# D6.8: Final EOSC Architecture

| | |
|---|---|
| Author(s) | Geneviève Romier, Eric Fede (CNRS) |
| Status | **Final** |
| Version | V2.4 |
| Date | 20/05/2019 |

Dissemination Level

| | |
|---|---|
| X | PU: Public |
| | PP: Restricted to other program participants (including the Commission) |
| | RE: Restricted to a group specified by the consortium (including the Commission) |
| | CO: Confidential, only for members of the consortium (including the Commission) |


Abstract

The objective of this deliverable is to refine the architectural design for the interoperation of various types of infrastructures, which could participate to the future EOSC. This architectural design was initially proposed in deliverable D6.2 "*EOSC architecture Design and Validation Procedure*". It also updates parts of the gap analysis presented in deliverable D6.1 "*e-Infrastructure Gap Analysis*". The final part of this deliverable provides recommendations to support the interoperability of the infrastructure. This "*Final EOSC architecture*" is complementary to deliverable D5.4 "*Final EOSC Service Architecture*".



The European Open Science Cloud for Research pilot project (EOSCpilot) is funded by the European Commission, DG Research & Innovation under contract no. 739563

| Document identifier: EOSCpilot -WP6-D6.8 | |
|---|---|
| Deliverable lead | **CNRS** |
| Related work package | **WP6** |
| Author(s) | Geneviève Romier, Eric Fede (CNRS)<br>Xavier Jeannin (Renater) |
| Contributor(s) | Afrodite Sevasti (GRNET)<br>Licia Florio, Christos Kanellopoulos (GÉANT)<br>Narayanan Krishnan, Brian Matthews (UKRI-STFC)<br>Volker Beckmann (CNRS) |
| Due date | **31/12/2018** |
| Actual submission date | **20/05/2019 (revised version 2.4)** |
| Reviewed by | Vincent Breton (CNRS)<br>Tiziana Ferrari (EGI)<br>Peter Wittenburg (RDA) – External Reviewer |
| Approved by | **Mark Thorley (UKRI)** |
| Start date of Project | **01/01/2017** |
| Duration | **28 months** |

## Versioning and contribution history

| Version | Date | Authors | Notes |
|---|---|---|---|
| **0.1** | **01/11/2018** | Geneviève Romier, Eric Fede (CNRS)<br>Xavier Jeannin (Renater) | First version |
| | | Afrodite Sevasti (GRNET) | Network Orchestration |
| | | Licia Florio, Christos Kanellopoulos (GÉANT) | AARC project results on AAI |
| | **15/11/2018** | Geneviève Romier, Eric Fede (CNRS) | Version draft 1.7 circulated on the WP6 mailing list |
| | **3/12/2018** | Narayanan Krishnan, Brian Matthews (UKRI-STFC) | Interoperability in Rules of Participation in WP2 and interoperability verifications |

| | | | |
|---|---|---|---|
| | **3/12/2018** | Geneviève Romier, Eric Fede (CNRS) | Version draft 2 circulated on the WP6 mailing list |
| | **6/12/2018** | Volker Beckmann (CNRS) | Updates for clarification |
| | **10/12/2018** | Geneviève Romier, Eric Fede (CNRS) | New version including comments from Nick Juty (The University of Manchester), Prof Keith G Jeffery (NERC), Brian Matthews and Narayanan Krishnan (UKRI-STFC), Ghita Rahal (CNRS) |
| **1.0** | **13/12/2018** | Geneviève Romier, Eric Fede (CNRS) | Version 1.0 ready for reviewing |
| **2.0** | **21/01/2019** | Geneviève Romier, Eric Fede (CNRS) | Including modifications and additions according to Vincent Breton's review |
| **2.2** | **06/02/19** | Geneviève Romier, Eric Fede (CNRS) | Including modifications and additions according to Tiziana Ferrari's review |
| **2.2.1** | **07/02/2019** | Volker Beckmann (CNRS) | Editorial work, proof reading |
| **2.3** | **21/02/2019** | Mark Thorley (UKRI) | Final typographic proof-read and edit |
| **2.4** | 02/04/2019 | Geneviève Romier, Eric Fede (CNRS) | Updates following review by External Reviewer |

# TABLE OF CONTENTS

# LIST OF FIGURES

## EXECUTIVE SUMMARY

This deliverable (D6.8), as a final document of the task T6.1, presents the final version of the infrastructure interoperability architecture based on the previous D6.1 (gap analysis) and D6.2 (preliminary infrastructure architecture) deliverables. Because this deliverable reports on the infrastructure interoperability architecture, essential issues linked to the FAIR approach such as a lack of explicit data organisation, metadata interoperability and semantic interoperability are not covered. These data interoperability issues are discussed in deliverable D6.9: Final report on Data Interoperability[1].


**Figure 1: Main gaps identified**

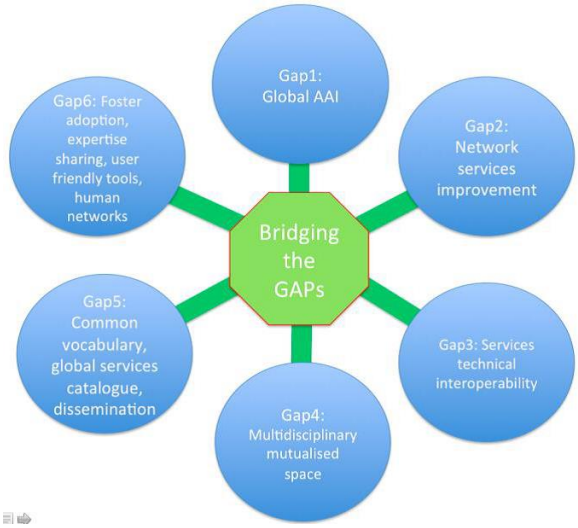


**Figure 2: Main bridges to be built**

Six sets of gaps preventing exploitation and usage of existing e-infrastructures and distributed resources were identified in D6.1. These, together with the main bridges to be built, are presented in Figure 1 and Figure 2.

This study on gaps and bridges, initially identified for infrastructures, has been completed in D6.8 using the project developments, results and issues that have occurred during the lifespan of the EOSCpilot project and the updates of questions, proposals, and approaches mentioned in both previous documents. The gaps have been detailed with the associated risks for infrastructures interoperability, and summarised in Figure 3.

[1] https://www.eoscpilot.eu/content/d69-final-report-data-interoperability

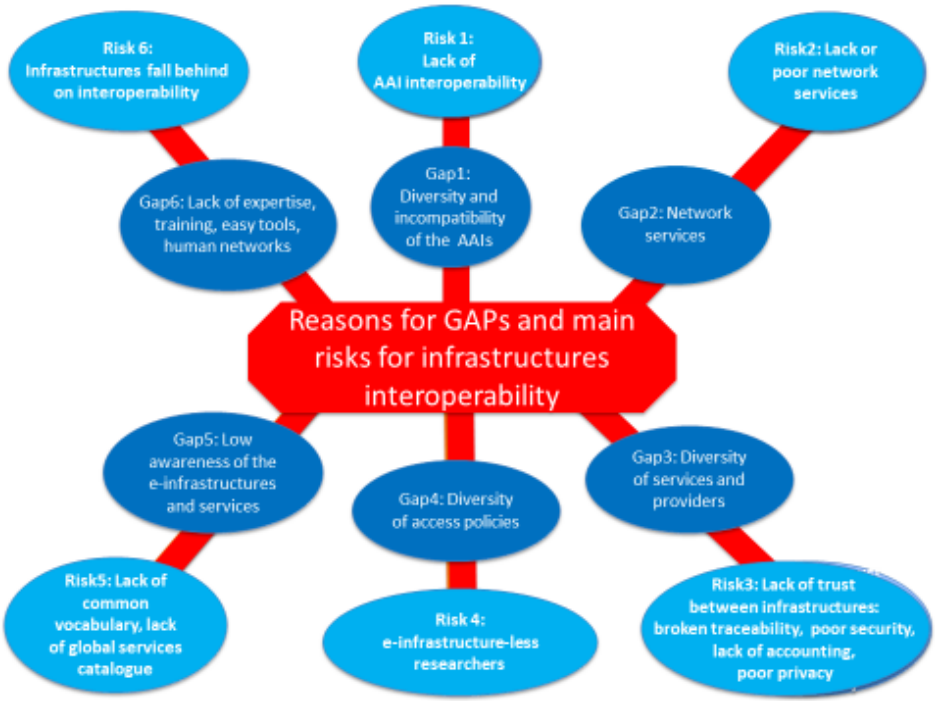


**Figure 3: Main risks by identified gap for infrastructure interoperability**

We have developed answers for the goal of "Ensuring interoperability" that are organised in three sections:

- "Technical aspects", which are elements identified as important to ensure the infrastructure interoperability,
- "Infrastructure interoperability in the rules of participation", and,
- "How to verify infrastructure Inter-operability".

We complete this document with six sets of "Recommendations" to ensure infrastructures interoperability. Each set of recommendations addresses one of the six identified gaps as shown in Figure 4:

**Recommendation 1**: *Implementation of a federated AAI[2]: adopt the AARC Blueprint Architecture and follow the associated guidelines.*

**Recommendation 2**: *Foster the network collaboration between EOSC and the NRENs in order to anticipate new capacity requirements.*

**Recommendation 3A:** *Infrastructures should describe their services in a standardised way, such as by using the eInfraCentral template. This should include services that act as service components in the provisioning of higher-level services.*

**Recommendation 3B:** *Infrastructures should identify their services' interoperability dependencies and to correctly describe these interdependencies.*

**Recommendation 3C:** *Infrastructures should be able to provide complete accounting data, to take care of the users' privacy in a GDPR compliant way, to provide traceability and to collaborate actively or at least follow the guidelines of all federative groups that work on the EOSC global infrastructure (AAI, security, privacy, global traceability, global information system…)*

***Recommendation 3D**: The EOSC itself should organise security channels to ensure security and privacy at EOSC level, and ensure global traceability, and to set up an EOSC information system.*

[2] AAI – Authentication and Authorisation Infrastructure.

**Recommendation 4A**: *In order to support access to the EOSC for groups of researchers that, because of access policies, cannot use any other infrastructure, EOSC should build a mutualised space (agnostic to discipline and geography) in order to serve all researchers, from all disciplines, in all countries, and to provide them with services that cannot be fulfilled by other means.*

**Recommendation 4B**: *The infrastructures that the EOSC is composed of should consider how to harmonise access and usage policies, in order to minimise the conditions that users need to accept to be able to access resources within the infrastructure, thus encouraging interoperability and reuse across the participating providers.*

**Recommendation 4C**: *EOSC should propose incentives to support funding agencies in making the resources they fund more openly available. EOSC should work with the infrastructures accessible through excellence-based applications to explore how to facilitate the interoperability between these infrastructures and EOSC for the benefit of all users.*

**Recommendation 4D:** *Infrastructures should define and publish applicable Service Terms of Use including acceptable usage policies in the EOSC Service catalogue as proposed by eInfraCentral. Machine readable-licences in interoperable formats are encouraged, to allow automatic brokering and access services across infrastructures to be supplied.*

**Recommendation 5A**: *The EOSC should promote the setup of a common vocabulary. Infrastructures should publish their services in a catalogue such as the EOSC-hub catalogue that is foreseen to be the first step of the EOSC Service catalogue*.
This recommendation is related to and reinforces recommendation 3A.

**Recommendation 5B**: *Disseminate and widely promote the EOSC at all levels of the European Research Area, including European infrastructures, national organisations, funding agencies and research laboratories, in order to improve the global knowledge of the computing landscape.*

**Recommendation 6A**: *Infrastructures should use the "EOSC Skills and Capability Framework" (D7.3)[3]. This describes the framework that can be used to identify what skills are required of individuals and organisations wishing to foster interoperability.*

**Recommendation 6B**: *Foster networking activities to share expertise and experience among infrastructure technical staff in order to create the conditions for technical experts to find and mutualise the right way to support their users and to interoperate. Foster the development and use of portals across different systems, and of comparable user-friendly tools that allow users to easily deploy their data analysis pipelines. Be careful to provide users with enough skilled technical staff to guide them in their use of the EOSC.*

Figure 4 gives an overview of our main recommendations.

[3] https://www.eoscpilot.eu/content/d73-skills-and-capability-framework

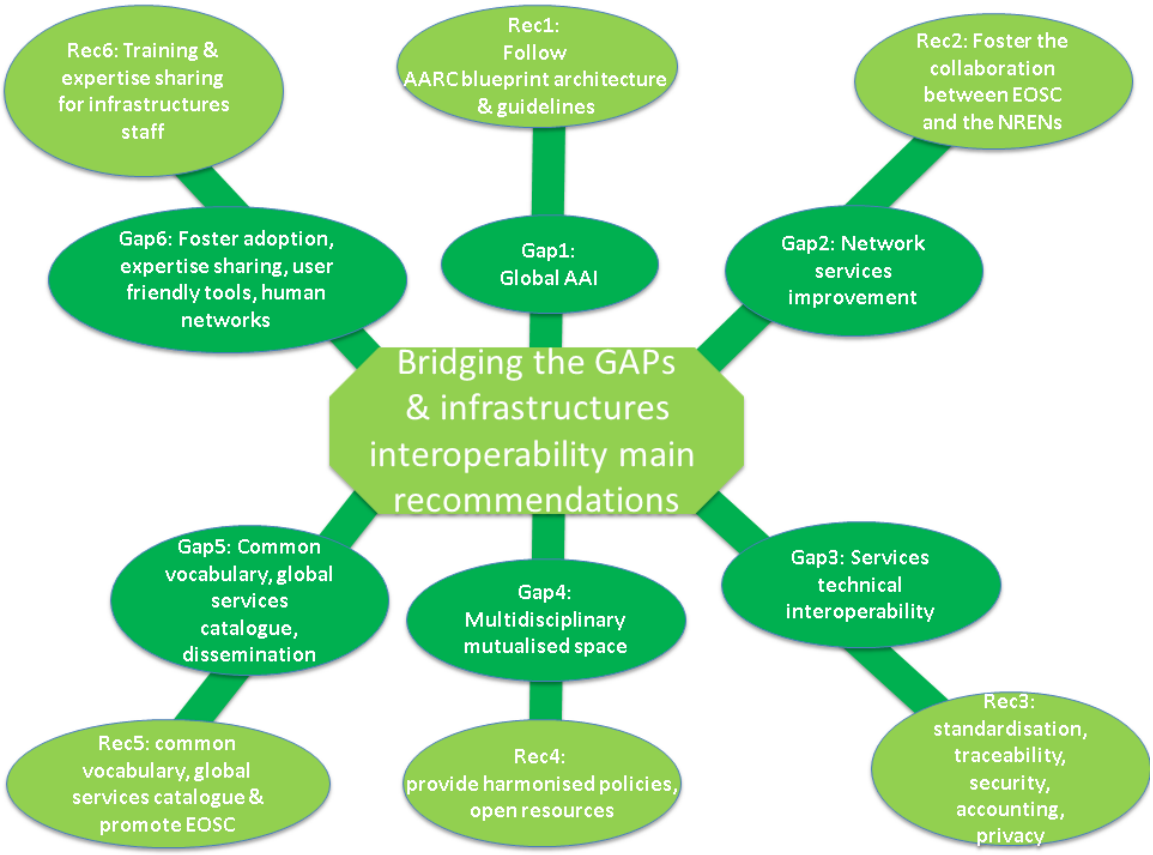


**Figure 4: Main recommendations for infrastructure interoperability**

# 1 INTRODUCTION

The European Open Science Cloud (EOSC) will federate IT infrastructures and will be a fundamental component of Europe's future research infrastructure and Horizon2020/HorizonEurope actions. The EOSCpilot project brings together a broad range of stakeholders from Research Infrastructures (RIs) to e-infrastructure services and e-infrastructure providers. The WP6 work package of the EOSCpilot project addresses the issue of interoperability, and in this context recommends solutions to interconnect various types of infrastructures provided by the many actors within European research. Interoperability of infrastructures covers many aspects in order to accomplish the challenges of an efficient construction of deployment processes and mechanisms.

The goal of this deliverable (D6.8) is to propose an interoperability architecture for the EOSC. It updates and completes the "e-Infrastructure Gap Analysis"[4] (D6.1) and the "EOSC architecture design and validation procedure"[5] (D6.2).

The content of the different parts of the document are classified by the gaps already defined in D6.1. In order to keep the deliverable short and easy to read, previous and external work will be cited and completed if necessary.

In addition, our approach is to reference throughout the text other documents and milestones and to avoid rephrasing them. This is especially relevant to the work carried out so far in task T6.1. Thus, we recommend reading this final EOSC architecture document after reading the executive summaries of D6.1[2] and D6.2[3].

The structure of this document is as follows:

1. The first section introduces the document and describes its structure.
2. The second section presents the Context and Objectives, by:
    a. Setting the context of the work already done in the WP6 work package notably task T6.1;
    b. Giving the scope and the limits of the study relatively to the work done inside EOSCpilot work packages.
3. The third section is dedicated to the risks and potential mitigations related to e-infrastructure interoperability.

    The gaps identified during the initial phase of the project (deliverable D6.1) are updated if necessary, though in cases where no update has been necessary only a reference to the previous document is provided. This section presents new input from the infrastructures that was gathered with the help of a questionnaire (see ANNEX A), sent to the same recipients as the first survey conducted to prepare the gap analysis (deliverable D6.1), and also, via work package 4, to the Science Demonstrators. The lessons learnt from the Science Demonstrators gathered from their reports and the test beds of task T6.3 are discussed, as well as input from the other EOSCpilot work packages and external contributions to e-infrastructure interoperability, which address the gaps identified in D6.1. To be as readable as possible this section is organised according to the identified gaps.
4. The fourth section covers "Ensuring interoperability", which was introduced in section 5 of D6.2. In this deliverable we complete the work with section 4.1 looking at elements identified as important to ensure the infrastructure interoperability; section 4.2 considers the findings that work package 2 expressed in deliverable D2.5 "Recommendations for a minimal set of Rules of Participation"; And, section 4.3 tackles the question of "How to verify infrastructure interoperability?". It provides a list of recipes or procedures to perform, and also a set of good practices to be adopted. In keeping with the previous section, this one is also organised along the identified gaps.
5. The fifth section provides recommendations, organised by the gaps and bridges identified in D6.1.

[4] https://www.eoscpilot.eu/content/d61-e-infrastructure-gap-analysis
[5] https://www.eoscpilot.eu/content/d62-eosc-architecture-design-and-validation-procedure

6. The final section presents the conclusions of our work.
7. Annex A includes the questionnaire that sent to e-infrastructures and Science Demonstrators.

As identified in the first gap analysis (D6.1) a common vocabulary is required in order to bridge gap 5, and this was reinforced by findings elsewhere in the project. We are pleased to use the first version of the *EOSCpilot Glossary* proposed by WP5 as mentioned in the introduction (see Annex B).

# 2 CONTEXT AND OBJECTIVES

This deliverable "Final EOSC architecture" (D6.8) is the main deliverable of task T6.1 of the "EOSC interoperability" work package (WP6), and it deals with e-infrastructure interoperability. The scope of the task was to perform an e-infrastructure gap analysis and to propose an e-infrastructure interoperability architecture. This deliverable updates the previous "EOSC architecture design and validation procedure" (D6.2). It also updates the findings of the “e-infrastructure Gap analysis” (D6.1) where new inputs or new solutions are identified.

The focus of the “e-infrastructures Gap Analysis” (D6.1) was to perform a gap analysis of the current issues preventing exploitation and usage of existing e-infrastructures and distributed resources, on technical and political barriers that prevent the interconnection of e-infrastructures, and on the application of the FAIR principles, with the aim to provide architecture that will allow overcoming these gaps. Six main gaps were identified during the gap analysis and the main findings were classified according to these gaps. The main findings are summarised in section 4.4 of D6.1. Six main bridges were proposed to fill the six gaps. For convenience, we reproduce here the figures from D6.1 describing the main gaps identified and the main related bridges to be built in order to achieve e-infrastructure interoperability in the EOSC. As far as possible, the content of this document is organised according to the same gap classification.

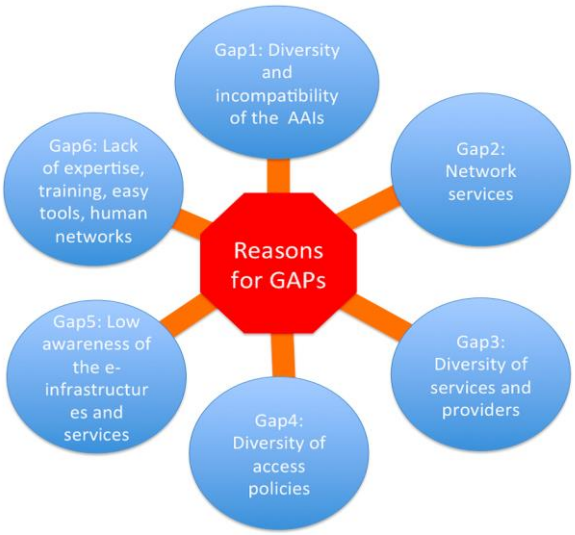


**Figure 5: Main gaps identified**


**Figure 6: Main bridges to be built**

The objectives of “EOSC Architecture Design and Validation Procedure” (D6.2) were to describe the framework that needs to be put in place to allow the interoperability between the e-infrastructures and the Research Infrastructures (RIs) involved in the EOSC project, and to propose a first version of the infrastructure interoperability validation procedure.

It should be noted that the “EOSC architecture” as defined in the EOSCpilot glossary (service-oriented) is a product of work package 5 and will be described in deliverable D5.4 ("Final EOSC Service Architecture[6]").

EOSC Research and Data interoperability is the topic of task T6.2. Deliverables produced by this task, D6.3, “First report on Data Interoperability: Findability and Interoperability[7]”, and D6.6, “Second report on Data Interoperability: Accessibility and Reusability[8]”, are already available. This series of reports will be completed by deliverable D6.9 "Final report on data interoperability[9]".

[6] https://www.eoscpilot.eu/content/d54-final-eosc-service-architecture
[7] https://www.eoscpilot.eu/content/d63-1st-report-data-interoperability-findability-and-interoperability
[8] https://www.eoscpilot.eu/content/d66-2nd-report-data-interoperability
[9] https://www.eoscpilot.eu/content/d69-final-report-data-interoperability

However, in the present document we also reference the results obtained by the other work packages such as WP2 (Governance), WP3 (Policy), WP4 (Science Demonstrators), and WP7 (Skills) when there is a link with the bridges identified and recommendations proposed. In order to keep D6.8 concise, previous and external work will be cited and completed if necessary.

The main objectives of the present document are to complete the work presented in D6.1 and D6.2. It defines the architectural design of the interoperation of different types of infrastructures, taking into account the risks related to e-infrastructure interoperability and their potential mitigations. It proposes elements to ensure and, where possible, verify infrastructure interoperability. It also proposes sets of recommendations, both short and long term, to facilitate, set up, and maintain infrastructure interoperability.

# 3 RISKS AND POTENTIAL MITIGATIONS RELATED TO E-INFRASTRUCTURE INTEROPERABILITY

This section is an update of the first two deliverables (D6.1 and D6.2) of task T6.1

## 3.1 New inputs from e-infrastructures

A questionnaire was sent to e-infrastructures (list available in D6.1) during the early phase of the EOSCpilot project to identify the issues preventing them exploiting and using existing infrastructures, and also to gather their best interoperability practices. The answers to this questionnaire were then used as input to the infrastructure gap analysis (D6.1). Before finalising the proposal for an EOSC architecture to ensure e-infrastructure interoperability, and more than 18 months after the first survey, it was important to find out from the e-infrastructures whether the interoperability landscape had changed. A new set of questions was developed to target this (see Annex).

Unfortunately, only three responses were received, however they confirmed that the identified gaps are still seen as relevant and that D6.1 did not miss any important points. However, a concern was raised that "The proposed bridges are mainly technical and the sociological and political gaps are not really treated."

## 3.2 Lessons learnt from Science Demonstrators

The risks and potential mitigations related to e-infrastructure interoperability expressed by the Science Demonstrators (SD) that were set up during the first 18 months of the project can now be taken into account - unlike in the first version of the deliverable (D6.2). This new input is presented by 'gap', as a listing by Science Demonstrator would not be an efficient way to show this information.

The inputs from task T6.3 are also taken into account in this section. These inputs are related to the "Pilot for distributed authorization and authentication" and the"WLCG-AARC2-EOSCpilot". The final outcomes of the Science Demonstrators are reported in deliverable D4.4[10].

"Preliminary general recommendations" were collected from Science Demonstrators during the First Stakeholder event on Nov 28, 2017 in Brussels (see D4.3, page 11):

*According to the concerns raised by the Science Demonstrators, the following areas need to be addressed:*

- *The Identity provisioning in a federated EOSC model,*
- *The need for incentives for scientific communities to adopt interoperable standards,*
- *The need for support of high-bandwidth data connections between data centres,*
- *The need for the integration of High-Throughput Computing (HTC) computing resources with large data storage resources,*
- *A way for the EOSC to operate with different underlying data models and strategies to achieve interoperability without collapsing data models,*
- *The need for cross-discipline data interoperability,*
- *The trade-off between performance and resources in a cloud computing model,*
- *The need for incentives to make better progress towards the implementation of the FAIR principles,*
- *The discrepancy between long-term, multi-decade commitments in ESFRIs and other projects and short life-times of EC funded projects: According to a consensus opinion of Science Demonstrators the benefits through Open Science and Open Data as a significant multiplier for knowledge creation through a large number of additional scientists being enabled to work on these data by far outweighs the risk of potential misinterpretations or misuse of Open Data.*

[10] https://www.eoscpilot.eu/content/d44-consolidated-science-demonstrator-evaluation-report. D4.4 was not available during the writing of this deliverable.

*Initial concrete recommendations include:*

- *Progress towards better interoperability shall not be made by the standardization of the well-established community specific tools, procedures and formats, but by the work on interfaces that achieve compatibility through conversions.*
- *The mechanism by which hardware resources for general usage shall be provided has to be clarified, since the established community-specific e-infrastructures cannot be expected to be opened for general free use on behalf of their respective budgets.*
- *Addressing issues with copyright laws since they can block access to data for open use.*
- *Strategies and concepts for long-term (multi-decade) data preservation need to be developed.*

In the following sections, we cite the different documents studied and the main citations relating to the gaps.

### 3.2.1 GAP 1: Diversity and incompatibility of the AAIs

D4.3 Virtual Earthquake and Computational Earth Science e-science environment in Europe (EPOS/VERCE) Science Demonstrator Report page 33, "Missing functionality or services"

They indicate that the following service would be needed: "Liferay EGI Check-in Integration (Registration / confirmation of account, RCAuth.eu online CA for X509 certificate)."

D4.3 PROMINENCE Demonstrator Report page 39, "Missing functionality or services"

"*The use of Virtual Organization (VO)*[11] *for accessing the EGI Federated Cloud*[12] *will be a significant impediment to take up. Fusion used to have a VO but many external users considered the registration process too onerous and would not likely take up a service requiring this*". The issue is not really about interoperability itself but about the use of tools mandatory to be a member of a VO. One can note that a VO is a means to allow a user to use a large set of services and the underlying resources.

### 3.2.2 GAP 3: Diversity of services and providers

In its answer to our new questionnaire, the "VisuaMedia service" Science Demonstrator[13] describes elements to bridge this gap. It explains that "*our VisualMedia service could be considered as an extension to the services already provided by EOSC, filling a gap (web-based publishing and visualization of complex visual media files, to support cooperative and remote data analysis)*".

### 3.2.3 GAP 6: Lack of expertise, training, easy tools, and human networks

D4.3 "*Leveraging EOSC to offload updating and standardizing life sciences datasets and to improve studies reproducibility, reusability and interoperability*" Science Demonstrator Report page 45 *"Missing functionality or services"*:

This states the need for a *"Catalogue of security requirements across resource providers"*. The issues encountered were: "- *Ability of resource providers to provide security levels is unclear – the SD suggested a portfolio of security possibilities for each resource provider would be helpful - Size of computational resources needed is not yet clear ➔ exact requirements will be determined and fed back to shepherd*".

## 3.3 Contributions from other EOSCpilot work packages

Because D3.6 "Final Policy recommendation[14]" was not available at the time of writing this report, the "Draft Policy Recommendations" produced by WP3 is referenced as [D3.3][15] and used in the context of this section.

[11] https://wiki.egi.eu/wiki/Glossary_V1#Virtual_Organisation
[12] https://www.egi.eu/federation/egi-federated-cloud
[13] https://eoscpilot.eu/social-sciences-and-humanities-visualmedia-service-sharing-and-visualizing-visual-media-files-web
[14] This report has since been published, see: https://www.eoscpilot.eu/content/d36-final-policy-recommendations
[15] https://www.eoscpilot.eu/content/d33-draft-policy-recommendations

### 3.3.1 GAP 1: Diversity and incompatibility of AAIs

[D3.3], Open Science: 1. Adopt the AARC (*Authentication* and Authorisation for Research and Collaboration) framework for enabling an interoperable AAI infrastructure is clearly targeting this gap.

### 3.3.2 GAP 2: Network Services

#### *3.3.2.1 Connectivity:*

Taking into consideration the EOSC user requirements concerning data access and manipulation, networks are fundamental for the EOSC infrastructure. It has been demonstrated that the internet provides much lower network quality in terms of performance and reliability than dedicated research and education networks. The network EOSC user traffic will be concentrated between the academic data centres that support the EOSC, between academic data centres and the EOSC users, and between academic data centres and commercial clouds that users may use. National research and education networks (NRENs) and the pan-European GÉANT network, are vital in this context.

Research and educational networks (NRENs and GÉANT) offer the best route between academic data centres and enforce an overprovisioning capacity policy whereas internet connection offers much less capacity and potentially poor interconnection depending on commercial agreement.

The connectivity between the EOSC academic data centres will be optimal in terms of route thanks to the NREN. We recommend fostering the collaboration between the EOSC and the NRENs in order to anticipate new capacity requirements.

The NRENs are able to provide Virtual Private Networks (VPNs), which can, in certain case especially if the data exchange is important, improve dramatically the data exchange performance and reduce EOSC data centres sites security capital expenditure. One problem identified for interoperability between sites is the security policy level. Certain data centres have a very strict security policy and they allow data exchange only through a DMZ (De-Militarized Zone), which can lead to a long delay to give access to remote users to the data stored in these data centres. A VPN allows data centres exchanging with the assurance that the site sending them their data is a member of the EOSC, thus, a VPN usage can be an appropriate answer to this security policy problem. Nevertheless, as it is in first place a security policy issue, this problem has to be addressed by the EOSC security policy. Finally, if the European data centres, which are members of the EOSC, identify the requirement to create a federated infrastructure, NRENs will be able to interconnect them thanks to an appropriate network service.

The NRENs and GÉANT work also in order to offer an optimal connectivity to commercial cloud infrastructures. GÉANT developed a specific service called “GÉANT Cloud VRF” and NRENs try to have the best connectivity with commercial cloud providers for instance by setting up private peering. This could be reinforced if the EOSC requires it.

The connectivity between academic data centres and the user sites is also an important point and the “last mile” connectivity is already identified as a frequent source of network trouble by many academic data centres. Solving this problem requires identifying and highlighting which part of the network is the root of the problem. Thus, this problem can be addressed by deploying a monitoring solution.

#### *3.3.2.2 Reliability of the network:*

One of the big advantages of NREN networks in comparison with the internet is the reliability of connectivity provided. The EOSC will be a complex federation of “resources” delivering data storage and computing services. Data remote access requires a both powerful and reliable network. The success of the EOSC is also dependent on the level of confidence that the user will have in this federated environment. Network monitoring is necessary at operational level and at business level (Service Level Agreement). The monitoring solution inside each NREN depends on which type of connectivity will be used but the monitoring solution could be deployed in the EOSC site like in LHCONE project using PerfSONAR. This solution is relatively widely deployed in NRENs too; this will facilitate the easy identification the root of potential problems or of performance issues.

#### *3.3.2.3 Orchestration, network and cloud:*

Providing an orchestration solution for e-infrastructure interoperability for the orchestrated delivery of network and cloud services requires a multi-dimensional approach. The e-infrastructures need to adhere to a common basic set of information-modelling principles and entities, functional elements, process definitions and inter-provider application programming interfaces (APIs) and at the same time share product specifications that map to their internal implementations of product-underpinning services. The GÉANT orchestration framework addresses all of these aspects, introducing:

- Minimum requirements for the operations and business support systems of participating service providers.
- Orchestrated processes across service providers (e.g. order qualification, order fulfilment, service activation).
- Inter-service-provider APIs for the different types of service provider interactions (e.g. business agreement establishment, order management, service delivery).
- Coordinated order management, by exposing one-stop shop capabilities to the user. The orchestration framework is designed in such a way that any participating domain/service provider in the delivery chain can trigger an orchestration instance or undertake the orchestration role. Equally, it does not limit future possibilities for distributed orchestration.

The approach is heavily influenced by and largely compliant with relevant standards and best-practices, namely the TM Forum (TMF) Information Framework (SID) and Open APIs as well as the Metro Ethernet Forum (MEF) Lifecycle Service Orchestration specifications, which ties in with the on-going collaboration between TMF and MEF for convergence of specifications for service orchestration across multiple providers.

The orchestration framework introduces the following significant benefits:

- Modelling and advertising of offerings/services programmatically (e.g. via YAML, XML, JSON files).
- No need for exchanging spreadsheets/service definition documents offline.
- Managing business agreements and terms of service use programmatically.
- Enabling service chaining and composition.
- Incorporating federated AAI functions as part of the orchestrated workflows.
- Accommodating EOSC but also commercial service providers' existing APIs.
- Create wrappers compliant to the framework's APIs.
- Dynamic on boarding of service providers and users.
- Any provider that is able to expose the APIs can be part of the service delivery chain.
- Eliminating manual tasks –where possible.

One significant use case for the framework, is the service delivery of dedicated (e.g. L2 VPN) network connectivity between an NREN-connected institution/researcher facility and a computing/storage/cloud resources EOSC provider via GÉANT and other intermediate network providers (see Figure 7).

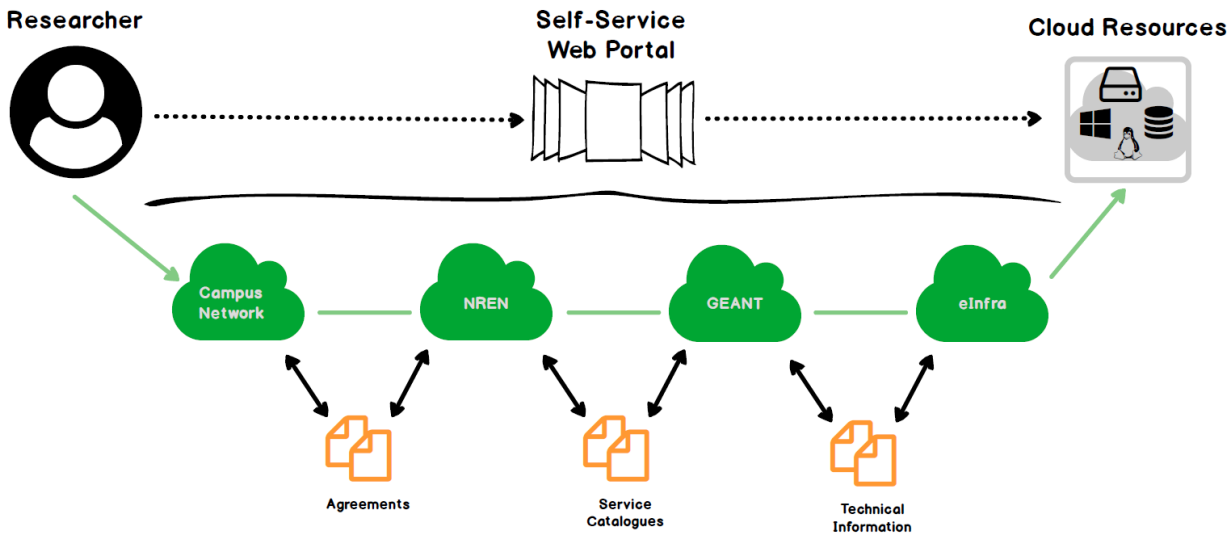


**Figure 7: Network connectivity network providers**

Using orchestration, the previously manual processes for establishment of dedicated secure VPN connectivity with multiple parties involved (institution NOC, NREN, GÉANT, EOSC compute/storage provider) can be significantly expedited. Using orchestrated end to end VPN establishment (already available from / offered by the participating network service providers i.e. individual NRENs and GÉANT) and invoking appropriate APIs of the EOSC providers' IaaS/PaaS solutions, it is possible to provide to EOSC users end-to-end service delivery in much reduced timescales (order of minutes).

The ultimate goal is to streamline service delivery, by minimizing lead times as much as possible, providing to the user:

- A one-stop-shop experience for end-to-end cloud connectivity.
- Visibility and transparency regarding the state of their service orders as they are handled by all the providers involved.

The orchestration framework requires the participating service providers to expose a minimum set of standards-based APIs as follows (Figure 8):

| Context | Description | API |
|---|---|---|
| **Party** management (partners, institutions, suppliers) | Capabilities usually found in an organization's CRM | Party management API |
| **Agreement** management | Managing information about business agreements in a structured format | Agreement management API |
| **Service** catalogue | Managing information about customer and resource facing service offerings (to be shared with partners), usually found in an ITSM system | Service catalogue API |
| **Product** catalogue | Information about product offerings-to be shared with customers/users, usually found in an ITSM system | Product catalogue API |
| **Order** management | Managing information about orders, usually found in an ITSM system | Order management API |

**Figure 8: Service providers' standards-based application programming interfaces (APIs)**

In special cases, Configuration & Activation and Resource Inventory APIs are also needed for end-to-end service delivery orchestration.

### 3.3.3 GAP 3: Diversity of services and providers

[D3.3] Open Science. Adopt a minimum metadata schema and limited number of APIs to be considered as standard for services, infrastructures and other resources in the EOSC Service Catalogue. This point is a good step to foster the services technical interoperability. However, evolutions have to be possible to avoid technical stagnation and to avoid barriers to new comers.

### 3.3.4 GAP 4: Diversity of access policies

WP3 in deliverable D3.1[16], page 20 (2.3 Focusing on Specific Policy Areas (a) Policies for infrastructures and services) the most relevant policies related to infrastructures are detailed.

### 3.3.5 GAP 6: Lack of expertise training, easy tools and human networks

WP3 in D3.1, page 21 (2.3 Focusing on Specific Policy Areas (c) Policies for improving skills and supporting the development of open educational resources) gives a list of policy initiatives that "aim at the introduction of the skills necessary for conducting open science as well as the introduction of open educational practices across all types and levels of education. "

[D3.3] Open Science 7 recommendation Develop, support and promote an EOSC Skills and Capability Framework as a common reference point targets this gap.

According to deliverable D7.2[17], page 12, "EOSCpilot WP7 focuses on the skills requirements of various professional groups that share responsibility for stewardship." The Annex A.1 of D7.2 gives details on necessary skills for these professional groups. No required skills related to infrastructures staff are mentioned in that document. But, the recommendations of WP7 are of great interest for the EOSC adoption by users and communities. These recommendations address the need to improve the skills of users identified in the D6.1 GAP 6.

## 3.4 Other contributions to infrastructure interoperability

Some other issues not identified in the previous section could contribute to infrastructures interoperability. Many can be considered as a set of “good practices”. Whilst some of them come from official and robust frameworks, some others are just common sense.

Following the same strategy of consistency with the rest of the document, these elements have been identified by gap. Nevertheless, some generic consideration from the European context can be provided.

The *New European Interoperability Framework is part of the Communication* (COM(2017)134)[18] has been adopted by the European Commission on 23 March 2017. The framework gives specific guidance on how to set up interoperable digital public services, summarised with the help of 47 concrete recommendations. As presented in D6.2 chapter 3, the EOSC takes this framework into account in designing its approach to interoperability and its validation.

The European Commission staff working document "Implementation Roadmap for the European Open Science Cloud" (SWD(2018) 83 final)[19]) was issued in March 2018, and the information in this document has been illustrated and updated with the help of presentations[20] in May 2018. A potential model is presented in chapter 2.1. Its architecture is presented as a "federation of existing and planned research data infrastructures, adding a soft overlay to connect them and making them operate as one seamless European research data infrastructure.... The EOSC federating core is understood to be constituted by EOSC shared

[16] https://www.eoscpilot.eu/content/d31-policy-landscape-review
[17] https://www.eoscpilot.eu/content/d72-interim-report-and-catalogue-eosc-skills-training-and-educational-materials
[18] European Interoperability Framework –Implementation Strategy: (COM(2017)134
[19] http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf
[20] http://ec.europa.eu/research/openscience/pdf/eosc_strategic_implementati...

resources and by a compliance framework including notably the Rules of Participation". In this report, inputs from the *New European Interoperability Framework* are referenced as [NEIF].

The **High Level Expert Group** published its "Prompting an EOSC in practice"[21], *Interim report of the EOSC High Level Expert Group* in June 2018. This report gives recommendations related to the EOSC implementation, engagement and steering. Infrastructures interoperability is not often directly mentioned in this report but underlying many recommendations. In this report inputs from the **High Level Expert Group** are referenced as [HLEG].

GÉANT, PRACE, EGI, EUDAT and OpenAIRE are members of the consortium of the eInfraCentral European project[22]. Their participation as large e-infrastructures ensures that the projects' results reflect this e-infrastructures point of view. eInfraCentral focuses on a service description template, a service catalogue and a portal (that gives access to a market place) that are positioned as the catalogue and the portal of the emerging EOSC. This is a first step to bridge important gaps identified in D6.1 such as GAP 3, GAP 4 and GAP 5. In this report, inputs from eInfraCentral are referenced as [eInfra].

In addition the *Authentication and Authorisation for Research and Collaboration* (AARC) project[23], referenced as [AARC], and the EOSC-hub project[24], referenced as [EOSCHub], produced valuable results to bridge the identified gaps.

### 3.4.1 GAP 1: Diversity and incompatibility of AAIs

[NEIF] The diversity and incompatibility of the AAIs of the current diverse infrastructures that provides the research communities with services that was clearly identified as a main gap in D6.1 are taken into account in this roadmap as the first of the five main types of services to be offered to the European researchers. *"A unique identification and authentication service and an access point and routing system towards the resources of the EOSC."*. The document states that "Work to integrate and federate such services has already began in Horizon 2020 Work Programme 2016-2017, with the EOSC-hub project and other related projects expected to deliver services under the EOSC."

[AARC] It is recognized that users should be able to access services that they need logging in once (via their institution and/or their preferred infrastructure) and access the services they need for their research. All the complexity should be transparent to the users.

There is consensus on deploying **federated access**, as it

- preserves privacy,
- it is secure, and
- it reduces the number of credentials that users are expected to have.

There has been significant investment to deploy it at large. The research and education community has been championing federated access for more than a decade now; initially building *eduroam*[25], a global infrastructure to enable federated access first to (wireless) network and then via eduGAIN[26]. eduGAIN, the infrastructure operated by GÉANT that enables the exchange of identity information between services providers and identity providers worldwide. eduGAIN builds trust between national R&E identity federations so users and services in different countries can interact. With more than 5000 entities (about 2900 institutions and 2100 services), eduGAIN enables users in one of the participating institutions to access all eduGAIN services.

[21] https://ec.europa.eu/info/sites/info/files/conferences/eosc_summit_2018/prompting_an_eosc_in_practice_eosc_hleg_interim_report.pdf
[22] http://einfracentral.eu/
[23] https://aarc-project.eu/
[24] https://www.eosc-hub.eu/
[25] https://www.eduroam.org/
[26] https://edugain.org/

The relationship between the users' home institutions and service providers, which is typically found in the national identity federations and eduGAIN, now becomes a relationship between a research community, the users' home institutions and service providers.

To address the challenges above, AARC has defined a BluePrint Architecture (BPA)[27] to guide research collaborations and e-infrastructure to build interoperable authentication and authorization infrastructures (AAIs) to support their own users. The AARC BPA defines a reference architecture for authentication and authorization infrastructures (AAI) that best fits the needs of international research collaborations and the EOSC vision. The AARC BPA builds on top of eduGAIN and adds the functionalities required to support common use cases within research collaborations, such as access to non-web services and access to resources based on community membership. The AARC BPA champions a proxy architecture in which services in research collaboration can connect to a single point, the proxy, which itself takes the responsibility for providing the connection to the identity federations in eduGAIN, thus reducing the need for each service to have to individually connect to a federation/eduGAIN.

To ease the deployment of the AARC BPA, AARC has provided guidelines, a set of technical and policy implementation guidelines, along with a policy development 'kit'. The guidelines help research collaborations choose the best way of identifying their users and keeping track of them when they move jobs through their career, whilst remaining associated with the collaboration. The guidelines also provide template policies to help research collaborations organise their community to access data, computing and network services without researchers being bothered with Terms and Conditions and tick-boxes all the time. AARC also defines guidelines to manage data protection and the sharing of attributes: the "Snctfi"[28] framework helps research AAI operators to deploy an AARC BPA compliant AAI that is also compliant with the *General Data Protection Regulation (GDPR)*.

The AARC project has validated the AARC BPA and policy frameworks via pilots with different research collaborations. The AARC BPA is also being deployed by the following research and e-infrastructures: EUDAT, EGI, GÉANT, Elixir, LIGO, DARIAH, XSEDE, Life science community.

AARC results are already proving to be a solid foundation for the EOSC.

The AARC project has created the *AARC engagement group for infrastructures* (AEGIS) with the aim to bring the AAI infrastructures' operators together to endorse and deploy the AARC reference blueprint architecture and the relevant guidelines in order to ensure the interoperability and alignment of all the AAI elements in EOSC and beyond. AEGIS has been endorsed recently in the collaboration agreement between the GÉANT (GN4) and EOSC-hub programs, in which the infrastructures involved in the delivery of EOSC have committed to support the works of AEGIS beyond the AARC project.

[EOSCHub] The EOSC-hub project works on AAI: As stated in its deliverable D5.1 *Initial maintenance and integration plan for federation and collaboration services*[29], "The integrated EOSC-hub AAI will build on existing AAI solutions that follow the architectural and policy recommendations defined in the AARC project". The integration plan is presented in this deliverable, and is a step towards a global AAI for infrastructures.

### 3.4.2 GAP 2: Network Services

[NEIF] GÉANT is listed in the Annex 2 – Actions from WP 2016-2017 of the Research Infrastructure Work Programme have paved the way for the establishment of the EOSC. This ensures that the network services improvement necessary to bridge the GAP 2 identified in D6.1 will be realised.

### 3.4.3 GAP 3: Diversity of services and providers

[NEIF] The diversity of services and providers needs technical interoperability services to be designed. This point is not detailed in the *Implementation Roadmap for the European Open Science Cloud* document. This is because the document takes the point of view of the user than the one of the infrastructures that have to

[27] https://aarc-project.eu/architecture/
[28] https://wiki.geant.org/display/AARC/Snctfi
[29] https://www.eosc-hub.eu/deliverable/initial-maintenance-and-integration-plan-federation-and-collaboration-services

connect their services together. However, it proposes to make available "an access point and routing system towards the resources of the EOSC", irrespective of discipline or national boundaries. This is more detailed in the chapter (d) Access and interface: "Work on the EOSC access and interface has already begun under Horizon 2020 Work Programme 2016-2017: the EOSC-hub project will pilot the common platform and the access to EOSC services, while the eInfraCentral project provides a first catalogue and access to e-Infrastructure services.

The entry points to the EOSC would be similar but not equivalent, and typically would consist of a web-based user interface, or front-end, which can be tailored to the specific needs and context of particular user communities. In addition, it would consist of a common platform building on the EOSC-hub project and further developed in the INFRAEOSC-06-2020 call a) and b), that would be accessible to users via machine-to-machine interfaces and which offers access to shared EOSC resources and to the full range of EOSC services".

Technical interoperability is evolving with the evolution of the services and the underlying technologies. This is a ongoing process that will continue with the help of these projects.

[HLEG] *"I2 – Define EOSC interoperability standards so that services can be interconnected and federated to be as effective as possible and be based on existing open standards; I7 – Promote the development of services as independent, interoperable and exchangeable building blocks to foster the future accreditation of innovative and/or efficient alternatives."*

[eInfra] The eInfraCentral Service Description Template guarantees that the services proposed by the different infrastructures are described in a similar way.

The current version (V1.12) of the *Service Description Template* [30] provides definitions of the service features/attributes organised in blocks, example values and their specific format, recommendations as well as whether the attribute is mandatory or optional for the implementation of a number of features in a service catalogue.

Unfortunately, from the point of view of infrastructures interoperability we are missing complementary information. For example:

Services are based on technologies that are compliant (or not) to standards. This important information is currently not included in the description template.

The service provider is referenced by its name as a free text described as "The organisation that manages and delivers the service and with whom the customer signs the SLA". Given examples are "e.g. GÉANT, PRACE, EGI, EUDAT, OpenAIRE, etc." This is useful but probably not accurate enough:

- A typing error in the name is possible;
- A link to a description of the provider could give more details related to the way the service is provided (which entities in a federation for example, in which countries, by academic or commercial or other types of providers, if this service is based on a procurement...);
- Depending of the catalogue functionalities, the search of the set of services provided by a provider will be easy or not. This point is important because e-infrastructures build generally a set of interoperable or compatible services. At least they propose a Single Sign On or a unique entry point to their set of services and often global related services (accounting, monitoring, documentation, user support...).

The eInfraCentral D2.2 deliverable "eInfraCentral in the context of the European Open Science Cloud" "highlights and elaborates on the strategic fit between the work done in eInfraCentral and the emerging EOSC.". It presents in a synthetic way the eInfraCentral products and ambitions. The adoption of

[30] https://github.com/eInfraCentral/docs/raw/master/eInfraCentral-JNP-ServiceDescriptionTemplate.pdf or https://jnp.gitbooks.io/service-description-template-v1-12/

such products could be a first step to the services interoperability. The Service Description Template could be broadly discussed in a wider and open framework (e.g. The Research Data Alliance).

### 3.4.4 GAP 4: Diversity of access policies

[NEIF] Despite the vision that:

*"The EOSC will allow for universal access to data and a new level playing field for EU researchers"* with *"Easy access through a universal access point for ALL European researchers"* and the implementation of a universal entry point *"Acting as a universal entry point for all potential users, the portal would have a full-fledged user interface supported by the common platform. A universal entry point usually guarantees that all users have access to the full range of services, irrespective of geographical location or scientific affiliation"*;

This is not sufficient if the access policies to the resources accessible through this portal do not give access rights to the portal users. These access policies depend mainly on the financing entities and the terms of use of each infrastructure (or service) as identified in our D6.1 paragraph 4.2.

[eInfra] The information about the terms of use of the services in the catalogue is useful. It doesn't solve the issue of the diversity of the access policies but gives necessary information.

### 3.4.5 GAP 5: Low awareness of e-infrastructures and services

This gap concerns the low awareness of the e-infrastructures and services in the research laboratories. To bridge this gap, we proposed in D6.1 and D6.2 to set up common vocabulary, global services, catalogues and to foster dissemination on the EOSC.

[NEIF] The implementation roadmap will contribute to build many bridges in this domain with the development of initial catalogues of services such as the one[31] built by eInfraCentral project. The glossary[32] developed by EOSCpilot work package 5 could be a candidate to build a common global glossary.

The dissemination of all the projects of the Horizon 2020 work programmes will contribute to the awareness of the EOSC among the researchers.

[eInfra] eInfraCentral contributes to a better knowledge of the European e-Infrastructures with the publication of the list of "on-going EU funded e-infrastructure projects ". This list and its presentation is currently available[33]. In addition, its catalogue will provide a better awareness of their associated services.

### 3.4.6 GAP 6: Lack of expertise training, easy tools and human networks

[NEIF] The EOSC portal will give an easy access to all researchers. The document cites the training sessions organised by almost all projects and the competence centres organised around certain communities.

[HLEG] *I12 – "Build a workforce able to execute the vision of the EOSC by ensuring data stewards, data and infrastructure technologists and scientific data experts who are trained and supported adequately."*

[31] http://einfracentral.eu/basic-page/common-service-catalogue
[32] https://www.eoscpilot.eu/eosc-glossary
[33] http://einfracentral.eu/community/e-infrastructures-knowledge-base

# 4 ENSURING INFRASTRUCTURE INTEROPERABILITY

Ensuring permanent and persistent interoperability of e-infrastructures is probably an impossible task, as the only way to ensure interoperability of infrastructures would be to validate all the cases of usage of the infrastructures. However, we can identify the key elements and/or concepts important for e-infrastructure interoperability. Based on these core elements and concepts, some recipes or procedures can be defined and performed.

Ensuring interoperability has to be envisaged basically through the technical aspects and considering a set of rules.

## 4.1 Technical and policy aspects

Within the context of EOSC several elements have been identified as potential risks for e-infrastructures interoperability. Consideration of these elements needs to keep in mind the risk of decreasing or completely breaking the interoperability between infrastructures.

a) A lack of AAI interoperability

Many researchers work in international collaborative groups, driven by common goals and grants. Hosted in different organisations and countries, they need services that are specific to their research community and which are not just aligned with their institution identity and permissions. On the other hand, researchers are already affiliated with an organisation, and they already have credentials that they use to access resources on a daily basis. The authentication and authorisation of users to access resources available within the EOSC is a very important aspect. Accessing services via the EOSC poses some challenges, as the resources in EOSC will be offered by different infrastructures, will require the support of different protocols and different authorisation rules.

Users, however, should be able to use their institutional credentials to access resources available via the EOSC; equally, mechanisms should be in place to support users that may not be able to rely on their institutional credential for whatever reason.

In the past each research or e-infrastructure deployed their own AAI to manage their users and their resources. Users that needed to access services offered by different infrastructures needed to have accounts on each of the infrastructures. This is now changing, as it is recognised that users should be able to access any services that they need by logging in only once (via their institution and/or their preferred infrastructure).

It is clear that infrastructures that are not able to join federations such as eduGAIN to give access to their services do not fulfil the main interoperability requirement. The impact on such infrastructures is double: they are not able to be part of the EOSC and the use of these infrastructures is not possible for external user communities. In that case interoperability does not exist.

b) A lack of network services or poor network services, including:

- poor network quality level, or
- usage of non-standard protocols.

As described in detail in 3.3.2, connectivity, network reliability and orchestration network cloud are key for network interoperability between infrastructures. The network services quality has a direct impact on the services provided by the infrastructure. When the infrastructure provides several basic services involved in a complete scientific workflow, a high level of network services is necessary.

c) A lack of trust between infrastructures, mainly concerning items below

- Traceability

  Traceability is key concerning a large set of elements. On a distributed infrastructure as large as the EOSC, the capability to follow the usages and workflows on all affected infrastructures is mandatory. If this traceability cannot be ensured between infrastructures the trust between these infrastructures is compromised, which is a high-level risk concerning their interoperability.

  This notion of traceability should be considered at not only the computing services level. Also, traceability of the procedures and processes on every infrastructure should be enabled. If traceability is broken data integrity and authenticity may not be guaranteed. The impact may be the loss of trust by other infrastructures and user communities.

- Security

  Security aspects are the first risk to create mistrust between infrastructures. As with traceability, this consideration is about computing services, process and procedures, and also about data, with the security of data being a corner stone of the trust between infrastructures. Deliverable D6.9 "*Final report on data interoperability*" covers this aspect of interoperability. The impact is major.

- Accounting

  All aspects relating to the accounting of resource usage need trust in the information delivered by the neighbouring infrastructures. A real and usable accounting environment, including, for example, the case of workflows across different infrastructures, requires that all the elements required for accounting are available. A lack of accounting data means that integrity of the accounting process is corrupted and accounting is not consolidated.

- Privacy aspects and GDPR compliance

  EOSC as a European infrastructure has to be GDPR compliant. EOSC infrastructures are a large set of infrastructures and services which are not only provided by European entities. In addition potential users are not only working in the European context. Therefore, it is important to ensure that the trust between infrastructures includes GDPR aspects. Lack of GDPR compliance is clearly a break in interoperability.

In summary, in order to allow infrastructures to work together, a certain level of trust is mandatory. This global level of trust is built on different sub-component such as traceability, security, accounting, privacy aspects and GDPR compliance. A lack of trust has a major impact on the interoperability.

d) A lack of information

- About services (end-user services and technical services) and their description (catalogue)

  As detailed in sections 3.3.3 and 4.3.3 of this report it is a requirement that infrastructures' services are described in a common way in catalogues. Infrastructures must comply with the description recommendations and publish their services in the future global catalogue of the EOSC (according to the conditions). First steps are currently the EOSC-Hub and the eInfraCentral catalogues. Services not described in the final catalogue (end-user services and technical services) will not be findable and thus infrastructures that provide these services will not be interoperable with the other infrastructures.

- Information system from infrastructures

Infrastructures generally provide their users with a consistent set of interoperable or compatible services that do not run in an isolation. At least they propose a Single Sign On or a unique entry point to their set of services and often global related services (accounting, monitoring, documentation, user support...). Also, the access rights are often given at the infrastructure level to the subset of services the user or the community is granted access to. The dissociation of services out of their infrastructure-set may be a reason for interoperability breach. In addition, using services provided by different infrastructures may be difficult if they are not technically interoperable (AAI, network, protocols used…). It can be impossible to get a global view on the usage of the services if the infrastructures are not sufficiently interoperable in, for example, the domains of security, accounting, monitoring and privacy.

Figure 9 gives an overview of these risks:


**Figure 9: Main risks for infrastructures interoperability**

## 4.2 Infrastructure interoperability in the rules of participation in WP2

We review the Rules of Participation to highlight issues relating to the gaps identified in service and data interoperability.

Deliverable D2.5 "Recommendations for a minimal set of Rules of Participation"[34] identifies and proposes a minimal set of rules that should be satisfied to join the EOSC. The rules have been designed in harmony with widely accepted working practices in already established organisations, embracing the principles of openness, transparency and inclusiveness, though not explicitly embracing interoperability itself. As the infrastructures of EOSC service providers will have to comply with this set of rules, it is important to assess how the rules contribute to infrastructure interoperability. In this section we study how the Minimal Rules of Participation address the major infrastructure interoperability gaps identified in D6.1.

### 4.2.1 Gap 1: Diversity and incompatibility of the AAIs

In deliverable D2.5 **Rules specific to Core Resource of EOSC** are defined as "*the set of services and processes that are needed for EOSC operations to integrate and enable access to the various resources federated in the EOSC*". The deliverable also acknowledges that Core Resources of EOSC are the '*glue*' of EOSC, as in Figure 10. One major example of such Core Resource is Authentication and Authorisation Infrastructure (AAI). The rules also recommend that "*such core services are bound by Service Level Agreements and interoperability aspects"*. Thus, core services within the EOSC are expected to maintain interoperability as overseen by the governance of the EOSC.


**Figure 10: Importance of EOSC Core Resources in the EOSC. This plays a central role in EOSC Compliance and supports both EOSC and external resources[35].**

### 4.2.2 Gap 2: Network services

**Quality of service** recommendation stipulates the minimal set of quality guidelines that are being developed within the EOSC. As mentioned above, network services interoperability refers to connectivity, reliability and orchestration of network and cloud. The recommended quality guidelines will require adhering to standards such as international ones like ISO and/or domain specific standards, and would enable users to make informed choices based on quality, performance and capacity. Such quality 'certifications' would also build trust with users.

### 4.2.3 Gap 3: Diversity of services and providers

This gap is addressed through the following recommended rules of participation:

[34] https://eoscpilot.eu/content/d25-recommendations-minimal-set-rules-participation
[35] https://eoscpilot.eu/content/d22-draft-governance-framework-european-open-science-cloud

EOSC Services **shall be registered** in an EOSC Compliant or compatible service catalogue visible to the global EOSC gateway. This recommendation also stipulates that the service catalogue should contain interoperability information, supported through published metadata schemas and data formats, along with other information, for example, service availability, functionality, openness of licences, privacy, term of use and contractual frameworks.

**Machine-readable metadata** are a corner-stone for service interoperability enabled through EOSC catalogues. Provision of machine-readable metadata raises the level of interoperability between services, allowing machine-to-machine harvesting and aggregation, searching and brokering. This should also include details specific to communities that the service caters to. Such details will enable users to assess services on characteristics that are specific for their line of work. For example, quality indicators, FAIR-ness of data, licence restrictions etc.

Interoperability can also be driven by aligning with the *EOSC Implementation Roadmap*[36] published by the European Commission. There is a clear focus *on the use of tools, specifications, catalogues and standards (framework for FAIR research data)* and hence, certifications like CoreTrustSeal[37] and ISO 20614:2017[38] are recommended, where applicable and appropriate. Furthermore, projects like FAIRsFAIR[39] that have activities that assess FAIRness of data/service will provide additional certification services to ensure that diversity between services and providers do not hinder interoperability.

### 4.2.4 Gap 4: Diversity of access policies

As per the recommended Rules of Participation, **Terms of Use and Acceptable Usage Policies** need to be established for services to be incorporated into EOSC framework. Infrastructures/Service Providers will need to make publicly available terms of use including Access Policies - such as who is able to access the services/datasets, how long for, who are these used by/aimed for, how long will the data be available, security and privacy considerations, SLAs etc. EOSCpilot WP 3 EOSC Policy also recommends use of machine-readable licences for data services wherever possible, to allow automatic brokering and access services.

### 4.2.5 Gap 5: Low awareness of e-infrastructures and services

The key rule of participation applicable for this gap is that services should be registered into the **EOSC compatible service catalogue**. Services provided by Research or e-Infrastructure would need to be enlisted with the necessary indicators of performance, quality and functionalities. By accessing services via the portal, awareness of their capabilities can be enhanced.

### 4.2.6 Gap 6: Lack of expertise training, easy tools and human networks

As part of the minimal Rules of Participation to the EOSC, the recommendations require all services/tools to be accompanied by corresponding documentation, support and training materials and contact channels. These are aimed to **ease usability** of the services. The infrastructures are also encouraged to publish their pledge to FAIR principles – whether through certifications or through organizational commitments. This would increase the interoperability and usability of their services.

The Skills and Competency work package (WP 7) has considered extensively the implications of the EOSC services to service providers and has conducted skills landscape analysis and created a competency model (deliverable D7.1[40]). The Skills and Capability Framework (deliverable D7.3[41]) describes the framework that infrastructures can use to identify what skills are required of individuals and organisations wishing to use their services and how to address any skills gaps. The Figure 11 (from D7.3) illustrates the process for filling skills gaps within the EOSC framework.

[36] http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf
[37] https://www.coretrustseal.org/
[38] https://www.iso.org/standard/68562.html
[39] FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-5-2018-2019 (c), grant agreement 831558. https://www.fairsfair.eu
[40] https://www.eoscpilot.eu/content/d71-skills-landscape-analysis-and-competence-model
[41] https://www.eoscpilot.eu/content/d73-skills-and-capability-framework

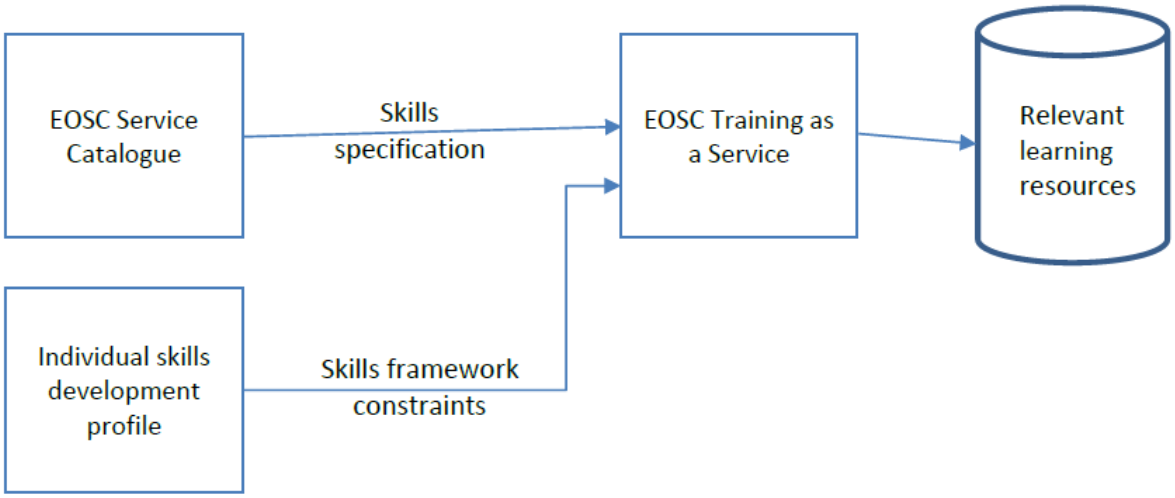


**Figure 11: Overview of the three-step procedure to identify relevant learning resources. (Ref: D7.3)**

In summary, the Minimal Rules of Participation address the major interoperability gaps as follows:

| Gap | Minimal rule of participation recommendation |
|---|---|
| **GAP1: Diversity and incompatibility of AAI** | Rules specific to Core Resources of EOSC |
| **GAP2: Network services** | Quality of service |
| **GAP3: Diversity of services and providers** | Registry in EOSC Compatible Service Catalogues |
| **GAP4: Diversity of access policies** | Terms of Use and Policies |
| **GAP5: Low awareness of e-infrastructures and services** | Portability<br>Registry in EOSC Compatible Service Catalogues |
| **GAP6: Lack of expertise training, easy tools and human networks** | Ease of usability (Relation to users)<br>FAIRness and Reproducibility (Relation to users) |

## 4.3 How to verify e-infrastructure interoperability

Some aspects of the infrastructure interoperability were already considered on the Deliverable D6.2[42]. We consider them as still valuable.

### 4.3.1 Interoperability Auditing

Interoperability auditing forms the basis for verifying the interoperability of infrastructures within the context of EOSC. The responsibility for conforming to the elements in the criteria rests with the infrastructure and service providers. Assessment of interoperability should be done within the New European Interoperability Framework [43] that is broadly recommended for European public services. These include: *return on investment, total cost of ownership, level of flexibility and adaptability, reduced administrative burden, efficiency, reduced risk, transparency, simplification, improved working methods, and level of user satisfaction*. An interoperability auditing process would need to include the following aspects:

[42] https://www.eoscpilot.eu/content/d62-eosc-architecture-design-and-validation-procedure
[43] https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

- **AAI:** the Interoperability Audit should check that the infrastructure is compliant with AAI services as recommended in the AARC Guidelines and following the AARC Blueprint Architecture (BPA). This enables the services provided by the infrastructure to provide a good user experience when using different EOSC services. This would lower the barrier for users needing to manage multiple identities and having to address multiple user authentication challenges.
- **Traceability**: The audit will ensure that service and data-access and usage traceability mechanisms are available in the infrastructure. Infrastructures should also have accounting systems to monitor the usage of the services, data and tools.
- **Service Quality Management**: Infrastructures enabling services through EOSC would have to adhere to standards of service quality management as recommended by the minimal rules of participation. While there are many standards of service management available, the infrastructure could make an informed decision to adopt the best practice/standard that is relevant to their service operation like ISO/IEC 27001[44], ITIL[45] or FitSM[46]. FitSM is a lightweight standard aimed at facilitating service management in IT service provision, including federated scenarios. Deliverable D5.3, *EOSC Federated Service Management Framework*[47], discusses service management for the EOSC with reference to FitSM.
- **Service catalogue**: The service catalogue is a keystone of interoperability. The interoperability audit would have to ensure that the service conforms to the conditions of entry into the service catalogues, such as compliance with minimal metadata schemas, and that links to the services/tools, quality standards, service availability are up-to-date.
- **FAIR Data:** EOSCpilot Work Package 3 (Policy) has established the *Open Science Policy Toolkit* (deliverable D3.5[48]). These are a set of third-party tools/services that are designed to facilitate the development and implementation of Open Science policies by the different stakeholders operating in the EOSC – including Research Infrastructures. Complementary to the Policy Toolkit are the *Open Science Monitor specifications* (deliverable D3.2[49]) measuring FAIRness, Openness, skills and impact. These play an important role in delivery and monitoring the effectiveness of interoperability. According to D3.2, measuring FAIRness needs to consider the following factors:

| FAIR principle | Factors |
|---|---|
| **Findability** | Visibility of research data, service, tools |
| **Accessibility** | Protocols of access, location of deposition/publication, understandability by users, access costs, embargo periods |
| **Interoperability** | Machine readability |
| **Reusability** | Open licences, provenance |

Interoperability verification will need to be conducted within the context of the overall Governance structures of the EOSC. The governance structure setup by the EOSC should periodically evaluate infrastructures on the basis of interoperability to maintain the level of service standards and the applicability of the EOSC across all domains and applications of research. The EOSCpilot has proposed recommendations for the governance of the EOSC. This includes an Executive Board, with overall responsibility for an action plan for scientific data interoperability to ensure the usage of the FAIR principles. In addition EOSC governance is supported by Technical Working Groups, and we recommend that such technical working groups appraise infrastructures/services/tools on their conformity to interoperability standards. Using the

[44] https://www.iso.org/isoiec-27001-information-security.html
[45] https://www.axelos.com/best-practice-solutions/itil
[46] https://fitsm.itemo.org/
[47] https://eoscpilot.eu/content/d53-eosc-federated-service-management-framework
[48] https://eoscpilot.eu/content/d35-open-science-policy-toolkit
[49] https://eoscpilot.eu/content/d32-eosc-open-science-monitor-specifications

above auditing criteria, the working groups would develop audits depending on the community and user needs of interoperability of the infrastructures.

### 4.3.2 Verification Checklist

In order to summarize Interoperability verification we can consider two levels of verification that have to be assessed for infrastructure interoperability. The first one is an internal assessment of the infrastructure undertaken by the infrastructure itself; the second is an external assessment of the infrastructure by other entities in the EOSC. For both cases a checklist can be defined in order to help the infrastructure provider to verify that the infrastructure is likely to be interoperable.

**Internal assessment of Infrastructure Interoperability:**

- The infrastructure is compliant with the EOSC policy and rules.
- The services provided by the infrastructure are correctly detailed and are available on the infrastructure catalogue and are ready to be integrated in the EOSC catalogue. For more details on the services themselves, please refer to the work of Work Package 5.
- The infrastructure is compliant with AAI services such as recommended in the AARC Guidelines[50] and following the AARC Blueprint Architecture (BPA).
- The infrastructure is connected to and accessible via the network.
- The infrastructure access policy allows external usage.
- Traceability mechanisms and accounting of the usage of the infrastructure are available.
- Security and privacy are managed at the infrastructure level, and the infrastructure is GDPR compliant.
- All the mandatory infrastructure information required for integration in to the EOSC is correctly defined and published.
- The infrastructure supports interfaces and formats that are compatible with the recommendations of EOSC Governance bodies.
- The infrastructure staff are aware of the requirements of infrastructure interoperability.

**External assessment of Infrastructure interoperability:**

- The infrastructure can be reached from other infrastructures/services.
- The information provided by the infrastructure is complete and accessible from other infrastructures.
- It is important to verify that the infrastructure can be used on a sample workflow, in-place-of another infrastructure providing the same functionalities.

[50] https://aarc-project.eu/guidelines/

# 5 RECOMMENDATIONS

Based on the gaps and bridges to be built identified during the work already performed in the e-Infrastructure Gap Analysis (deliverable D6.1[51]) and based on the framework proposed in the EOSC Architecture Design and Validation Procedure (deliverable D6.2 [52]), recommendations are proposed to ensure infrastructure interoperability in the context of the EOSC. These recommendations have now been refined and completed.

For each of the six gaps identified we propose a series of recommendations.

## 5.1 GAP 1: Diversity and incompatibility of the AAIs

As presented on the AARC website[53]:

*The AARC Blueprint Architecture[54] (BPA) is a set of software building blocks that can be used to implement federated access management solutions for international research collaborations. The Blueprint Architecture allows software architects and technical decision makers to mix-and-match tried and tested components to build customised solutions for their requirements.*

*The current version (AARC-BPA-2017) consists of four component layers grouped by functional roles:*

- ***User Identities**: Services which provide electronic identities that can be used by users participating in International Research Collaborations.*
- ***Identity Access Management**: Defines an administrative, policy and technical boundary between the internal/external services and resources.*
- ***Authorization**: Contains elements to controls the many ways users can access to services and resources.*
- ***End-services**: Where the external services interact with the other elements of the AAI.*

*The AARC Guidelines[55] complement the AARC Blueprint Architecture (BPA) and the policy best practices recommended by the AARC project. The guidelines can apply to any topic that helps to advance Federated Identity Management for research and collaboration.*

*The AARC Guidelines help communities and infrastructures to implement and operate an AAI for research and collaboration more effectively and in an interoperable way.*

**Recommendation 1**: *Implementation of a federated AAI: adopt the AARC Blueprint Architecture and follow the associated guidelines.*

## 5.2 GAP 2: Network Services

Research and educational networks (NRENs and GÉANT) offer the best route between academic data centres and enforce an overprovisioning capacity policy, whereas standard Internet connections offer much less capacity and potentially poor interconnection, depending on commercial agreements. Therefore, NRENs offer optimal connectivity in terms of routes between the EOSC academic data centres.

See section 3.3.2 for more details.

**Recommendation 2**: *Foster the network collaboration between EOSC and the NRENs in order to anticipate new capacity requirements.*

[51] https://eoscpilot.eu/content/d61-e-infrastructure-gap-analysis
[52] https://eoscpilot.eu/content/d62-eosc-architecture-design-and-validation-procedure
[53] The AARC and ARCC2 projects are coordinated by GÉANT.
[54] https://aarc-project.eu/architecture/
[55] https://aarc-project.eu/guidelines/

## 5.3 GAP 3: Diversity of services and providers

This deliverable focuses on infrastructures and is different to deliverable D5.4 - "*Final EOSC Service Architecture*" developed by WP5. Our recommendations are limited to the context of infrastructure interoperability. However, also see the recommendations about services that are outlined in D5.4.

**Recommendation 3A:** *Infrastructures should describe their services in a standardised way, such as by using the eInfraCentral template. This should include services that act as service components in the provisioning of higher-level services.*

As developed in 3.4.3 e-infrastructures build generally a set of interoperable or compatible services. These services may be interdependent in the sense that they are often linked together with other services such as an infrastructure level accounting, monitoring, security or Single Sign on.

**Recommendation** *3B***:** *Infrastructures should identify their services' interoperability dependencies and to correctly describe these interdependencies.*

**Recommendation 3C:** *Infrastructures should be able to provide complete accounting data, to take care of the users' privacy in a GDPR compliant way, to provide traceability and to collaborate actively or at least follow the guidelines of all federative groups that work on the EOSC global infrastructure (AAI, security, privacy, global traceability, global information system…)*

**Recommendation 3D**: *The EOSC itself should organise security channels to ensure security and privacy at EOSC level, and ensure global traceability, and to set up the EOSC information system.*

## 5.4 GAP 4: Diversity of access policies

Despite the fact that the diversity of access policies is an obstacle to the interoperability, the solution to this issue is not always within the scope of the architecture of infrastructures interoperability.

However, partial solutions identified in deliverable D6.1 (*e-Infrastructure Gap Analysis*) section 4.2.1 are relevant to support access to the EOSC for groups of researchers that, because of access policies, cannot use any other infrastructure - they are mainly part of the "long tail of science".

**Recommendation 4A**: *In order to support access to the EOSC for groups of researchers that, because of access policies, cannot use any other infrastructure, EOSC should build a mutualised space (agnostic to discipline and geography) in order to serve all researchers, from all disciplines, in all countries, and to provide them with services that cannot be fulfilled by other means.*

This 4A recommendation targets mainly the long tail of science or new emerging communities during their set up period.

**Recommendation 4B**: *The infrastructures that the EOSC is composed of should consider how to harmonise access and usage policies, in order to minimise the conditions that users need to accept to be able to access resources within the infrastructure, thus encouraging interoperability and reuse across the participating providers.*

Infrastructures are financed by funding agencies that generally define the access/usage rules to be applied to the resources for which they provide funding. Scientific workflows that require access to different infrastructures may be impossible to run because of incompatible rules.

**Recommendation 4C**: *EOSC should propose incentives to support funding agencies in making the resources they fund more openly available. EOSC should work with the infrastructures accessible through excellence-based applications to explore how to facilitate the interoperability between these infrastructures and EOSC for the benefit of all users.*

**Recommendation 4D:** *Infrastructures should define and publish applicable Service Terms of Use including acceptable usage policies in the EOSC Service catalogue as proposed by eInfraCentral. Machine readable-licences in interoperable formats are encouraged, to allow automatic brokering and access services across infrastructures to be supplied.*

## 5.5 GAP 5: Low awareness of the e-infrastructures and services

Low awareness of the e-infrastructures and services is a difficulty for infrastructures interoperability. The solutions largely exceed the scope of this deliverable, however, we recommend the EOSC as already mentioned in deliverable D6.1:

- Build and foster a common vocabulary (at least a glossary) that is a necessary basis to allow infrastructures to understand each other.
- Build an EOSC Service catalogue. This is an on-going work in the framework of EOSC-hub and eInfraCentral projects.

**Recommendation 5A**: *The EOSC should promote the setup of a common vocabulary. Infrastructures should publish their services in a catalogue such as the EOSC-hub catalogue that is foreseen to be the first step of the EOSC Service catalogue*.
This recommendation is related to and reinforces recommendation 3A.

**Recommendation 5B**: *Disseminate and widely promote the EOSC at all levels of the European Research Area, including European infrastructures, national organisations, funding agencies and research laboratories, in order to improve the global knowledge of the computing landscape.*

## 5.6 GAP 6: Lack of expertise training, easy tools and human networks

The lack of expertise training, easy tools and human networks are obstacles to the interoperability that were identified in deliverable D6.1 (sections 4.2.2 and 4.2.3).

**Recommendation 6A**: *Infrastructures should use the "EOSC Skills and Capability Framework" (D7.3)[56]. This describes the framework that can be used to identify what skills are required of individuals and organisations wishing to foster interoperability.*

**Recommendation 6B**: *Foster networking activities to share expertise and experience among infrastructure technical staff in order to create the conditions for technical experts to find and mutualise the right way to support their users and to interoperate. Foster the development and use of portals across different systems,*

[56] https://www.eoscpilot.eu/content/d73-skills-and-capability-framework

*and of comparable user-friendly tools that allow users to easily deploy their data analysis pipelines. Be careful to provide users with enough skilled technical staff to guide them in their use of the EOSC.*

Figure 12 gives an overview of the main recommendations.

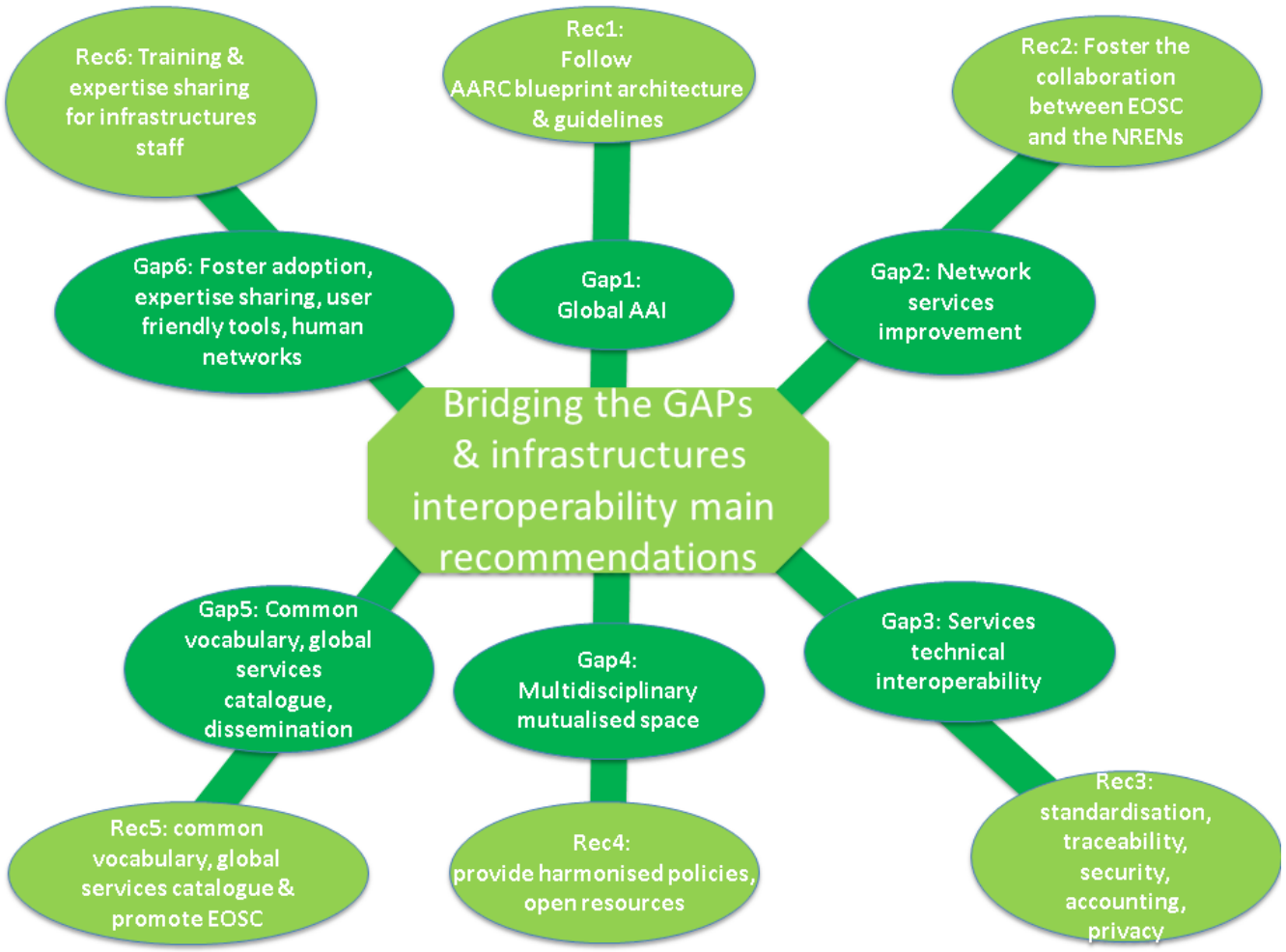


**Figure 12: Main recommendations to ensure infrastructure interoperability**

# 6 CONCLUSION

Interoperability of e-infrastructures is a corner-stone to the success of the EOSC as it will allow the implementation of workflows through using services across infrastructures. At the same time, this interoperability will allow users to build these workflows from a large set of mostly user friendly, useable and powerful components. This interoperability will help support the adoption of the EOSC by users. As this deliverable covers issues of infrastructure interoperability architecture, no conclusions or recommendations related to the issues of data interoperability are mentioned. However, they are fully discussed in deliverable D6.9: Final report on Data Interoperability.

In this report we have taken into account the results from the previous deliverables D6.1 and D6.2. We have identified the main risks that occur if the interoperability gaps are not taken into account. We have considered approaches to auditing and verification, and we also provide six sets of recommendations to ensure interoperability between infrastructures at the level of the infrastructures and globally across the EOSC.

Interoperability can be guaranteed through a set of interoperability conditions. Missing any one of them may completely break the interoperability. Infrastructure interoperability relies on technical solutions and/or best practices and also on human expertise, sharing and collective decisions. Even if the first three gaps (i.e. AAIs, networks, and diversity of services) seem to be more technical and more related to this interoperability architecture, parts of the solutions are directly linked to a global EOSC organisation. The last three gaps are less technical and more organisational, however, they are key issues for the EOSC adoption by user communities, particularly by the communities working in the "*long tail of science*". This is a vital aspect that has to be considered to ensure the success of the EOSC.

# ANNEX A. QUESTIONNAIRE SENT TO INFRASTRUCTURES AND SCIENCE DEMONSTRATORS

The following questionnaire was sent to infrastructures and the science demonstrator projects (with the help of the Management of Work Package 4).


## Inputs to the WP6 "Final EOSC (interoperability) architecture".

The EOSC Interoperability work package of the EOSCpilot project develops and demonstrates the interoperability requirements between e-Infrastructures, domain research infrastructures and other services providers needed in the European Open Science Cloud. A first survey have been conducted in the first part of 2017 and thanks to all of you, have provided the main input for two documents: "e-infrastructure gap analysis" (https://www.eoscpilot.eu/content/d61-e-infrastructure-gap-analysis) and "EOSC Architecture Design and Validation procedure" (https://www.eoscpilot.eu/content/d62-eosc-architecture-design-and-validation-procedure). You may consult a summary at https://www.eoscpilot.eu/themes/wp6-interoperability.

The objective of this second survey is to provide an update that is to be used as input for the "Final EOSC (interoperability) architecture".

We would like to get your input by the end of August.

| | |
|---|---|
| E-infrastructure, site, Science Demonstrator: | |
| Contact: | |
| Date: | |
| Version of the document: | |

1. Do you think that the identified gaps are relevant for you?
2. From your point of view what did we missed about infrastructures interoperability in this report?
3. From your point of view are the proposed bridges (solutions to bridge the gaps) to build relevant enough?
4. Are there specific scientific use cases that may encounter issues of interoperability and that you would like us take especially into account in this architecture interoperability study?
5. Comments.

# ANNEX B. GLOSSARY

The EOSC Glossary is available from the EOSCpilot website at https://www.eoscpilot.eu/eosc-glossary. The glossary will evolve with the EOSC through the EOSCpilot Glossary Working Group and in forthcoming EOSC projects.