers (blockchain), trusted execution environments, zero-knowledge proofs, and automated risk management and compliance tools.
- **Operation of the system, including run-time evaluation of the risks associated with each transaction and invocation of specific technologies needed to mitigate transaction-related risks.** Mohammadi et al (2018)[102] conclude that effective data protection can only be optimized at run-time, in order to respond to the specific risk factors associated with each data exchange or sharing transaction.

The Big Data Value Association (BDVA) provides a valuable guide[103] to the required approach,

[100] Peter Kairouz, H. Brendan Mcmahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, et al.. Advances and Open Problems in Federated Learning. 2019. https://hal.inria.fr/hal-02406503

[101] Cappiello, Cinzia & Gal, Avigdor & Jarke, Matthias & Rehof, Jakob. (2020). Data Ecosystems: Sovereign Data Exchange among Organizations. Report from Dagstuhl Seminar 19391. https://drops.dagstuhl.de/opus/volltexte/2020/11845/. e-SIDES (Ethical and Societal Implications of Data Sciences), Deliverable 3.2, Assessment of existing technologies (2018), https://e-sides.eu/resources/deliverable-d32-assessment-of-existing-technologies. UN Global Working Group on Big Data 2018, UN Handbook on Privacy-Preserving Computation Techniques, http://publications.officialstatistics.org/handbooks/privacy-preserving-techniques-handbook/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf

[102] Mohammadi, Nazila & Mann, Zoltan & Metzger, Andreas & Heisel, Maritta & Greig, James. (2018). Towards an End-to-End Architecture for Run-Time Data Protection in the Cloud. 514-518. doi.org/10.1109/SEAA.2018.00088.

[103] Timan, T. & Z. Á. Mann (eds) (2019) "Data protection in the era of artificial intelligence. Trends, existing solutions and recommendations for privacy-preserving technologies", October 2019. BDVA.

noting "[t]echnologically, this means that data architectures and logics need overhaul". In this context, the BDVA makes several recommendations:

- **Develop regulatory sandboxes.** These sandboxes would allow experimentation and scaled-up testing of privacy-preserving technologies, while ideally exempting the enterprises that need data from the responsibility to prove that they have all the necessary security measures in accordance with the legal precepts.
- **Continued support for research, innovation and deployment of privacy-preserving technologies.** Both the BDVA and e-SIDES reports enumerate a number of recent or ongoing projects to develop, refine and deploy these technologies in practical application domains. While these technologies are at various technological readiness levels (TRL), and not mature enough for "production" deployment, they are nevertheless promising and would benefit from continued investment and support for early stage adoption and deployment.
- **Support and contribution to the formation of technical standards for preserving privacy.** Such standards would provide risk assessment tools, test suites for validation of performance, as well as evaluation of data for sensitive content.

The BDVA's recommendations are reflected in this document's recommendations below. Note in particular that these recommendations align with the proposed response to the Green Paper's Major Challenge M3: "Secure and trusted data access, sharing and processing across different organisations".

## 7.3 Tight Coupling of Data and Data Processing

Both the EUSD and Gaia-X conceptually separate their treatment of data spaces from an underlying federated cloud capability. However, data is stored, physically, in that same federated cloud capability. For some data, for example, Internet-of-Things data from specific devices, data volumes may not be significant, and divorcing the data from its storage, transmission and processing can be effective as a conceptual approach. However, as data volumes increase, and as the data in a "data space" are more widely distributed across multiple data owners and data storage systems, the actual physical location of the data is increasingly relevant to orchestrate efficient analysis of that data, as well as to maintain the security and privacy of that data.

Demchenko et al. (2014)[104] capture this redistribution of both storage and computational capabilities in their proposed architecture for a big data ecosystem. Mazumdar et al (2019)[105] note "it remains a challenge to optimally store and place or migrate ... huge data sets across data centres (DCs). In particular, due to the frequent change of application and DC behaviour (i.e., resources or latencies), data access or usage patterns need to be analysed as well." Their analysis shows that there are no clear, or standard, solutions to this problem and that more research and development are needed. Finally, several important privacy-preserving technologies identified in section 7.2.2 above, such as homomorphic encryption and secure multi-party computation, entail significant computational requirements, so computational resources must be available where the data is stored in order for the data to be kept private.

All of these issues are aggravated by exponential increases in data volume. For example, earth observation satellite data from Copernicus are too large to transfer from where they are stored to where a customer might process them, so special data processing services are needed to

https://www.bdva.eu/sites/default/files/Data%20protection%20in%20the%20era%20of%20big%20data%20for%20artificial%20intelligence_BDVA_FINAL.pdf

[104] Demchenko, Yuri & Laat, Cees & Membrey, Peter. (2014). Defining architecture components of the Big Data Ecosystem. 104-112. doi.org/10.1109/CTS.2014.6867550.

[105] Mazumdar et al. (2019) A survey on data storage and placement methodologies for Cloud-Big Data ecosystem, J. Big Data 6:15, https://doi.org/10.1186/s40537-019-0178-3

process the data where they are stored and deliver (smaller) results to a customer. Similar problems are growing in public health, life science, astronomy, as well as in industry. Optimized solutions have been developed in climate and earth sciences[106] as well as other fields of science and may represent promising approaches for similar “big data” problems in society.

While keeping data and data processing separate might allow underpinning data spaces and federated clouds to focus on significant issues around standards and harmonized operations, the reality is that there are important technical dependencies that require both levels to be configured, managed and orchestrated together.

## 7.4 Technical Services Required to Support Federated Data and Data Spaces

Supporting federated data services, and enabling the data spaces contemplated by the EUSD, will require a range of common technical services. These can be grouped into two categories: general data-enabling services, as well as specific services required to enable the secure processing of sensitive data.

### 7.4.1 Services Enabling Data Spaces

The experience of research infrastructures such as EGI identifies a number of key enabling services that need to be provided to support data spaces:

- Tools for adding data resources to the data space (secure upload or data transfer, optimized for large datasets, or linking to existing versions of the data resource).
- Tools for adding or updating metadata descriptions and annotations for each data resource, including defining policies around visibility and acceptable data usage.
- Defining licensing, intellectual property rights, compensation mechanisms as appropriate. For FAIR research data these do not usually apply, but other types of data will often have these attributes.
- Identifying and making accessible any software or tools required to “understand”, access or process a data resource.
- Creating persistent identifiers (PIDs) for each data resource. Depending on the type of PID to be created, there may be certain rules about accessibility and use.
- “Publishing” digital resources (including version control, exposing the resource for controlled access, exposing metadata to allow searchers to find the resource)
- Metadata harvesting -- collecting new or updated metadata and using that information to enable search and discovery.
- Tools for searching data spaces for certain kinds of data. These tools may require authentication (some data may not be visible to un-authenticated users, or even to some classes of authenticated users), and typically operate on metadata provided for each data asset.

[106] Sandro Fiore, Cosimo Palazzo, Alessandro D'Anca, Donatello Elia, Elisa Londero, Cristina Knapic, Stephen Monna, Nicola M. Marcucci, Fernando Aguilar, Marcin Płóciennik, Jesús E. Marco De Lucas, and Giovanni Aloisio. 2017. Big Data Analytics on Large-Scale Scientific Datasets in the INDIGO-DataCloud Project. In Proceedings of the Computing Frontiers Conference (CF'17). Association for Computing Machinery, New York, NY, USA, 343–348. https://doi.org/10.1145/3075564.3078884; Hu, Fei & Yang, Chaowei & Schnase, John & Duffy, Daniel & Xu, Mengchao & Bowen, Michael & Lee, Tsengdar & Song, Weiwei. (2018). ClimateSpark: An in-memory distributed computing framework for big climate data analytics. Computers & Geosciences. 115. doi.org/10.1016/j.cageo.2018.03.011.

- Metrics capture and reporting. What are users searching for, how often is a particular dataset returned as a search result, how often is it viewed and accessed?
- Managing metadata schemas and vocabularies.

These services build on more general federated services such as identity management, authentication and authorization services, mechanisms to generate and validate trust certificates (e.g. X.509) to support certain security requirements, as well as a variety of data storage services (object, file and block storage, database management systems of different kinds, streaming data ingest and management).

**Provenance Registries**: Another service that is not currently commonly required, but would be desirable, in a research environment is maintenance of provenance information about each data resource. Initially a data resource's provenance will be provided by the first provider of the data, explaining how the data was created or collected, the specific methods used, as well as introductory information about how to use the data. However when an existing data resource is used by a data consumer to produce a new data resource, the provenance of that new data resource should incorporate the provenance of the original data sources, as well as details of how the new data resource was created (methods, tools used, etc., limitations). The research community is building awareness of the need for this kind of provenance data, and it is sometimes generated as part of automated scientific workflows (which can automatically record the input data sources and processing steps and then generate provenance data for each output generated by the workflow step). As discussed above, this generic service is required when certain kinds of sensitive data are processed, but it is listed here since it is more generally valuable.

**Data replication, verification, archiving and backup:** A number of services are required to ensure the integrity and availability of each data resource and manage policies regarding its retention. In personal life, data integrity might be ensured by using a backup service. In a production IT environment, this function may translate into storing multiple copies of a data resource (replication), cross checking them to determine if any have been corrupted (verification) and replacing corrupted versions with clean versions. In parallel there is also an archiving process to ensure continued accessibility of the resource for a defined (or indefinite) period. Depending on usage patterns, data resources may be migrated to lower cost storage systems, including tape storage, or alternatively may be replicated in more places to increase availability for users[107]. Very high usage patterns may require the use of data caching and content distribution networks, such as are used for streaming video, to achieve the desired availability.

**Big Data Analytics and Machine Learning Frameworks**: These include analytics services found in many IT environments (Hadoop, Spark, Hadoop Data File System (HDFS)), as well as statistical packages like R. Many machine learning frameworks are available, including TensorFlow and PyTorch.

There are also specialized "big data" analytics platforms optimized for particular domains, such as Ophidia[108], optimized for analysis of climate data.

### 7.4.2 Services to Support Privacy Preservation

Section 7.2 listed a number of promising privacy preserving technologies:

- homomorphic encryption,
- secure multi-party computation,
- differential privacy,

[107] Replication also has implications for data processing efficiency, as noted by Mazumdar et al (2019).
[108] http://ophidia.cmcc.it/

- federated machine learning,
- distributed ledgers (blockchain),
- trusted execution environments,
- zero-knowledge proofs, and
- automated risk management and compliance tools.

These technologies exhibit varying technological maturity, so they have not all reached the point where production-ready IT implementations can be described. Nevertheless, it is reasonable to expect that most will be implemented and packaged as software services that can be set up at multiple sites where they can be exposed as services and accessed by customers to support privacy preservation.

In the case of federated machine learning, TensorFlow Federated provides a standard framework for setting up a federated machine learning network. Multiple production-grade distributed ledger solutions are also available.

# 8 CHALLENGES AND RECOMMENDATIONS

**S-F Challenge 1: Coordinated/federated approaches must be structured around the objectives of their stakeholders, balancing community focussed initiatives with pan-European solutions.** The European landscape for cloud- and data-driven innovation is complex and fragmented, with many potential use cases, customers, providers, innovators and stakeholders. The EUSD itself addresses a range of requirements and opportunities across nine sectors of the economy. The needs of different stakeholder communities must be balanced against the need for common or aligned solutions. The effectiveness of a EUCF will depend strongly on the clarity of its value proposition and how it is constituted to realize that value proposition.

**S-F Recommendation 1.1: Develop detailed business cases for identified use cases in each of the nine sectoral data spaces described in the EUSD that quantify the societal gains and costs to achieve the desired benefits and ascertain feasibility and related ICT innovation needs.** Identify existing initiatives and high-impact use cases (e.g. existing health data hubs), elaborate specific data sharing use cases building on existing good practices. Identify data and cloud resources that might be good candidates for sharing and re-use through federated structures. Quantify specific gains and costs for businesses, research organizations and Public Administration to use federated cloud and data services as a platform for cross-sector data sharing involving private data, public data and governmental data. Ideally this would follow process similar to that described by the High-Level Expert Group on Business-to-Government Data Sharing[109] (page 48) which starts by identifying the problem to be solved, the conditions for data re-use, possible compensation structures, and then considers the optimum model for data access. Conduct a requirement and gap analysis to identify services needed from a EUCF, as well as the stakeholders that would need to be involved in their delivery and supervision. This analysis may identify clusters of requirements where solutions at the correct technological readiness level are available and could be deployed, for example to ensure secure and interoperable access to data and cloud services. This analysis will also identify gaps in existing solutions that should be prioritized for additional research and/or development. [Deployment, Research]

**S-F Recommendation 1.2: For each business case, select the most appropriate federation business model that fulfils the requirements while providing the best value with the least effort.** Consider whether the proposed action is a response to a market failure, which can be corrected with a temporary action, or whether continued support is needed to correct a systemic problem.

**S-F Recommendation 1.3: Create an open infrastructure and testing capability that could flexibly support demonstrations, proofs of concept and pilots of how federated cloud and federated data solutions could be assembled, operated, managed and governed, including collection of data that would validate the business cases developed in S-F Recommendation 1.1.** Before establishing a formal EUCF, invest to encourage participation in federated cloud and data sharing pilots and create a dedicated virtual support and training centre. This would bring together a number of customer organizations (e.g. public administrations, industries and research organizations with a specific data sharing use case) to define requirements, which would then be addressed by providers integrating existing tools (i.e. no research and minimal software development), and solutions would be supervised using best-practice federated governance models. As skills are a major asset for a successful repurposing and re-architecting of applications by potential cloud customers, a support centre will be necessary to provide the technical expertise needed for an effective use of the pilot infrastructure. Particular attention should be paid to provider costs and assessments of value received, in order to identify sustainable business cases for continuing operation. Based on the analysis presented in the H-CLOUD Green Paper, promising use cases can be found in

[109] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64954

each of the public administration, transport/mobility and health care sectors, for which sectoral data spaces are all proposed in the EUSD. [Deployment]

**S-F Recommendation 1.4: Support the creation of multiple EUCF-affiliated initiatives** and their cross-domain collaboration, which will specify domain-specific use cases, objectives and beneficiaries, federation partners and stakeholders, governance and decision-making mechanisms, scope of possible federated activities, and applicable business models. [Deployment]

**S-F Recommendation 1.5: Develop a lightweight model for the EUCF** as an umbrella coordinating body of sector- or use-case-focussed European federated cloud initiatives, supporting coordination of their research and innovation activities, cross-sector collaboration on interoperability, facilitating best practice operations, and providing relevant shared services such as certification activities. [Deployment]

**S-F Recommendation 1.6: Implement a EUCF with a phased approach that flexibly aligns activities across multiple domains, and that allows achievement of "quick wins".** Recommendations 1.1 through 1.4 highlight the many topics that must be addressed before a EUCF can be organized and functional. Pilot projects and demonstrators will clarify requirements and identify applications and use cases where an early version of the EUCF can achieve success, which in turn will build credibility and support. [Deployment]

**S-F Recommendation 1.7: Set up the EUCF following known organizational recommendations.** Based on research into multi-organizational alliances, a EUCF should adopt the following approaches [Deployment]:

- Form a central organization focussed on the coordination efforts required to make a EUCF work,
- Establish and follow well-documented procedures (for example, governance protocols) to improve its effectiveness,
- Establish communications mechanisms that support a broad set of interfaces among the EUCF, its primary and affiliated members, typically through working groups and committees,
- Establish monitoring and evaluation mechanisms to track EUCF performance against its objectives,
- Identify funding mechanisms and a business plan to ensure sustainability of the EUCF's activities and particularly its central coordinating body.

**S-F Recommendation 1.8: Evolve existing best practices and standards (e.g. ITIL, FitSM, etc.) for federated service management** to ensure federated cloud initiatives have reference requirements, processes, procedures and policies that ensure the compatibility of service delivery and planning across different initiatives. Develop "starter kits" to assist with implementation of each federated business model, with sample templates for required governance and service management processes (definitions, roles, process maps, etc.). [Deployment]

**S-F Challenge 2: Defining, Evolving, Selecting, Agreeing On and Managing the Architecture, Technical Standards and tools for Federated Clouds and distributed data access and exchange.** Creating a distributed yet federated, technically effective data-processing system is an active subject of research -- many technical approaches are being studied, and many technical approaches are in use in the community, but they are not converging into "standards" because the underlying technologies are rapidly evolving and because the scope of integration is expanding from the data centre out to the more heterogeneous edge computing environment. Where distributed capabilities need to work together, there must nevertheless be an agreement on the framework of the system (the

architecture) and the standards to be adopted within that framework, even in the face of this rapid change. Establishing a platform architecture enables technical discussions to be modularized and compartmentalized and facilitates agreement on standards that enable specific services to interoperate.

**S-F Recommendation 2.1: Develop and evolve a Federated Cloud Reference Architecture (FCRA).** To the extent possible incorporate the NIST CFRA, EGI and Gaia-X's technical architectures, and evolve it to ensure conformance of emerging federated cloud initiatives. This should specifically characterize how practical compliance frameworks and service catalogues would align with the EUSD's contemplated "Cloud Rulebook" and "Service Marketplace" concepts. [Deployment]

**S-F Recommendation 2.2: Create and maintain a federated cloud interoperability framework as an evolving suite of technology, standards and tools that are consistent with the FCRA allowing interoperation within a given federation and across multiple federations, compliance with European values and identify components that are interoperable.** This suite of components would help EU customers navigate the many options for cloud-based solutions and would help formalize how they are described and the possibilities for integration. This recommendation is similar to "product labelling" recommendations in other industries, making it clear what is included in a given component, how it works, how it works with other components, and any limitations on performance.

**S-F Recommendation 2.3: Coordinate research and innovation activities for funding in Horizon Europe by aligning cross-domain cross-use case research and innovation activities of common interest for different federation stakeholders to increase synergies, innovation potential and avoid duplication across the industry, research and public administration sectors.** This recommendation is meant to increase innovation capacity of the EU programmes by ensuring that EU funded research and innovation has a large potential of adoption by multiple stakeholders across different sectors. [Research]

**S-F Challenge 3: Federated data has great potential to support secure, private sharing of data held by many different organizations.** Best practice roadmaps are urgently needed to ensure federated data sharing initiatives are established and operated efficiently while preserving and ensuring the highest level of trust that affected sensitive data will be kept private and secure.

**S-F Recommendation 3.1: Guidelines for implementing different data sharing approaches using federated data platforms.** The best practice roadmap for federated data sharing should include specific guidelines for using federated platforms to enable the different forms of operational approaches to be easily adopted by data sharing communities.

**S-F Recommendation 3.2: Efforts to increase semantic interoperability for data within and across sectors are critical and must also include harmonization of data usage models to enable automated, yet secure and appropriate, data sharing.** Existing data usage models should be systematically analysed in hopes of finding common philosophies and features that can form the basis for a common, cross-sector usage model that could facilitate greater levels of properly governed data sharing.

**S-F Recommendation 3.3: Develop regulatory sandboxes.** These sandboxes would allow experimentation and scaled-up testing of privacy-preserving technologies, while ideally exempting the enterprises that need data from the responsibility to prove that they have all the necessary security measures in accordance with the legal precepts.

**S-F Recommendation 3.4: Continued support for research, innovation and deployment of privacy-preserving technologies.** Both the BDVA and e-SIDES reports enumerate a number of recent or ongoing projects to develop, refine and deploy these technologies in practical application domains. While these technologies are at various technological readiness levels (TRL), and not mature enough for "production" deployment, they are nevertheless

promising and would benefit from continued investment and support for early stage adoption and deployment.

**S-F Recommendation 3.5: Support and contribution to the formation of technical standards for preserving privacy.** Such standards would provide risk assessment tools, test suites for validation of performance, as well as evaluation of data for sensitive content.

**S-F Recommendation 3.6: Continued support for research, innovation and deployment of** distributed data analytics tools, as well as data placement tools, that minimize security privacy risks and maximize speed, computational and network efficiency as well as energy efficiency.