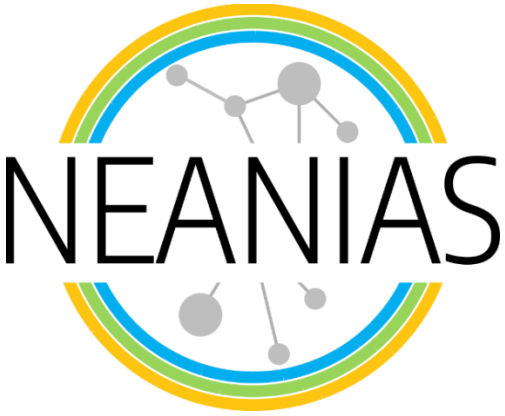


NEANIAS
Novel EOSC services for Emerging Atmosphere, Underwater and Space Challenges

NEANIAS Whitepaper: Best practices for sustainable services on EOSC

Ioannis Neokosmidis (INCITES), Vangelis Logothetis (INCITES), Theodoros Rokkas (INCITES), George Loumos (INCITES), George Papastefanatos (ATHENA), Christos Tsiakaliaris (ATHENA), Nikos Chondros (NKUA), Petra Eleni (NKUA), Aggeliki Gkamiliari (ATHENA), Eva Sciacca (INAF), Georgios Papanikos (CITE), Georgios Kakaletris (CITE), Makis Ntouskos (ATHENA)

NEANIAS is funded by European Union under Horizon 2020 research and innovation programme via grant agreement No. 863448

Disclaimer

NEANIAS is a Research and Innovation Action funded by European Union under Horizon 2020 research and innovation programme, via grant agreement No. 863448.

NEANIAS is project that comprehensively addresses the 'Prototyping New Innovative Services' challenge set out in the 'Roadmap for EOSC' foreseen actions. It drives the co-design, delivery, and integration into EOSC of innovative thematic services, derived from state-of-the-art research assets and practices in three major sectors: underwater research, atmospheric research and space research. In each sector it engages a diverse set of research and business groups, practices, and technologies and will not only address its community-specific needs but will also enable the transition of the respective community to the EOSC concept and Open Science principles. NEANIAS provides its communities with plentiful resource access, collaboration instruments, and interdisciplinary research mechanisms, which will amplify and broaden each community's research and knowledge generation activities. NEANIAS delivers a rich set of services, designed to be flexible and extensible, able to accommodate the needs of communities beyond their original definition and to adapt to neighboring cases, fostering reproducibility and re-usability. NEANIAS identifies promising, cutting-edge business cases across several user communities and lays out several concrete exploitation opportunities.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently 28 member states of the European Union. It is based on the European Communities and the member states' cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice, and the Court of Auditors (http://europa.eu.int/).

Table of Contents

Abstract

NEANIAS is an EU project that co-designs, develops and deliver innovative thematic services for European Open Science Cloud (EOSC) communities in three major sectors: Underwater research, Atmospheric research and Space research. NEANIAS aims to maximize the industry/research engagement by identifying and addressing all the challenges ensuring thus the sustainability of the services in question. These challenges are stemming not only from the funding needs but also from requirements of EOSC, the availability of data, the need for a common governance as well as from security and privacy related issues.

This white paper is based on the lessons learnt and the experiences gathered during the project. It describes the best practices for sustainable services on EOSC which could be a useful tool for all the stakeholders of the EOSC ecosystem.

1. Introduction

The services developed by NEANIAS are mainly provided by academic partners following the free model, meaning that most of them are offered free of charge. However, all services (with and without revenues) are accompanied by a number of costs related to their operation, maintenance and update. Taking into account that in most cases revenues are not able to cover expenses, self-(financial)-sustainability seems to be impossible.

However, sustainability also encompasses other elements such as the use of existing infrastructures, the availability of data, a governance model, security and privacy issues as well as EOSC inherent requirements and limitations.

In this white paper, best practices derived through NEANIAS journey for sustainable services on EOSC are described. A study of the European funding schemes has been performed, identifying key schemes available in Europe, at both national and transnational levels, that would meet the needs of NEANIAS services.

Then, the use of existing infrastructures is discussed. Different options are provided along with their advantages and disadvantages. A new approach to Kubernetes provisioning is also presented as the most promising for Kubernetes-based deployments. National and regional funding to access existing infrastructure is also reviewed. In addition, EU funding for usage (virtual access) is examined as a means to cover both operational and infrastructure costs. Different models for calculating Virtual Access costs for funding are presented.

Governance of common/shared infrastructure is also discussed as a means for simplification of the communication between services providers (thematic and core/dependent services). The availability of data is then investigated. Proposals to improve data availability are provided. The approaches followed in NEANIAS are also given.

Finally, security and privacy issues are discussed. The elements of Authentication and Authorization Infrastructure (AAI), logging and accounting are presented. The impact of security and privacy elements on users trust, engagement, accessibility and flexibility of registrations as well as on services cost and charging schemes is also examined.

The best practices provided in this white paper can be proved a valuable tool for service providers, decision/policy makers and other stakeholders of EOSC ecosystem.

2. Funding for Services and Infrastructure

2.1. EU and National Funding for the Services

Since the NEANIAS services are mainly provided by academic partners, it is not expected to have revenues as per the business models described. However, all services (with and without revenues) are accompanied by the costs for operation, maintenance and upgrade. Thus, some of the services, especially those without revenues, are not able to cover their expenses. Hence, in order to ensure the financial sustainability of services, several funding mechanisms should be investigated.

This document analyses funding tools available to the public and private sector at the European level. The funds to be mentioned are:

- European Union funds
- Regional funds
- Funds from Institutional projects
- National Research funds

2.1.1. European Union funds

In 2020, the European Union approved its budget for the period 2021-2027. At the same time, the major economic crisis generated by the Covid-19 pandemic in its Member States required extraordinary funding by the EU to sustain their economies, which translated into the adoption of the Recovery Plan for Europe. The greatest novelty of the Recovery Plan for Europe is the NextGenerationEU, a €750 billion temporary recovery instrument that will allow the Commission to raise funds on the capital market. The Recovery plan for Europe will offer a total of €1,8 trillion to help rebuild a post-Covid-19 Europe. Within this stimulus package, €143,4 billion are earmarked for the single market, innovation and digital questions.

It should be noted that to attribute the funds of the Recovery Plan, the Commission is mainly supplementing a number of its existing programs for the 2021-2027 period, such as Horizon Europe or the InvestEU Programm which were bolstered up.

2.1.1.1. The InvestEU program 2021-2027

This program is important because it brings together most of EU financial instruments currently available. The program expands the successful model of the Investment Plan for Europe, the Juncker Plan. With InvestEU, the Commission will trigger at least €650 billion in additional investment.

The InvestEU fund is composed of:

- The EU long-term budget which has attributed €15.2 billion to the InvestEU fund;
- An estimated € 47,5 billion guaranteed from the EU budget and financial partners' resources which comprises € 11,25 billion for research and innovation and the same amount for SMEs; and
- The Junker plan, which will crowd-in public and private investors to reach a total estimated investment of at least € 650 billion.

2.1.1.2. Horizon Europe

Horizon Europe is the new 2021-2027 research program which has a budget estimation of €97,6 billion, €3,5 billion of which will be allocated under the InvestEU Fund.

The Global Challenges and Industrial Competitiveness pillar (€52,7 billion) directly supports research relating to societal challenges, reinforces technological and industrial capacities, and sets EU-wide missions to tackle problems. It also includes activities pursued by the Joint Research Centre (€2,2 billion) which supports EU and national policymakers with independent scientific evidence and technical support. The Open Innovation pillar (€13,5 billion) focuses on market-creating innovation via the European Innovation Council (€10 billion). The first Horizon Europe Strategic Plan (2021-2024) was published at the beginning of 2021.

2.1.2. Regional Funds

The EU's Regional Policy targets all regions and cities in the European Union in order to support job creation, business competitiveness, economic growth, sustainable development, and improve citizens' quality of life.

Regional Policy is delivered through two main funds: the European Regional Development Fund (ERDF) and the Cohesion Fund (CF). Together with the European Social Fund (ESF), the European Agricultural Fund for Rural Development (EAFRD) and the European Maritime and Fisheries Fund (EMFF), they make up the European Structural and Investment (ESI) Funds.

The European Regional Development Fund and Cohesion Fund 2021-2027 will be regulated by a **new single regulation covering the ERDF and Cohesion Fund funds**, in accordance with a proposal of the Commission currently under review by the EU Parliament. In January 2021, the European Institutions have entered into the *Trilogue* negotiations i.e. tripartite meetings between Parliament, the Council and the Commission to approve the regulation.

2.1.3. Funds from Institutional projects

2.1.3.1. ESA (European Space Agency) projects

ESA's Partnership Projects provide the satellite communication industry with the environment to introduce innovative space-based solutions systems into the commercial market. ESA has implemented several Partnership Projects through its ARTES program (Advanced Research in Telecommunications Systems). EU companies that wish to partner with ESA can respond to an invitation to tender, a call for ideas, or a call for proposals.

2.1.3.2. EUREKA

Eureka is an intergovernmental decentralized organization for pan-European research and development funding and coordination. EUREKA is an open platform of 48-member countries including the EU, for international cooperation in innovation. EUREKA 'Clusters' programs are long-term, strategically significant industrial initiatives. They operate with a large number of participants to develop inclusive technologies of key importance for European competitiveness mainly in ICT, energy and more recently in the biotechnology and automation

sectors. Eureka Clusters have had an impact on the ability of the European microelectronics sector to compete with other continents.

The Eureka Clusters are:

- CELTIC NEXT: Information and Communications Technologies
- EURIPIDES: Smart Electronic Systems
- ITEA 3: Software Innovation
- METALLURGY EUROPE: New metals
- SMART: Advanced manufacturing program
- PENTA: Micro and nanoelectronics enabled systems and applications
- EUROGIA2020: Low-carbon energy technologies

2.1.4. National Research Funds

Apart from the nationally administered funds mentioned before, that are still linked to the European Union, several European countries have their own national research programs that have a substantial budget for research financing. Going through the list of individual countries is time-consuming; nevertheless, all of the research funding organizations have an international character, therefore they should (in principle) be in line with EOSC's objectives and vision.

2.2. Funding for Research Infrastructures

In the course of the previous Structural Funds Programming Period, the local competent authority of each partner for Research, Technological Development and Innovation Policy, such as General Secretariat for Research and Innovation (GSRI) in Greece, completed the development of a National Strategy for Research Infrastructures and a Multiannual Financial Plan for Research Infrastructures, highlighting each country's priorities for long-term investments in large-scale research infrastructures. It must be noted that developing a National Strategy and a Multiannual Financial Plan for Research Infrastructures was an ex-ante conditionality for receiving Structural Fund financing under the 2014-2020 Programming Period for all European countries. In Greece, specifically, the Financial Plan covers 28 national-scale, mostly distributed, research infrastructures.

In every European country, at local level, the GSRI is supported by the Policy Support Facility of the European Commission's Directorate-General for Research and Innovation in developing the Strategy for Research Infrastructures under the 2021-27 Programming Period. An international expert group has been tasked with evaluating the performance of the Research Infrastructures comprised in the Multiannual Financial Plan, as well as their contribution to the economy and society at large, their relevance to the Smart Specialization Strategy, their international visibility and development prospects.

It is important to point out that the notion of research infrastructures is not confined to buildings and equipment, but it also encompasses human resources, know-how, information, networking and all intangible assets required for their operation and optimal use. Every European country designs and implements a transparent and robust procedure, based on

European standards, to document the infrastructure needs of the country's research community and productive sectors, to promote necessary synergies and collaborations, to assess and, eventually, identify integrated, national-scale, open-access infrastructures with outward-looking, sustainable and innovation-supporting features, implemented and put into operation by priority starting from the 2014-2020 Programming Period and beyond.

EOSC would be viewed as a result of a top down (governmental) and a bottom-up initiative (covering projects and research centers). It mainly needs a global European strategy and support to be conveyed in every European country and extra push from the top-down level to further formalise and enhance the support process. Towards the success of this particular approach (both top down and bottom-up approach) from all European countries, relative ministries should construct a combined alliance and network.

In national level, what is needed is a common national strategy covering globally all governmental institutions, ministries and organizations, with all relevant ministries working together under a common base lead by every country's strategy starting from the Competence authority's strategy and the Digital Governance's one. This strategic approach in national level should be in compliance to the General European Policy. In addition to that both in national and in European level, a coordination schema is needed involving representations of the basic core infrastructures, the thematic infrastructures, the relevant and connected organizations of all ministries and of course the community engagement with particular stakeholders' representatives.

A binding glue is needed with a strong coordination also together with EOSC ecosystem, such as EOSC Association, and local European partners' initiatives of Open Science, such as OPENAIRE NOADs, OSIs (local Open Science Initiatives) as HOSI in Greece.

2.3. EU funding for usage – Virtual Access

One of the models applied by EC for supporting the sustainability of services that attract researchers and scientists is the model of Virtual Access. In brief, under this model, services may be funded by the EC in the context of projects by supporting new users that onboard the service.

Behind Virtual Access funding, there are well justified terms:

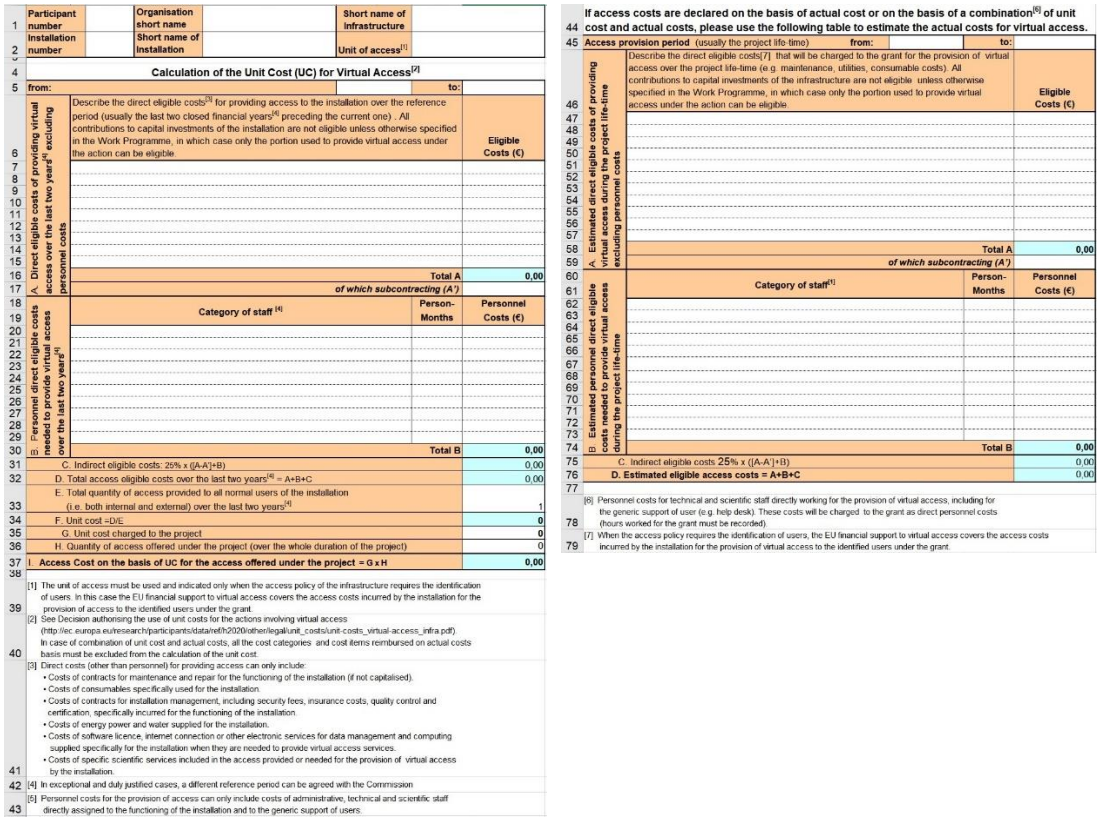
- Virtual Access may be provided under a trans-national model contrasted to a national model that should be seeking other resources for support.
- Access to the users must be free.
- Access must be accounted via formal means of the infrastructure.
- Access must be targeting the user groups defined by the respective call as those may be further specialized by the proposal that is being funded.

To support Virtual Access funding, certain decisions have been taken by the EC (e.g. [1], [2]) and related terms have been added in the H2020 Grant Agreement template (see H2020 Annotated Grant Agreement [3]).

Under the model of Virtual Access, users may benefit from a multitude of Infrastructure components such as:

- Compute equipment mainly behind large research datacentres of EU, under the model of Platform-as-a-Service (PaaS) or Infrastructure-as-a-Service (IaaS), along with its hardware maintenance, power and cooling provisions.
- Storage resources for data both for temporary deposition and processing as well as long term persistence.
- Sophisticated equipment such as high cost or sparse instruments, abstracted via services. Telescopes, antenna or sensor arrays, drones etc are some examples.
- Data sets, algorithms and other assets offered on top of infrastructures.
- Generic and highly specialized software services providing access to tools, data, algorithms etc.
- Network and communication resources, directly or indirectly related to the access of the aforementioned infrastructure components.

In the following figure, the tool provided by EC for calculating the Virtual Access costs in prior calls is shown (can be found in [4] as OpenXML document).



1	Participant number		Organisation short name		Short name of infrastructure	
2	Installation number		Short name of installation		Unit of access[1]	
4	Calculation of the Unit Cost (UC) for Virtual Access[2]					
5	from:				to:	
6	A. Direct eligible costs of providing virtual access over the last two years[4] excluding personnel costs	Describe the direct eligible costs[3] for providing access to the installation over the reference period (usually the last two closed financial years[4] preceding the current one) . All contributions to capital investments of the installation are not eligible unless otherwise specified in the Work Programme, in which case only the portion used to provide virtual access under the action can be eligible.				Eligible Costs (€)
16					Total A	0,00
17					*of which subcontracting (A')*	
18	B. Personnel direct eligible costs needed to provide virtual access over the last two years[4]	Category of staff [5]			Person-Months	Personnel Costs (€)
30					Total B	0,00
31	C. Indirect eligible costs: 25% x ([A-A']+B)					0,00
32	D. Total access eligible costs over the last two years[4] = A+B+C					0,00
33	E. Total quantity of access provided to all normal users of the installation (i.e. both internal and external) over the last two years[4]					1
34	F. Unit cost =D/E					0
35	G. Unit cost charged to the project					0
36	H. Quantity of access offered under the project (over the whole duration of the project)					0
37	I. Access Cost on the basis of UC for the access offered under the project = G x H					0,00
38						
39	[1] The unit of access must be used and indicated only when the access policy of the infrastructure requires the identification of users. In this case the EU financial support to virtual access covers the access costs incurred by the installation for the provision of access to the identified users under the grant.					
40	[2] See Decision authorising the use of unit costs for the actions involving virtual access (http://ec.europa.eu/research/participants/data/ref/h2020/other/legal/unit_costs/unit-costs_virtual-access_infra.pdf). In case of combination of unit cost and actual costs, all the cost categories and cost items reimbursed on actual costs basis must be excluded from the calculation of the unit cost.					
41	[3] Direct costs (other than personnel) for providing access can only include: • Costs of contracts for maintenance and repair for the functioning of the installation (if not capitalised). • Costs of consumables specifically used for the installation. • Costs of contracts for installation management, including security fees, insurance costs, quality control and certification, specifically incurred for the functioning of the installation. • Costs of energy power and water supplied for the installation. • Costs of software licence, internet connection or other electronic services for data management and computing supplied specifically for the installation when they are needed to provide virtual access services. • Costs of specific scientific services included in the access provided or needed for the provision of virtual access by the installation.					
42	[4] In exceptional and duly justified cases, a different reference period can be agreed with the Commission					
43	[5] Personnel costs for the provision of access can only include costs of administrative, technical and scientific staff directly assigned to the functioning of the installation and to the generic support of users.					

44	If access costs are declared on the basis of actual cost or on the basis of a combination[6] of unit cost and actual costs, please use the following table to estimate the actual costs for virtual access.				
45	Access provision period (usually the project life-time)		from:	to:	
46	A. Estimated direct eligible costs of providing virtual access during the project life-time excluding personnel costs	Describe the direct eligible costs[7] that will be charged to the grant for the provision of virtual access over the project life-time (e.g. maintenance, utilities, consumable costs). All contributions to capital investments of the infrastructure are not eligible unless otherwise specified in the Work Programme, in which case only the portion used to provide virtual access under the action can be eligible.			Eligible Costs (€)
58				Total A	0,00
59				*of which subcontracting (A')*	
60	B. Estimated personnel direct eligible costs needed to provide virtual access during the project life-time	Category of staff[1]		Person-Months	Personnel Costs (€)
74				Total B	0,00
75	C. Indirect eligible costs 25% x ([A-A']+B)				0,00
76	D. Estimated eligible access costs = A+B+C				0,00
77					
78	[6] Personnel costs for technical and scientific staff directly working for the provision of virtual access, including for the generic support of user (e.g. help desk). These costs will be charged to the grant as direct personnel costs (hours worked for the grant must be recorded).				
79	[7] When the access policy requires the identification of users, the EU financial support to virtual access covers the access costs incurred by the installation for the provision of virtual access to the identified users under the grant.				

Figure 1: Virtual Access Calculation Spreadsheet (source: EC)

In this template, one can see two of the early models for calculating Virtual Access costs for funding:

- **Unit Cost**: This is the primary model, where the cost of the Virtual Access is calculated on the basis of the cost per unit of access, taking into account the costs that the

offering had prior to the fund periods of operation. Under this model, the provider is funded for the "units" that their service is offering in the context of the grant.

- **Actual Cost**: This is a transitional model for services that do not have enough data to estimate their unit costs. Under this model, the actual costs of personnel and other direct costs related to the offering of the service are calculated for the period of the project and get funded accordingly.

Under the Virtual Access model, taking into account the breadth of components that can be offered, the chances are that most computer services that can attract a growing user basis, can have their opportunity for an extra bit of sustainability push. It has to be noted though that the VA model of funding will only cover the costs initially planned by the provider at the time of the proposal. This means that if the Key Performance Indicators (KPIs) related to access will not be achieved, the received funding might be influenced. On the other hand, in case of overachieving of KPIs, the additional costs of increased access will not be covered.

Although Virtual Access is a great opportunity to increase the adoption of a service, it has to be noted that this is directly related to the provisioning of the access and not to the sustainability of the service (e.g. the development of new features of the software that supports a service). Virtual Access covers several types of costs, such as:

- Direct personnel costs for providing access.
- Costs of contracts for maintenance and repair for the functioning of the installation (if not capitalised).
- Costs of consumables specifically used for the installation.
- Costs of contracts for installation management, including security fees, insurance costs, quality control and certification, specifically incurred for the functioning of the installation.
- Costs of energy power and water supplied for the installation.
- Costs of software licences, internet connection or other electronic services for data management and computing supplied specifically for the installation when they are needed to provide virtual access services.
- Costs of specific scientific services included in the access provided or needed for the provision of virtual access by the installation.

It should also be clarified that the "**unit of access**" needs to be defined by the provider and maintained both in their service offering and accounting as well as in the cost claims. A common unit of access seems to be the "user". However, this is usually not enough to capture systems' utilisation. Units of access commonly seen are various definitions of "users" (e.g. identified persons, service clients, individual browsers etc), deployments, API invocations, experiments run, projects supported, documents or data artifacts deposited/published etc.

Another significant element of Virtual Access funding is that it can be obtained via participation to a call that supports Virtual Access, having the service provider as one of the beneficiaries of the respective proposal (examples of such opportunities: [5], [6]). This can be straightforward if a service is provided by a single legal entity, but it becomes complicated if a service is a joint offering of more than one beneficiary. In this case, having a legal entity that will be responsible for the service is the proper scheme for pursuing Virtual Access funding. If this is not the case, then the addition of more than one beneficiary to the proposal is needed,

adding complication to the management and separation of work in the context of the proposal.

3. Other Factor Affecting Services Sustainability and Maintainability

Apart from the financial issues, there is a number of other elements influencing the sustainability of the services. Inherent requirements and limitations of EOSC are amongst these elements.

3.1. EOSC related Factors Influencing Services sustainability

EOSC is an environment for hosting and processing research data to support EU science. The ambition of the EOSC is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation and educational purposes.

The heart of the EOSC Project is the EOSC Catalogue. It is the repository component offering the necessary programmatic interfaces for the addition, modification, and access to information regarding providers, resources and user activity collected in EOSC portal. Users can browse by scientific domain, resource category or provider and also find helpful material for the right use of the catalogue (if needed). This component offers the underlying storage functionality and the interoperability tools for the programmatic access, registration, manage (CRUD) of providers, services, and catalogues. It also offers the necessary API functionality for the interoperability of service catalogues from individual providers or aggregators (e.g., thematic, or regional catalogues) with the EOSC portal.

In order (for the catalogue) to be operational, the catalogue itself and the data as well should comply with the FAIR principles. In particular:

- F: making content findable, not just the data sources/content providers. Foresee an alert system based on the selected filters to which the user can subscribe for future notifications
- A: identifying the access type, both at the metadata level and at the content provider level (in case authentication is required)
- I: provide details on the data format and maybe usage guidelines
- R: provide information on the license type and any relevant legal information

Even though the EOSC catalogue aims to become one of the most updated and user-friendly catalogues for all of the scientific community, there are still some limitations/guidelines that should be taken into account when considering uploading products in it.

- The service to be uploaded must fall within the remit of the EOSC activities, i.e., it brings value to users and facilitates them to implement Open Science.
- It must either be an online service, or a 'human' service, such as training and consultancy.
- The service must be mature, reaching ‘Technology Readiness Level’ 8 (TRL8).

- The compulsory fields of the service description template must be filled during onboarding.

In addition to the above-mentioned limitations, we encountered the following challenges.

- EOSC catalogue is not the only repository for most of the services populating it. Most of its services derive from other catalogues, thus updates should be made at all catalogues hosting the products at the same time, to avoid duplicating different versions or information. In order to address this challenge, and according to the EOSC interoperability framework, we exploited the APIs offered by EOSC which allow the population of EOSC catalogue directly from the NEANIAS one, should the service provider wishes so [7]. Based on EOSC guidelines, this option is available only for service providers which are already registered on EOSC portal and have uploaded their first service.
- The sustainability of a service is inter alia directly related to the number of users using it. The latter, apart from the nature and the characteristics of the service, includes the way(s) a user can request and get access to the service. The investigation of the needs of NEANIAS service providers resulted that NEANIAS services should be accessible in EOSC in one of the following ways [8]:
  - Open access: The service does not need user authentication/authorisation.
  - Ordering: The service requires an access order from the user.
  - Ordering with computing resources provided by the user: In this option the user must make a compute allocation in EOSC before submitting the NEANIAS order and include the allocation endpoint in the NEANIAS order.
  - Pay-for-use ordering: This option was assessed and the conclusion was that this will not be elaborated further in NEANIAS. The providers who need this option should register their service with a 'demonstration access' level in EOSC.

In order to support the ordering requirements of NEANIAS service providers, after investigating the options provided by EOSC Order management, we employed the following approach:

- Option 1 (open access) is enabled via redirection from the EOSC Marketplace.
- For Option 2 (ordering) the access order is generated by the EOSC Marketplace (MP) and sent to the NEANIAS provider who can enable access and approve the order in the MP. To support this option we chose to develop our own ordering mechanism on the NEANIAS service management IT system in order to achieve greater flexibility on granting/customising access to NEANIAS services.
- Up to this point, Option 3 above is not available at the EOSC catalogue. The solution we have come up with is **hosting NEANIAS thematic services in EGI-ACE cloud/Kubernetes resources**. The successful validation of this use case would open a new sustainability model for thematic service providers: The thematic service providers could make arrangements with EOSC Compute Platform providers for the hosting of the NEANIAS services beyond the NEANIAS lifetime, and delivering the services as a collaborative effort of NEANIAS and EOSC Compute Platform providers.

- Additional to the above, we would also like to have a **pay-for-use ordering** option (Option 4), which is currently investigated by service providers outside of NEANIAS project.

Another important factor related to the sustainability of services on EOSC is the reuse of ready-made components/ services, especially of those already available on EOSC ecosystem. NEANIAS consortium, from the very beginning of the project, decided to integrate NEANIAS services (thematic and core) with relevant EOSC services, where we could see value of the integration (e.g. to enhance our operations, to deliver more features to users) [9]. Following this approach, we were able to exploit services that on one hand have already proved their sustainability and on the other hand users and providers are already accustomed with them. The main integrations with EOSC services we considered include [7]:

- EOSC Portal / Service Catalogue. Through the integration of NEANIAS catalogue portal with EOSC Portal, NEANIAS services are listed in the respective directory and can be located and consumed by EOSC users.
- The integration of NEANIAS AAI with EGI – Check-in for authentication and authorisation brings NEANIAS services closer to the EOSC user base providing more streamlined access from the EOSC portal catalogue of services.
- For service availability and reliability monitoring purposes, we used a monitoring service based on the ARGO system provided to EOSC by the EGI e-infrastructure federation.
- Although not mandatory for NEANIAS thematic services, we chose to pursue the integration with Zenodo as it supports several features required by NEANIAS technical management, including discovery, storage (or linking), publication and description formalization of research artifacts as well as association with formal Persistent Identifiers (DOIs).
- For helpdesk purposes, we examined the integration with the EOSC ticketing system, however we chose to develop our own ticketing system on the NEANIAS service management system in order to have greater flexibility on how incidents and service requests are handled and monitored. Moreover, this approach allowed us to follow a more homogeneous process for handling external (i.e. coming from service users) and internal (i.e. originating from NEANIAS partners) requests and issues.

3.2. Use of existing infrastructure resources

As NEANIAS is a mostly academic-oriented project, the resources to deploy and run the services came mostly from research-oriented infrastructures. Still, the complexity of deploying and supporting services (at EOSC or not) are inherent regardless of the target audience. Such complexities include infrastructure technologies (e.g., OpenStack, Kubernetes) and processes to support service management (e.g., as per FitSM).

One possible host for an EOSC service is the research institute of the service provider. In such a case however, management of the infrastructure is left completely to the system administrators of the academic institution, requiring extensive knowledge of state-of-the-art infrastructure-level technologies such as OpenStack and Kubernetes. This has proved to be

cumbersome, as system administrators of research institutes are mostly concerned with inwards looking services instead of supporting outwards looking ones, such as the NEANIAS services.

Based on our experience, **we do not encourage this approach**.

Another approach is for the service provider to obtain resources from a national research infrastructure. If this is restricted to the virtual machine level, it is clear enough that can be viable. However, if the service provider is utilizing Kubernetes for the service deployment, then someone needs to install and administer such a cluster.

Even though this can be handled by the service provider, it places the burden of installing, administering, upgrading, and monitoring the cluster solely to the service provider. This is not to be taken lightly, as it requires skills that are distinct from the usual software engineering ones that are common when implementing software-based services.

Recently, EGI has started testing a new approach to Kubernetes provisioning, and that is the notion of the managed Kubernetes cluster. The idea is that the infrastructure provider is also the sole administrator of the cluster, and that service providers are just "users" that come and get enough access to deploy and manage their own services. This considerably reduces the load of infrastructure management from the service provider and thus can be seen as a strong advantage of this approach. The things to note as issues to tackle, in this case, are the restrictions that the cluster administrator imposes. Such examples are the inability to run root-based containers, storage availability, and vertical services such as log aggregation and management. In other words, the service provider now has to adapt his/her deployment to the restrictions imposed by the infrastructure provider. **Even with its restrictions, we feel that this approach is the most promising for Kubernetes-based deployments.**

NEANIAS is participating in the pilot of this EGI experiment with managed Kubernetes clusters, hosted by CESNET. We are hopeful that after this pilot, this model will spread across all EGI members and become ubiquitous.

3.3. Governance of common/shared infrastructure – Synergies/Collaborations

Providing a public service (let's call this 'S') at EOSC, is a complex process that will inevitably force the service provider to rely on third parties. Reliance can be direct, such as the infrastructure where S is deployed, or other services that S cannot function without, such as identity management, authentication, authorization, storage, log management, etc. It may also be indirect, such as ticketing and helpdesk systems, monitoring systems, etc.

Inevitably, this complicates things when it comes to properly administering S and having its users notified whenever issues arise. One example is downtime management: service S will be unavailable not only when its service provider needs to bring it down for maintenance (such as upgrades) but also any time one of its direct dependencies (say D) is unavailable. This is further exasperated when S's dependencies are managed by separate administrative domains than S.

Ideally, the service provider of service D, which is a direct dependency of S, will need to forward-warn the service provider of S about any planned downtime. This should be clarified

once the engagement of D by S is initiated, for example by studying D's Service Level Agreement (SLA).

Furthermore, announcement of such downtimes should be automated, so that S can reliably notify its users about planned unavailability periods. However, in our experience, this domain is still not mature enough and not completely automated, requiring manual effort to plan, announce, and account for planned downtime. This inevitably spans any employed monitoring system, because such a system will need to be informed about planned downtime so that it can accurately report service S's availability and reliability.

The above examples highlight the need for 'communication' between the service providers of S and D. However, quite often said service providers are using separate messaging platforms (e.g. Slack vs Teams) and different ticketing systems (e.g., Jira vs Redmine). This complicates things even more, as 'tickets' raised against one service, now need to be manually propagated across different systems to allow the target service provider to gain knowledge of said issue and process it accordingly. And, for synchronous communication of separate teams, engineers often need to resort to neutral third-party vendors just to allow collective consultation on addressing issues. At the NEANIAS project, due to its vast breadth, we have witnessed all of the above issues and we feel the need to warn future service providers about these, so they can prepare accordingly.

We are hopeful that in the future this domain (ticketing and group communication) will be standardized and automated, so that manual labor for such mundane tasks can be minimized. However, we are still not there yet.

3.4. Availability of Data

In EOSC landscape, **data are the core** based on which other elements are built. Taking its current definition [9] one can see that data is mentioned repeatedly in EOSC with a central role:

> *The European Open Science Cloud (EOSC) is an environment for hosting and processing research data to support EU science.*
>
> *The ambition of the European Open Science Cloud (EOSC) is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation and educational purposes.*

The conclusions derived from NEANIAS experience with data availability are mixed.

In the **SPACE research sector,** the Virtual Observatory (VO) ecosystem of standards and technologies, discussed and defined by the International Virtual Observatory Alliance (IVOA), has, since the setup of the Alliance itself (2002), focused on interoperability as a way to provide discovery of and access to distributed resources in the astrophysics and planetary science research domains and to enable re-use of those same resources. Therefore, VO standards provide a direct support of FAIR principles and, adding the fact that they are open in definition and governance, allow for an Open Science scenario in SPACE. VO also pre-dates

the FAIR principles formalisation [11] by a timespan long enough that let its architecture grow mature.

The NEANIAS SPACE community worked in close co-operation with the H2020 ESCAPE Project toward FAIR data and services. Integration of the IVOA Registry of Resources in EOSC has happened with the EUDAT B2FIND service harvesting, through the OAI-PMH protocol, DataCite metadata from the IVOA Registry. Work is ongoing towards adding more metadata, using the B2FIND metadata schema at harvesting, instead of DataCite only. The mapping from VOResource metadata to DataCite is not complete, the most important differences being about mapping per-protocol access URLs and tablesets. The Unified Astronomy Thesaurus top level concepts have been mapped to VOResource subject keywords. For more information please see [12].

In the **Underwater research sector**, particular difficulty has been encountered regarding the availability of multispectral multibeam echosounder data, utilized in service UW-MAP. The difficulty stems from the limited usage of the corresponding echosounder systems, as this technology (multifrequency survey mode) has been introduced only very recently. In order to overcome this problem, multispectral echosounder data publicly available in the context of a multispectral seabed classification challenge (R2Sonic Multispectral Challenge) have been used as demo and validation data of the service. Additionally, compatible data from previous surveys have been provided from partners of the project, in combination with data acquired within the duration of the project. A similar approach has been followed for the UW-MOS service.

In the **Atmospheric research sector** data availability varies. While there are significant volumes and data coming from earth observation missions, openly available and via well-established standards and sustainable services to conduct various analysis and visualization scenarios, specialized data required by services such as ATMO-STRESS and ATMO-FLUD are scarce. This scarcity is complemented with the lack of well-defined standards for the data in question, proprietary CSV schemas being the main means of packaging.

In the course of examining data availability NEANIAS work team concluded on a few reasons and counter measures for them:

- The **lack of open data** (sharing) culture in particular communities, as those have not transitioned to the Open Science culture. This is an area where NEANIAS and other EOSC initiatives contribute to, by onboarding services and communities and showcasing the benefits of FAIR data approaches.
- The **lack of standards** is a blocking situation as it limits the usability of data and does not provide an ease to employ framework for sharing them. Though this is a loop problem, as need and desire to share data will also push towards the adoption of de facto and formal standards. NEANIAS provides documentation for its data formats, to facilitate their adoption by its service users, while adopting well known standards from other sectors, where feasible, or proposing very simple to conform to data manifestations.
- Issues that have to do with **data privacy** come into focus in specific services where the data payload may represent formations or artifacts owned or managed by authorities or 3rd parties. This may be the case even for a seemingly public artifact,

such as an archaeological site. A long process of data clearance may be involved and the resulting dataset may be still under significant reuse restrictions.

- **Data quality** is another limitation for the availability of data. While data may be present there might be not enough evidence for their quality or appropriateness for specific usages. In this direction, NEANIAS has discussed on the need and relevance of data assets peer-reviewing and explicit referencing of those assets by publications, which is along Open Science's directions of evolution. The recognition of the value of data as a research product is being shaped and this will vastly improve the trustworthiness of data assets in the years to come, disclosing not only their quality but also the appropriateness for a particular case, via adequate provenance and licensing metadata.

Elaborating further on the availability of data, we see that there are a few directions that work could be done for improving this availability, beyond the ones already mentioned, which span outside NEANIAS domain of authority. These could include:

- The establishment of frameworks that would promote the relaxation of rules and costs for data utilized for academic or research or other not-for-profit purposes.
- Rewarding of data publication, via a number of means depending on the sector in case.
  - For instance, in the academic sector data referencing, peer-reviewing and inclusion in research performance and impact factors' establishment would be of significance for their publisher. This is an area where basic CC licenses fall rather short, not due to their omissions but due to lack of processes and instruments in the research publication ecosystem.
  - For shared infrastructure access, data sharing could translate into resources or other benefits etc. E.g. FAIR Open Data hosting could be free of charge etc.

As NEANIAS services require data for their operation and also have to conform with the aforementioned limitations and culture of the various domains in focus, several approaches have been followed depending on the case:

- **Utilisation of established open data repositories**: services may depend on well established, open and accessible data repositories that their availability is unquestionable and long proven. This holds, for example, in the case of space sector datasets and open earth observation data.
- **Bring your own data**: services requiring other data to operate, do not link to data sources for those data, but request the user to point to the dataset in question for performing a particular task.
- **Keep your own data**: Although NEANIAS services promote data sharing and offer the tools for doing so in a well established manner (i.e. documented data formats and Zenodo Publishing) they do not oblige the user to do so. Sharing data under a license the user picks is left to user's discretion, lifting a barrier for service adoption. It also guarantees that data are kept private when processed and visualised.
- **Sample data offering**: requiring users to bring their own data for services to operate is essential, however it would hinder the ability of services to attract users and prove their concepts. To overcome this, NEANIAS services supply their users with quality

sample datasets coming from field measurements, that allow one to fully comprehend both the way to feed the service and to understand its benefits.

3.5. Security and Privacy

The security and privacy aspects of a service integrating with the EOSC ecosystem has been elevated to one of the cornerstone requirements. The typical aspect of this requirement is evident from the fact that all service providers need to have available, early in the process of integrating with the EOSC ecosystem, a comprehensive and full declaration regarding their Privacy Policy and Terms of Use. The essence of the requirement is only manifested with the availability of the documents as they mostly enumerate a set of technical, operational and management related requirements, approaches and best practices that must be implemented and adhered to, thought the lifetime of the service and each user request cycle.

Focusing on some of the highly relevant and important aspects enabling the secure and privacy respecting operation of a service, we consider the following coarse grain areas of functionality, typically required by most services, and their impact on the sustainability of the services:

- **Authentication and Authorization Infrastructure (AAI)** – The infrastructure enabling the authentication of individual users and services by a trusted party, typically enabling federation of users
- **Logging** – The ability to trace service request progress, operation, informational aspects on the behaviour of the service, debugging context as well as recoverable or critical issues that the service is experiencing while operating
- **Accounting** – The high-level tracking of accountable actions, as defined by the service provider, be it resource utilization, service actions or distinct operations

Looking into one of the core issues a service provider needs to address, the process of *authentication* is at the heart of the security aspects being also part of the privacy consideration. One of the driving incentives for a service to be part of the EOSC ecosystem is the expansion of its user base. The EOSC ecosystem has provided the means for service providers to make the most of the multiple identity providers users may utilize in order to be authenticated, alleviated a lot of the technical burden to remain compliant across protocols and providers, linking and federating users, handling many of the privacy considerations relevant to the user identity management and, at the end, lifting the barrier for user authentication. In parallel to that, given the adoption and compliance of the EOSC provided authorization mechanisms to the GDPR mandates, it drives and helps the integrating service providers towards this direction too, enforcing practices and raising awareness for the aspects of security and privacy that they need to comply with. This is also an important step towards the sustainability of the services, as GDPR (General Data Protection Regulation) compliance is an essential step towards sustainability bests practices compliance. Early integration with the facilities offered through EOSC as peripheral or even core EOSC services has proven valuable within the NEANIAS services, has been a validation of the processes and security aspects of the internal service management system and is something to be pursued early in the onboarding process, as this will set the security and privacy standards on par with the EOSC

ecosystem in this aspect. Utilizing the potential offered by reusing existing institutional, corporate, or even social user accounts increases the sustainability of the service while making it more accessible to the end user. In addition to this, the ability offered to associate further attributes with user identities at the identity provider level and the possibility to forward these attributes downstream to the resource providers, gives additional flexibility on how to handle new user registrations, while it also increases the trust the resource providers have on the users, potentially elevating their access, depending on these attributes. This is a major gain for service providers and their service management and access granting procedures.

The authentication process is the starting point for securing access to the available resources though the authorization policies that can be enforced by service providers. It is often the case that a service developed to respond to a scientific need, addresses primarily the core challenge but leaves other aspects, such as authorization, not properly handled. In the course of the service operation this unavoidably leads to sustainability issues and even security breaches. A secure service increases user trust and retainment, knowing their operations, data and processes are not in danger of being leaked or tampered with. Highlighting some of the risks of not employing proper authorization within a service, from the aspect of being able to sustain the service based on its intended usage within limits set by the enabling hosting infrastructure, especially one that may be utilizing shared resources as is common in the EOSC ecosystem, one can distinguish the following:

- Malicious usage of resources (e.g. attacks to 3rd parties)
- Misuse of resources (e.g. mining, non-environmentally correct usage, long term allocation of temporary storage with unused files etc).
- Unfair usage of limited resources without ability to account for or manage the usage (heavy use of resources by specific users)
- Protection of resource provider from unexpected / unintended costs (e.g. too much network traffic or storage or power or cooling or VMs etc).
- Prevent usage of resource in non-compliant contexts (e.g. commercial exploitation of academic resources).

Another aspect of security and privacy that we have found within NEANIAS to have served as for the sustainability and ease of management of our services is that of centralizing the authentication and authorization process (where it could be applied) within the boundaries of an internally managed NEANIAS AAI service, federating access with identity providers (social and EOSC ecosystem) and granting needed access though our Service Management System. This has enabled us to be more agile in changes when and where required or deemed appropriate. For example, GDPR and other regulatory or contractual restrictions may impose restrictions on what processing can be made, by whom and where data and processing may be located. Such restrictions of course may change the value chain and would require additional considerations.

The simplification of troubleshooting a service behaviour is essential for lowering its cost and offering a high quality of service to its clients. The quality of the software and the infrastructure that underlies it form the foundation of this aspect, however as several factors and unforeseen issues are commonly found to impact software behaviour, extensive logging is an essential instrument that lowers the cost of operating a service. On the cloud ecosystem,

where transient resources can be utilised at low costs, logging is no longer just an action of the software, but requires a *logging service*, i.e. infrastructure of its own to collect data and information from several components. Delivering such a system that can organise at a central point of reference all the logs of infrastructure resources is an almost mandatory requirement for a sustainable service offering.

Authentication is not essential only for security. Authentication is also a prerequisite for enabling a service into a marketable product and linking its usage to a revenue stream, as well as to a cost center. Authentication of a user is a prerequisite so that the actions of a user can be mapped to usage of resources and be associated to costs that incur from this usage.

In the same direction, nowadays it is quite common that usage is the main element behind service charging, under the model of "pay-as-you-go" where a client is charged depending on the usage one performs. Depending on the pricing model of the offering, details of the usage may be utilised, and here is where the *accounting component* comes into play to support a fair and sustainable pricing model. The accounting component could also support the application of quotas and gradual restriction of resource consumption by authenticated users. Authentication and its several details may be able to support institutional allocation of resources and cost / charges linking, but it is security that safeguards the business and sustainability rules and prevents an actor to exceed one's boundaries in utilising resources.

4. Conclusions

Although financial aspects are the main elements affecting services sustainability on EOSC, there are several other factors also contributing to it.

In this white paper, all the factors influencing services sustainability on EOSC were described while best practices per factor, derived from the experience gained during NEANIAS project, were provided.

Funding schemes for both services and research infrastructures were identified and analysed. **Virtual Access** was discussed as one of the most promising funding schemes for usage. However, it should be noted that although Virtual Access helps providers increasing services adoption, it is directly related to the provisioning of the access and not the sustainability of the service.

Fortunately, the sustainability of a service is inter alia directly related to the number of users using it as well as the way(s) a user can request and get access to the service. In NEANIAS, several ways were examined such as **open access**, ordering, ordering with computing resources provided by the user and pay-for-use ordering. The first two ways are enabled by EOSC. For the third option, NEANIAS thematic services were hosted in EGI-ACE cloud/Kubernetes resources. The pay-for-use ordering option is currently under investigation.

The **reuse of ready-made components/ services**, especially of those already available on EOSC ecosystem, was discussed as an important factor related to the sustainability of services on EOSC. It should be highlighted that NEANIAS consortium, from the very beginning of the project, decided to integrate NEANIAS services (thematic and core) with relevant EOSC services that have already proved their sustainability and/or are familiar to users and providers.

However, providing a public service at EOSC, is a complex process that will inevitably force the service provider to rely on third parties. Inevitably, this complicates things when it comes to properly administering the service and having its users notified whenever issues arise. Thus, there is a need for 'communication' between the respective service providers. It was stressed that the use of separate messaging platforms and different ticketing systems further complicates things. **Standardization and automation of ticketing** and group communication was proposed as a means to minimize the manual labour for such mundane.

The **availability of data** per thematic sector investigated in NEANIAS was then discussed. The reasons behind the lack of data were analysed and measures to address them were proposed. Lack of open data (sharing) culture, lack of standards, data privacy, and data quality were some of the reasons discussed. The establishment of frameworks that would promote the relaxation of rules and costs for data utilized for academic or research or other not-for-profit purposes as well as rewarding of data publication, via a number of means depending on the sector in case were two proposals in order to improve data availability.

The **security and privacy** aspects of a service integrating with the EOSC ecosystem have been presented since these have been elevated to one of the cornerstone requirements. It was deduced that reuse of existing institutional, corporate, or even social user accounts increases the sustainability of the service while making it more accessible to the end user. The ability offered to associate further attributes with user identities at the identity provider level and the possibility to forward these attributes downstream to the resource providers, was shown to improve the flexibility on how to handle new user registrations increasing the trust the resource providers have on the users, potentially elevating their access, depending on these attributes.

Another aspect of security and privacy found within NEANIAS to have served as for the sustainability and ease of management of our services is that of centralizing the **authentication and authorization process** (where it could be applied) within the boundaries of an internally managed NEANIAS AAI service. In addition, the simplification of troubleshooting a service behaviour is essential for lowering its cost and offering a high quality of service to its clients. Delivering a logging system that can organise at a central point of reference all the logs of infrastructure resources was proved to be an almost mandatory requirement for a sustainable service offering.

Authentication was also proposed as a prerequisite for enabling a service into a marketable product and linking its usage to a revenue stream, as well as to a cost center. It should be highlighted that nowadays, usage is the main element behind service charging, under the model of "pay-as-you-go" where a client is charged depending on the usage one performs.

This white paper can be proved very useful for service providers onboarding their services on EOSC as well as to decision/policy makers and other stakeholders of EOSC ecosystem helping them to address several challenges and thus improve their services' sustainability.

References

[1] Decision authorising the use of unit costs for the actions involving virtual access [http://ec.europa.eu/research/participants/data/ref/h2020/other/legal/unit_costs/unit-costs_virtual-access_infra.pdf]

[2] DECISION authorising the use of unit costs for the costs of providing trans-national and virtual access in Research Infrastructures actions under the Horizon Europe Programme (2021-2027) and the Research and Training Programme of the European Atomic Energy Community (2021-2025) [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/unit-cost-decision-research-infrastructures_horizon-euratom_en.pdf]

[3] H2020 Programme - AGA – Annotated Model Grant Agreement, Version 5.2, 26 June 2019 [https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf]

[4] Template for Virtual Access Calculation of https://ec.europa.eu/research/participants/data/ref/h2020/other/legal/templ/h2020_tpl-calc-unitcost-va-infra_en.xls

[5] Research infrastructure services to support health research, accelerate the green and digital transformation, and advance frontier knowledge (2021) (HORIZON-INFRA-2021-SERV-01) [https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-infra-2021-serv-01-07]

[6] Horizon Europe Work Programme 2021-2022 - Research Infrastructures (European Commission Decision C(2022) 2975 of 10 May 2022) [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2021-2022/wp-3-research-infrastructures_horizon-2021-2022_en.pdf]

[7] NEANIAS project, Deliverable D8.6: Best practices for service onboarding the EOSC hub

[8] NEANIAS project, Deliverable D8.4: EOSC integration activities report (interim)

[9] NEANIAS project, Deliverable D8.1: EOSC integration plan

[10] EOSC Portal, about EOSC (Fall 2022) [https://eosc-portal.eu/about/eosc]

[11] Wilkinson, M. D. et al., “The fair guiding principles for scientific data management and stewardship,” Scientific Data 3 (2016)

[12] Molinaro, Marco, et al. "Supporting FAIR Principles in the Astrophysics Community: the European Experience." arXiv e-prints (2021): arXiv-2111.

List of acronyms

Acronym	Description
AAI	Authentication and Authorization Infrastructure
API	Application Programming Interface
CF	Cohesion Fund
CSV	Comma-separated Values
EAFRD	European Agricultural Fund for Rural Development
EC	European Commission
EFSI	European Fund for Strategic Investments
EMFF	European Maritime and Fisheries Fund
EOSC	European Open Science Cloud
ERDF	European Regional Development Fund
ESA	European Space Agency
ESF	European Social Fund
ESI	European Structural and Investment
EU	European Union
GDPR	General Data Protection Regulation
GSRI	General Secretariat for Research and Innovation
IaaS	Infrastructure-as-a-Service
IVOA	International Virtual Observatory Alliance
KPI	Key Performance Indicator
MP	MarketPlace
OSI	Open Science Iniatiatives
PaaS	Platform-as-a-Service
SLA	Service Level Agreement
SME	Small and Medium Enterprise
TRL	Technology Readiness Level
VA	Virtual Access
VM	Virtual Machine
VO	Virtual Observatory