H2020-INFRAEOSC-2018-3

NI4OS-Europe

National Initiatives for Open Science in Europe

Deliverable D6.5
Open call specification

Lead beneficiary(s):	GRNET (editor)
Author(s):	Evangelia Athanasaki, Nana Anastasopoulou, Andreas Athenodorou, Chrysovalantis Constantinou
Status –Version:	Final – h
Date:	March 22, 2022
Dissemination Level:	Public

Abstract: Deliverable D6.5 – Open Call Specification is a report on the implementation of the pan-European open call of NI4OS-Europe project, the process and the objectives of the call, and the available on-boarded EOSC services offered through the call.

The NI4OS-Europe Consortium consists of:

GRNET SA	Coordinator	Greece
ATHENA RC	Beneficiary	Greece
CYI	Beneficiary	Cyprus
UCY	Beneficiary	Cyprus
IICT	Beneficiary	Bulgaria
SRCE	Beneficiary	Croatia
RBI	Beneficiary	Croatia
KIFU	Beneficiary	Hungary
DE	Beneficiary	Hungary
ICI BUCURESTI	Beneficiary	Romania
UEFISCDI	Beneficiary	Romania
ARNES	Beneficiary	Slovenia
UMUKM	Beneficiary	Slovenia
IPB	Beneficiary	Serbia
UOB	Beneficiary	Serbia
RASH	Beneficiary	Albania
UNI BL	Beneficiary	Bosnia-Herzegovina
UKIM	Beneficiary	North Macedonia
UOM	Beneficiary	Montenegro
RENAM	Beneficiary	Moldova (Republic of)
IIAP NAS RA	Beneficiary	Armenia
GRENA	Beneficiary	Georgia

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
27.12.2021	a	Evangelia Athanasaki, Nana Anastosopoulou	Initial ToC defined
31.01.2022	b	George Artopoulos, Theo Christoudias, Zoe Cournia, Bojana Koteska, Andreas Athenodorou	Scientific fields defined
10.02.2022	c	Evangelia Athanasaki	Editing
20.02.2022	d	Andreas Athenodorou	Editing
25.02.2022	e	Branko Marovic	Provided contribution
05.03.2022	f	Andreas Athenodorou	Editing
18.03.2022	g	Evangelia Athanasaki, Ognjen Prnjat, Andreas Athenodorou	Final Editing
22.03.2022	h	Milutin Radonjic	Quality control check
06.04.2022	h	Evangelia Athanasaki, Andreas Athenodorou	Final Corrections

Table of contents

References

[1] Project NI4OS-Europe-857645 - Annex I - Description of the Action

[2] D6.3 "User communities' testcases"

[3] D6.4 "Service evaluation by user communities"

[4] D5.2 "First report on provider and repository integration"

[5] D5.3 "Update of service catalogue"

List of Figures

List of Tables

List of Acronyms

EC	European Commission
EOSC	European Open Science Cloud
ERA	European Research Area
ERAC	European Research Area and Innovation Committee
ESFRI	European Strategy Forum on Research Infrastructures
EU	European Union
EuroHPC JU	European High-Performance Computing Joint Undertaking
SEE or SE Europe	South East Europe
SEE-GRID	South Eastern European Grid
WP6	NI4OS-Europe Work Package 6: User engagement, training and demonstrations

Executive summary

What is the focus of this Deliverable?

The focus of this deliverable is to provide the details of the pan-European open call of the NI4OS-Europe project, targeted to offer the services on-boarded to EOSC to pan-European users. Within the scope of this deliverable, we provide the process and the objectives of the call, the description of the on-boarded services, as well as the available computational and storage resources provided via the generic services which have been on-boarded in the EOSC.

What is next in the process to deliver the NI4OS-Europe results?

The documentation presented in this deliverable will be used as a basis for providing the final report of the open call in deliverable D6.8 as per DoA.

What are the deliverable contents?

The deliverable describes the implementation of the open call for proposals of projects accessing NI4OS-Europe services provided to EOSC. These include thematic services, repositories, as well as generic services. Regarding the generic services, a detailed analysis of the resources provided for the open call is presented. This deliverable consists of all the information relevant to the implementation of the call. The description starts with a short introduction to the basic details of the call. Section 2 provides a summary of the adopted procedure and the objectives of the call. Section 3 provides all the information on the on-boarded services.

Conclusions and recommendations

This call provides a framework for scientists to use all the services on-boarded at the NI4OS-Europe service catalogue. This has a twofold purpose. Firstly, throughout the use of the services, NI4OS-Europe will ensure that all the services are production level and that their on-boarding is of high quality. In addition, it will provide the opportunity to European researchers to perform their work, extract and publish results within the framework of European Open Science Cloud. This deliverable reports on the preparatory activities conducted within task T6.5 "Open call support" to ensure the smooth progress of the open call.

Final report on the results of the review process of the submitted applications, which have been given access to computational resources, storage resources, and NI4OS-Europe services will be available in D6.8 "Report on the open call".

1. Introduction

This deliverable describes the setup and implementation of the pan-European open call for proposals for projects accessing NI4OS-Europe services, which are available in the framework of the European Open Science Cloud. This document will explain how researchers can get access to projects that require the use of NI4OS-Europe on-boarded services. This will enable the NI4OS-Europe consortium to test the usefulness as well as the quality of the onboarding of the associated services and at the same will enable researchers to create scientific output. The services which can be used are thematic services, generic services, as well as repositories. Throughout the open call, all researchers participating are expected to harmonize the management of their data according to the FAIR principles. In this way, they will practically perform Open Science.

The design and implementation of the call have been based on the outcome of deliverables D6.3 "User communities' testcases", D6.4 "Service evaluation by user communities", D5.2 "First report on provider and repository integration" and D5.3 "Update of service catalogue" where cross-border use-cases and their mapping to generic, thematic and repository services were provided, the first set of on-boarded services and their evaluation by the user communities were reported, as well as the updated service catalogue was presented.

The defined framework is being used in the context of the NI4OS-Europe project for providing access to users of the NI4OS-Europe thematic and repository services as well as the associated resources of the generic services.

The NI4OS-Europe service catalogue is composed of several types of thematic services supporting the broad scientific community. Although NI4OS-Europe has a highly interdisciplinary and cross-disciplinary character, it emphasizes the importance of thematic services in the four flagship scientific fields of Life Sciences, Climate Sciences, Digital Cultural Science and Computational Physics, which can provide solutions to timely problems. Nevertheless, NI4OS-Europe open call will also provide access to projects and thus thematic services belonging to other scientific topics. The generic services provide access to multiple types of computational services such as HPC, Cloud, storage services and underlying resources. Finally, repositories can be either institutional or generic. All those services have different characteristics in terms of availability and therefore access modalities.

All researchers can access seamlessly all the thematic services as well as repositories. However, accessing generic services such as HPC resources and storage is constrained by limitations related to the availability of the institutions owning the services.

The framework for the service provision is based on the following principles:

- To make services and resources provided by NI4OS-Europe service catalogue available to as many as possible researchers from the European region as well as the associate countries ensuring open, FAIR and unbiased access of services.
- To promote the usage of the underlying generic services provided by the NI4OS-Europe resource provider partners to the whole European region.
- To promote scientific collaboration and exchange of know-how between the research groups following FAIR principles.

- To open up the knowledge and data produced in the region to all researchers in Europe and beyond where possible.
- To provide the opportunity to researchers of all countries to have access to the EOSC on-boarded services offered by NI4OS-Europe.

All the above principles have been used to implement the NI4OS-Europe call.

2. Summary of the process and objectives of the calls

NI4OS-Europe offers a broad set of generic, thematic services as well as repositories in the European region, with the special focus on the scientific communities of Life Sciences, Climatology, Digital Cultural Heritage and Computational Physics. Generic services are in the areas of Compute resource provisioning (HPC and Cloud), Storage, and Data services provisioning. Repositories provide access to datasets, Software, as well as scientific manuscripts.

The call is addressed to scientists and researchers that work in academic and research institutions in EU member and associated countries.

Among all the proposals 80% should address open research topics in specific fields of Life Sciences, Climate research, Digital Cultural Heritage, and Computational Physics and 20% on any other scientific fields.

Via this call, NI4OS-Europe opens possibilities for scientists to access the advanced services on-boarded to EOSC (thematic, generic and repositories). The services provided have been tested in the use-cases ran by scientific teams that were assigned to this task. The services were then fine-tuned according to the feedback received from the demonstrators and now can be further tested via the open call.

The list of services and resources offered by the NI4OS-Europe project can be found at: https://catalogue.ni4os.eu/?_=/resources.

Access to underlying computational resources (generic) can be awarded for a maximum period of 6 months, while access to underlying storage resources may be provided for up to 8 months.

The call enables researchers from Europe to obtain access to the advanced services which have been on-boarded on the European Open Science Cloud with the support of NI4OS-Europe. At the same time, all research teams participating in the call will be practicing the FAIR principles throughout the whole lifetime of their work, thus, engaging with the rules of Open Science.

2.1. Applicable scientific fields

As explained in the previous section, 80% of the projects participating in the open call would belong to the flagship scientific communities while the other 20% include projects in any other scientific domain. Eligible projects in flagship scientific communities are the ones that address one of the following scientific and/or societal challenges:

In the field of Life Sciences:

- LS Area A: Modelling and Molecular Dynamics (MD) study of important drug targets.
- LS Area B: Computer-aided drug design.
- LS Area C: Analysis of Next Generation DNA sequencing data.
- LS Area D: Synchrotron data analysis.
- LS Area E: Image processing for biological applications.

In the field of Climate Research:

- CR Area A: Regional climate modelling to better understand and predict climate change and impacts, and phenomena such as dust storms.
- CR Area B: Air quality modelling, including atmospheric chemistry and air pollution transport.
- CR Area C: Weather forecast and extreme weather prediction, model development, application.

In the field of Digital Cultural Heritage:

- DCH Area A: Online services and access to repositories to enable studies of the cultural heritage assets in the region (e.g., searchable digital libraries; with support of meta-data and OCR for Latin characters).
- DCH Area B: Online visualization tools to drive breakthrough contributions to heritage enquiries (e.g., interactive visualization viewers of 3D digital assets with metadata integration, remote sensing datasets, agent-based model simulations).
- DCH Area C: Training material and Tools for the creation of interactive and immersive environments to support the GLAM industry (galleries, libraries, archives and museums) with limited resources to generate virtual exhibitions of their collections of artefacts.

In the field of Computation Physics:

- CP Area A: Computational fluid dynamics.
- CP Area B: Computational particle physics.
- CP Area C: Molecular dynamics.
- CP Area D: Electronic structure methods.
- CP Area E: Computational electromagnetics.
- CP Area F: Lattice gauge theories.
- CP Area G: Astrophysics.

The criteria for the evaluation of projects for accessing the available resources where:

- Scientific excellence.
- Scientific and/or social impact of the proposed research.
- The need for usage of the selected services and resources.
- The ability to provide project results (mainly data sets, but also services and software) as services for other future VRE users.
- Maturity and experience of the principal investigator and his/her team in the research field, as well as in using the selected resources and services.
- Feasibility of the project based on the technical evaluation and the availability of resources.
- Potential for collaboration among scientists in more than one eligible country for this call.
- Relevance with the above-mentioned scientific communities will be considered an asset.
- Gender balance will be taken into account.

The criteria above will be used for the evaluation of proposals during the review process.

NI4OS-Europe aims at a balanced provision of resources to the whole spectrum of scientific fields between the four target communities that this call addresses, as well as to as many as possible countries in the European region.

2.2. Eligibility

Eligible applicants (as Principal Investigators) are scientists affiliated with academic or research institutions in the following countries (in alphabetical order): EU member states: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and associated countries: Albania, Armenia, Bosnia & Herzegovina, Faroe Islands, Georgia, Iceland, Israel, Kosovo, Moldova, Montenegro, Morocco, North Macedonia, Norway, Serbia, Switzerland, Tunisia, Turkey, Ukraine, United Kingdom.

Collaborators in proposals might reside in any country provided that no specific geographical restrictions apply for access by the corresponding centres that offer resources in the various resource-providing countries.

Industrial partners may participate only as collaborators in a proposal that led by academic or research institutions from the eligible countries, and only if the aims and objectives of the project is open research with results to be published in research journals or conferences.

Applicants should commit to using the resources that will be allocated to them, as well as to providing reports of their work based on the proposed schedule (see below). Furthermore, scientists should acknowledge the use of the NI4OS-Europe services in all publications presenting results obtained from using the allocated resources.

Finally, all the projects participating in the open call should be EOSC compatible, meaning that all key components should be EOSC, all data and software used and produced throughout the lifespan of each project have been curated based on FAIR principles and lastly, resulting publications are being published in openly accessible journals. Therefore, the applications should reflect on the FAIR-ness of their outputs and intended or already used inputs, as well as the licensing model that they intend to apply on the produced data or other outputs. This should be done in a way that is in line with the requirements of the primary funding of the conducted research (if exists) but also the basic EOSC rules and FAIR principles. The projects that will be selected to participate to the NI4OS-Europe open call will be also asked to participate to the Pilot on Open Research Data in Horizon 2020 and submit a DMP, describing how their data will be managed.

Providing a draft idea of the research data management (RDM) process applied, will be appreciated during the evaluation of applications. Applicants are encouraged to consult the available training material and tools that are made available by NI4OS-Europe. Training material on FAIR data, which has been translated in mother tongues for all NI4OS-Europe countries, can be found here: https://training.ni4os.eu/course/view.php?id=52. When it comes to the use of

supporting tools, the applicants may state whether they regard tools such as [LCT](#) and [RoLECT](#) (described in 3.3.6) suitable, or – although this is not mandatory - if they have already used these tools to perform a preliminary licence clearance analysis or to draft a legal and ethics compliance report.

2.3. Application process

All proposals will be submitted electronically via the NI4OS-Europe survey tool.

The application form will also be available in a pdf format in order for applicants to have the full list of questions available. The applicants have to fill in the online form for their applications to be taken into account.

Support to applicants will be provided during the call: namely, NI4OS-Europe Access Team will be available to answer questions while the call is open.

2.4. Evaluation process – assignment of resources

All project proposals collected during the open call will undergo a technical and scientific review to determine the eligibility and suitability of applications for the requested services. Applications requesting HPC/Cloud/Storage resources will also be reviewed by independent scientific reviewers from the region, considering conflict of interest and fairness issues. Applications not requiring HPC/Cloud/Storage resources will undergo a more lightweight review from scientific community leaders or other scientists.

After the review process, the NI4OS-Europe access committee will prioritize the applications based on the criteria set in the scope and criteria of access of the call.

All applicants will be notified of the final results of the evaluation. Successful applicants will receive further details regarding the resources and the process to obtain user accounts for granted resources. The project will fully support the usage of provided services via the NI4OS-Europe helpdesk.

2.5. Important dates

2.5.1. Important dates for the Open Call

In the following list, we provide all the important dates concerning the call of applications for access to NI4OS-Europe services.

- Opening date: 11th of April 2022.
- Closing Date: 11th of May 2022.
- Clarification provided by applicants if needed: 13th of May 2022.
- Participation decision: July 2022.
- Generic service allocation Start Date of awarded proposals: July 2022.
- Allocation end date of award: January 2023 for computational services and March 2023 for some data services.
- Final report from successful projects: February 2023.

3. Available services

Available services within the NI4OS-Europe are categorized into repositories, thematic and generic services. Repositories group various literature and data collections that hold and preserve scientific information. Generic services provide generic capabilities and address technical needs common to multiple research areas, while thematic services are research community-specific services that provide value to the corresponding research groups.

Following the on-boarding procedure and plan described in the deliverable D5.1 – "Provider landscape analysis and provider categorization", up to now, we have recorded 92 services within the Agora catalogue, out of which 33 are repositories, 31 thematic, and 20 generic. Depending on the integration level with the project pre-production environment, all services recorded within the catalogue are further divided into on-boarded, candidate, and registered services. Such distribution is illustrated in Figure 1, and in the open call, we will offer only services marked as on-boarded or candidates. Practically, in the open call, we will offer 16 repositories, 21 thematic services, and 14 generic services.
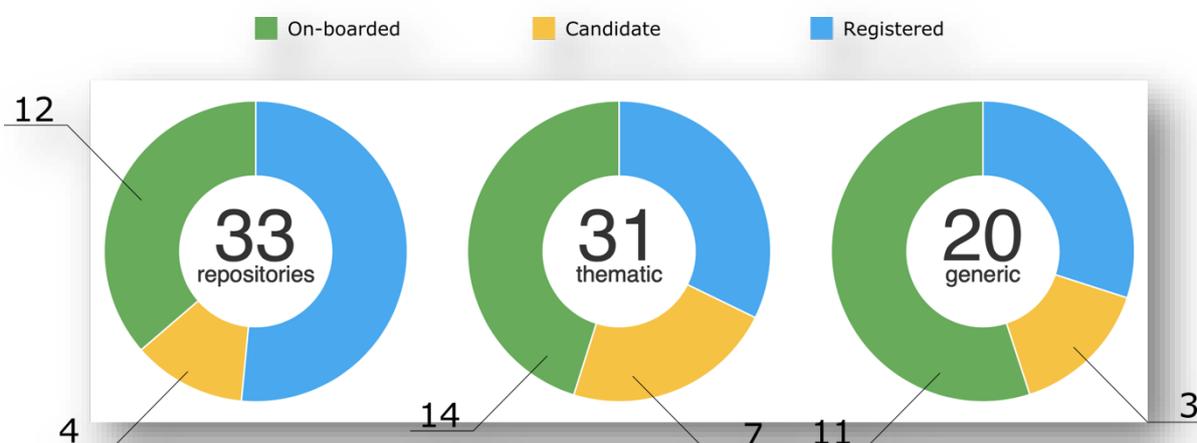
Figure 1: Distribution of available services registered within the catalogue

As described in the deliverable D5.2 – "First report on provider and repository integration", our on-boarding strategy included the scientific impact of the particular resources assessed within the WP6 and reported in the deliverable D6.3 – "User communities' test cases". Therefore, services in on-boarded and candidate status provide the capability to realize the scientific workflows created by WP6 scientific leaders and reported in the deliverable D6.3.

3.1. Available generic services

In the group of generic services, we have included services able to provide generic capabilities and address technical needs that are common to various research areas. The

services from this group support different aspects of the data lifecycle, from creation to processing, analysis, preservation, access, and reuse. Here we have generic HTC/HPC computing, Platform-as-a-Service (PaaS) solutions, generic storage, data preservation/curation/provenance, cloud containerization, and orchestration services, etc., and advanced data services discovery, management, processing, workflow management, etc.

We will provide access to 14 generic services, grouped into HPC, Cloud, data processing, storage, and other resources in the open call. These services are presented in the next five subsections. In addition, we provide a link to the catalogue page for each service that gives further information following the latest EOSC profiles specification.

3.1.1. HPC services

Up to now, the total amount of HPC services marked as on-boarded or candidate in the NI4OS-Europe catalogue is 5. All these will be offered to the researchers in the open call.

In addition, service providers allocated resources to support open call projects. In total, 6.2 million CPU-core hours, 402,000 GPU-card hours, and 10,000 Xeon Phi-card hours are allocated as listed in Table 1.

System	Allocation for the open call		
	CPU-core hours	GPU-card hours	Phi-card hours
ARIS	3,000,000	15,000	
Avitohol	1,000,000		10,000
Isabella	1,000,000		
PARADOX	500,000	2,000	
Cyclone	500,000	15,000	
Leo HPC	200,000	370,000	

Table 1: Allocation of the on-boarded HPC resources for the open call projects

ARIS

ARIS is a high-performance computing resource hosted by GRNET (National Infrastructures for Research and Technology). All compute nodes are connected via SLURM resources/workload manager. Access to the system is allowed only via SSH from specific IPs/networks to login nodes from which all data management/transfers and job submission could be performed.

The ARIS infrastructure consists of a total of five computing system nodes, based on Intel x86 architecture, interconnected into a single Infiniband FDR14 network offering multiple options and processing architectures. More specifically, the infrastructure consists of:

- Thin Nodes island based on the IBM NeXtScale platform and on Intel Xeon E5-2680v2 processors. It has 426 computing nodes and offers a total of 8,520 CPU cores.
- Fat Nodes island consisting of 44 Dell PowerEdge R820 servers. Each server offers 4 Intel Xeon E5-4650v2 processors and 512 GB of central memory.
- GPU accelerator nodes comprised of 44 Dell PowerEdge R730 servers. Each server contains 2 Intel Xeon E5-2660v3 processors, 64 GB of memory and 2 NVidia K40 GPU cards, and
- Xeon Phi accelerator nodes consisting of 18 Dell PowerEdge R730 servers, each containing 2 Intel Xeon E5-2660v3 processors, 64 GB of memory and 2 Intel Xeon Phi 7120P co-processors.
- Machine Learning node consisting of 1 server, containing 2 Intel E5-2698v4 processors, 512 GB of central memory and 8 NVIDIA V100 GPU card.

In addition, the system offers 2 PB storage space through IBM's General Parallel File System (GPFS). The infrastructure is complete with an IBM TS3500 library of maximum storage capacity of about 6 PB.

Avitohol

The supercomputer Avitohol was at 331st place in the TOP 500 list of supercomputers. It is built with HP Cluster Platform SL250S GEN8 (150 servers), Intel Xeon E5-2650 v2 8C 2.6GHz CPUs (300 CPUs), non-blocking InfiniBand FDR, 300 Intel Xeon Phi 7120P co-processors. It provides 412 TFlops of performance for diverse scientific and industrial applications. Users from science and industry with substantial computational needs use it to achieve their results faster and to solve bigger problems that are beyond the reach of ordinary clusters.

Isabella

Isabella consists of 135 worker nodes with 3100 processor cores, 12 GPUs and 756 TB data space. As a shared resource of all scientists in Croatia, it allows using significant computational resources in demanding data processing of scientific and research projects.

PARADOX

PARADOX-IV cluster represents the fourth major upgrade of the PARADOX cluster and became operational in September 2013. The cluster consists of 106 working nodes and 3 service nodes. Working nodes (HP ProLiant SL250s Gen8, 2U height) are configured with two Intel Xeon E5-2670 8-core Sandy Bridge processors, at a frequency of 2.6 GHz and 32 GB of RAM (2 GB per CPU-core). The total number of new processor cores in the cluster is 1696. Each working node contains an additional GP-GPU card (NVIDIA Tesla M2090) with 6 GB of RAM. With a total of 106 NVIDIA Tesla M2090 graphics cards, PARADOX is a premier computer resource in the wider region, which provides access to a large production GPU cluster and new technology. The peak computing power of PARADOX is 105 TFlops.

One service node (HP DL380p Gen8), equipped with an uplink of 10 Gbps, is dedicated to cluster management and user access (gateway machine). All cluster nodes are interconnected via Infiniband QDR technology, through a non-blocking 144-port Mellanox QDR Infiniband switch. The communication speed of all nodes is 40 Gbps in both

directions, which is a qualitative step forward over the previous (Gigabit Ethernet) PARADOX installation.

Cyclone

The Cyprus Institute provides the Cyclone cluster. It has a theoretical peak performance of 600 TFlop/s. It consists of 17 compute nodes and 16 GPU nodes and is the new production HPC system of The Cyprus Institute since its deployment in 2020. All nodes are equipped with two Intel Xeon Gold 6248 CPUs. The GPU nodes have 4 NVIDIA Tesla V100-SXM2 32GB each. The memory on the compute nodes is 96GB and on the GPU nodes is 192GB. The machine has a flash storage capacity of 135TB and is also connected to the 3PB HPC storage. An HDR100 Infiniband network connects all the system components.

Leo HPC

Leo HPC is a cluster-type machine located in Wigner DC, Budapest. The overall processing power of the 1344 CPU cores and the 252 GPUs is around 254 Tflops. Concerning the storage capacity, a total of 585 TB can be used for computational purposes. More than 380 projects have used the Leo cluster so far, and nearly 850,000 tasks have been submitted since its start. Over the years, approximately 70 disciplines have been registered, such as theoretical/computational chemistry, biology, physics, etc.

3.1.2. Cloud services

In total, five different cloud instances are on-boarded within the framework of the NI4OS-Europe project. The amount of allocated VM-cores for the open call purposes is given in Table 2.

System	Allocation for the open call
	VM-cores
FINKI	140
ICIPRO	64
Avitohol	80
GCoud.ge	64
ASNET-AM	16

Table 2: Amount of allocated VM-cores for open call purposes

FINKI

FINKI is an Openstack-based cloud infrastructure deployed at the Faculty of Computer Science and Engineering, UKIM. FINKI is hosted on 15 Huawei servers, each with 128GB RAM and 20 HT CPU cores, totalling 300 vCPU cores and 37TB SSD and 32 TB SAS storage. The system has been in production since 2017 as a National cloud system. The connectivity to the Internet is 1 Gbit through MARNET provided link to GEANT. Currently, the system hosts templates for all popular Linux distributions and Windows variations.

ICIPRO Cloud

ICIPRO offers Infrastructure as a Service service (IaaS) to Public Sector beneficiaries that need flexibility, modularity, dynamics, and access to state-of-the-art technologies. Currently, there are three types of subscriptions available for the virtualization platform, with the following characteristics: Bronze, Silver, and Platinum. Details can be found in https://www.icipro.ro/doc/EN%20-%20Manual%20Tenants.pdf.

Avitohol

The Avitohol Cloud service allows users to launch virtual machines on servers from the Avitohol supercomputer described above. It allows user groups to launch long-running virtual machines with substantial flexibility. It is used by diverse research groups, which need for both advanced computing and data storage.

GCoud.ge

GRENA is offering a cloud service – GCloud. Besides the standard IaaS (Infrastructure as a Service) platform, GRENA offers a wide range of applications that can be installed automatically within a few minutes. Additionally, we have integrated various useful services for developers to make their work easy. Details can be found at https://www.gcloud.ge/.

ASNET-AM Cloud

ASNET-AM cloud is based on Openstack and deployed at the Institute for Informatics and Automation Problems of NAS RA, Yerevan, Armenia. ASNET-Cloud provides Infrastructure as a Service (IaaS) to academia and stakeholders based on OpenStack middleware. The users launch virtual machines (1-64 CPU cores) via a dashboard distributed in the three zones (596 CPU cores in total).

3.1.3. Data processing services

Data analysis service or PARADOX Hadoop cluster consists of a single name node that runs the YARN resource manager, and three additional data nodes. The name node is hosted on a machine with a 4-core Intel Xeon E3-1220v3 CPU running at 3.1 GHz, with 4 GB of RAM, and 500 GB of local hard disk storage. Each of the data nodes, which perform the computation and storage, are hosted on machines with 24-core Intel Xeon E5-2620 CPUs at 2.4 GHz, with 64 GB of RAM and 2 TB of storage. In total, the cluster provides access to 60 CPU cores, 180 GB of RAM and 5.3 TB of storage in HDFS.

In the analysis of very large datasets, the movement of data can present a far more severe bottleneck than the actual computation. Therefore, the PARADOX Hadoop cluster is designed to overlap computation and data storage operations, i.e., to enable performing of computation on the same machine(s) that store the corresponding data.

For the open call purposes, 200,000 CPU-hours of the Data analysis service are allocated.

3.1.4. Storage services

ARIS archival

Data archiving is the practice of moving data that is no longer being used or are being used on a less frequent fashion into a separate storage device. It is a single set or a collection of historical records specifically selected for long term retention and future

reference. The Archival service is based on the tape-based tertiary storage of the ARIS HPC system.

Projects accepted in the open call will be offered with 10 TB of storage space.

Simple storage service

The Simple Storage Service (SSS) is a secure data storage service provided to researchers for storing and sharing research data as well as keeping it synchronized across different devices. SSS is functionally similar to Dropbox, Office 365, or Google Drive. Files are stored in conventional directory structures, accessible via WebDAV. User files are encrypted during transit and can be synchronized with local clients running Windows, macOS, or various Linux distributions. The service is based on the NextCloud platform.

Projects accepted in the open call will be granted 100 GB of storage space at the Simple storage service per created user account.

3.1.5. Other services

RePol – Repository Policy Generator

RePol – Repository Policy Generator is an open-source web application that guides users through the process of defining policies for repositories and web-based services. It helps in defining and maintaining comprehensive and clear repository and privacy policies. Generated privacy policies are suitable for any kind of service. RePol uses a step-by-step wizard and self-explanatory forms to guide users through the process. By choosing options in a form, users shape a policy document with predefined policy clauses formulated in line with the current best practice. The resulting policy document may be downloaded as an XML file, additionally customized, and integrated into the service or repository. The collected data and the key elements of the generated policy are provided in a machine-readable format. This allows for an automated interpretation of created policies and extraction of repository-level metadata for inclusion in registries, catalogues and various operational and data discovery tools. The resulting policy document may be downloaded, additionally edited, and integrated into a repository.

RePol is a free service, and it is publicly available. It does not require registration. We will fully support projects from the open call using this service. It is supported via the project's helpdesk.

LCT – License Clearance Tool

LCT aims to facilitate and automate the clearance of rights (copyright) for datasets, media, software and other content before they are released under an open licence or stored at a publicly trusted FAIR repository. It helps in the certification of datasets and other outputs in terms of licence compatibility analysis and selection as well as other related constraints. LCT helps in open research data management and certification for data repositories, aligning with activities and results of INFRAEOSC-5c.

LCT check for equivalence, similarity and compatibility between licenses if used in combination, particularly for derivative works. In one scenario, the user (data manager) aims to find an appropriate open-source licence for the set of the elements with separate licences or to select a licence for derivative work based on the content or components with various licences or that are not licensed at all. In the second scenario, the user declares

the desired out-license and verifies the compatibility of the existing in-licences with the derivative work, also getting the options for licensing of the existing unlicensed content.

The LCT user may register or use it as a guest. The data provided by the guest is not preserved after the report is produced. For the authenticated user, the content and all parts of the procedure are kept in the licensing clearance history and user history, allowing to resume the work or share the work and data with others. Clearance with LCT is not bound to the user who initiated the procedure and can be continued by others with granted access.

RoLECT – EOSC RoP Legal & Ethics Compliance Tool

This tool aims to help in addressing the need of researchers to publish in FAIR/open modes. RoLECT targets open and FAIR assessors' needs in terms of IPR, ethics and data protection compliance both at the policy and legal level and aligned with the EOSC governance. RoLECT aims at providing an aggregated procedure for legal and ethics compliance by integrating a set of model procedures including: model procedures for copyright acquisition, management and dissemination policies; model copyright clearance processes, documentation and tools; model data protection (GDPR compliant) processes, consent forms and data sharing agreements; decision support trees for data protection policies; model IPR and data protection documentation.

The intended use of the tool is to provide an aggregated guided assessment for EOSC Rules of Participation (RoP) focusing on legal and ethical aspects of compliance. Targeted users may be service providers, researchers and research organisations. The RoLECT platform will eventually evolve to automatically check the validity of the provided resources for at least the obligatory steps of the assessment.

RoLECT is fully operational and stable and will be aligned with the forthcoming input from the EOSC RoP Task Force.

3.2. Application-specific (Thematic) services

3.2.1. Application-specific services for Climate

RS2C – Remote Sensing Scene Classification

RS2C is a RESTful web service and web application for remote sensing scene classification based on convolutional neural networks. Currently, ResNet-50 pre-trained on ImageNet and fine-tuned on MLRSNet is used for classification. The web service is implemented in Python using TensorFlow Serving and Flask. The RS2C API provides methods for single- and multi-label classification.

3.2.2. Application-specific services for Digital Cultural Heritage

Clowder4DCH

The scope of the "**Clowder4DCH**" (C4DCH) service (https://clowder.hpcf.cyi.ac.cy/) is to offer a highly extensible active curation-based research data management platform. It enables users to form an online collaborative environment to support research communities and activities, and disseminate results. Clowder is a web-based data

management system that allows users to share, annotate, organize and analyze large collections of datasets. It provides support for extensible metadata annotation and distributed analytics for automatic curation of uploaded data. In our use case, C4DCH gives the user the chance to organize files in three different ways, datasets, spaces and collections.

CHERE

One-stop shop for digitizing cultural heritage.

CHERE Tools stands for Cultural Heritage Repository Tools and represents a set of web-based tools. It aimed at people working in cultural heritage preservation and digitization, but is not limited to those who use it as individual services since it can be used in a variety of ways. The service currently provides the following functions:

- Structure from Motion – reconstruction of a textured 3D object from a series of photographs of the object.
- Measurement of 3D objects – either reconstructed or user-provided.
- Conversion of BigBlueButton meetings' recordings to standalone video files.

3.2.3. Application-specific services for Life Science

FEPrepare

FEP prepare is a web server, which automates the set-up procedure for performing NAMD/FEP simulations. Automating free energy perturbation calculations is a step forward to delivering high throughput calculations for accurate predictions of relative binding affinities before a compound is synthesized, and consequently, save enormous time and cost.

DREAMM

DREAMM is a novel machine learning tool that predicts the protein-membrane interfaces of peripheral membrane protein, and optionally predicts binding sites near the predicted membrane-penetrating residues in protein conformational ensembles. As an output, the user can retrieve the membrane-penetrating residues in a .csv file and if the user's choice was to predict binding sites, a .zip file will be downloaded including the abovementioned .csv file, the binding pocket predictions, the visualizations, and summarized the binding site clustering results.

NanoCrystal

NanoCrystal is a novel web-based crystallographic tool that creates nanoparticle models from any crystal structure guided by their preferred equilibrium shape under standard conditions according to the Wulff morphology (crystal habit). Users can upload a .cif file, define the Miller indices and their corresponding minimum surface energies according to the Wulff construction of a particular crystal, and specify the size of the nanocrystal. As a result, the nanoparticle is constructed and visualized, and the coordinates of the atoms are output to the user.

ChemBioServer 2.0

ChemBioServer is a web server for filtering, clustering and networking of chemical compound libraries facilitating both drug discovery and repurposing. It provides researchers the ability to (i) browse and visualize compounds along with their

physicochemical and toxicity properties; (ii) perform property-based filtering of chemical compounds; (iii) explore compound libraries for lead optimization based on perfect match substructure search; (iv) re-rank virtual screening results to achieve selectivity for a protein of interest against different protein members of the same family, selecting only those compounds that score high for the protein of interest; (v) perform clustering among the compounds based on their physicochemical properties providing representative compounds for each cluster; (vi) construct and visualize a structural similarity network of compounds providing a set of network analysis metrics; (vii) combine a given set of compounds with a reference set of compounds into a single structural similarity network providing the opportunity to infer drug repurposing due to transitivity; (viii) remove compounds from a network based on their similarity with unwanted substances (e.g. failed drugs) and (ix) build custom compound mining pipelines.

Ingredio Application

Ingredio application is a natural processing language (NLP) application that offers a pipeline of three services related to the biomedical text. The application classifies biomedical text based on certain features of its content, extract compound names and infer causal relations from the text; however, it is experimental and is not meant to replace human curation. Its main use is to showcase how this can be used as a high-throughput and high precision language filtering software for large scale biomedical data.

OpenBioMaps

OpenBioMaps is a web-based, open-access database framework project maintained by the OpenBioMpas Consortium and the databases involved are at least partially open-access or contain free content. The OpenBioMaps provides an open-access web application, which is designed to create and use open-content biological databases, specifically for scientists and conservationists, and its customizable toolset allows for the easy access and management of data.

Reduce and Visualize Gene Ontology

REVIGO is a web server that summarizes long, unintelligible lists of Gene Ontology (GO) terms by finding a representative subset of the terms using a simple clustering algorithm that relies on semantic similarity measures. Furthermore, REVIGO visualizes this non-redundant GO term set in multiple ways to assist in interpretation: multidimensional scaling and graph-based visualizations accurately render the subdivisions and the semantic relationships in the data, while treemaps and tag clouds are also offered as alternative views.

eeghub.ge

EEGHUB.GE (http://eeghub.ge) is a Big Data EEG (Electroencephalogram-Brain Electrical Activity) online dataset in Georgia. The service is free for European (or national) researchers following the principles of findability and accessibility. Service has a convenient search engine, which allows users to identify any recordings that correspond to specific requirements. The recordings are easily accessible and can be downloaded for further exploitation. The target users are open-source groups of researchers/practitioners, lecturers/students, Scientific Organization, Hospitals, Universities, etc. It is envisaged that EEG collection "eeg.hub.ge" will support researchers in the field of neuroscience, psychophysiology, medicine, psychology, neurophysiology, cognitive and social science.

GRENA supported this research by providing technical assistance and computing resources for the establishment of an EEG online database portal EEGHUB.GE (http://eeghub.ge). The collected EEG data was systematized, curated, and stored in an online database in EDF format.

DICOM Network

"DICOM Network" service provide full set of functionalities for data collect, storage, distribution and exchange DICOM medical investigations. It's implements all the standard PACS interfaces as well as integrated security features. various imagistic investigations like tomography, Roentgen, ultrasound, angiography, etc. Familiarization and working experience accumulation by medical specialists in using such systems offer obvious advantages in imagistic investigations and forming treatment decisions, allow supporting collaborative work and appealing for support from the best local and foreign specialists who have extensive experience in the field.

MelGene

The MelGene database provides a comprehensive and regularly updated field synopsis and meta-analysis of all published genetic association studies performed in Cutaneous Melanoma (CM). In addition, dozens of up-to-date meta-analyses are available for all eligible polymorphisms with sufficient data.

3.2.4. Application-specific services for Computational Physics

Gaussian API

Gaussian regression (GPR), along with other emerging machine learning techniques, has become more and more popular in computational chemistry, physics, biology and life sciences. In conjunction with the molecular dynamics simulations (MD), these approaches are rather useful for the prediction of a wide variety of molecular and materials' properties and functionalities. However, due to the novelty of techniques, the procedures for their application as well as their validation are far from being standardized.

At the same time, GPR approach can be used to tune the parametrization in approximate computational techniques in a bias-free manner. A typical example of a popular and rather useful approximate technique is the Density functional tight-binding method (DFTB). This method aims to approximate the Density functional theory (DFT) approach, providing comparable accuracy at only a fraction of its computational expense. With DFTB, simulations on time- and length-scales, which are unfeasible with DFT, have become reality.

Aside from the above-mentioned advantage, DFTB also allows direct access to electronic properties, unlike other empirical approximate methods. This is a great advantage of DFTB, compared to other computational approaches with similar computational costs. Still, all these advantages come into play at the expense of the complication that empirical parameters need to be introduced. Due to this drawback, the method transferability (as compared to other more exact quantum chemical approaches) is significantly reduced. The DFTB segment related to its electronic structure part allows for the construction of workflows for transferable parametrization. However, the repulsive part of the potential is rather tricky to treat in this context. This service provides methods for fitting the repulsive part of the potential based on the GPR approach. The training data that can be used are DFT-DFTB energy or force residues. The methods can be applied to many elements at

once, i.e. for computing the repulsive potentials in the case of molecules that contain multiple "organic elements", such as C, H, O, etc.

Schrödinger API

In many subdisciplines of computational molecular sciences, computational physics, chemistry, biology, materials science, exact treatment and analysis of a wide variety of phenomena has to rely on the rigorous quantum description of the underlying processes. Numerous phenomena taking place in the nano-world are inherently quantum in nature. Their description and, more important, quantitative treatment, therefore, requires the usage of the apparatus of quantum mechanics. The basic paradigm of today's "mainstream" quantum mechanics is the Schrödinger equation, which is considered as a "quantum analogue" to the famous Newton's second law equation in classical physics. The effort required to solve the Schrödinger equation is heavily dependent on the dimensionality and complexity of the problem itself (e.g. the exact form of the Hamiltonian, number of the relevant degrees of freedom of the studied system etc.). Numerous methods have been proposed in the literature to achieve the mentioned aim. However, the available codes are most often user-hostile, the procedures for computation and generation of relevant data are non-standardized, and there is a clear lack of in-depth, thorough comparison of performances of various methods for solving the Schrödinger equation for various purposes. The proposed service will provide user-friendly (as much as possible) computational platforms for the solution of time-independent Schrödinger equation, implementing several algorithms that uses the Hermite discrete variable representation technique (DVR) approach.

3.2.5. Application-specific services for other sciences

OMApp (domain: Engineering & Technology)

OMApp is a cloud application for automatic image mosaicking and georeferencing. The application is designed to support several users, whereby each user can upload a set of captured images via a web interface, begin their processing and make an overview of already created maps. OMApp uses numerous open-source image processing tools and libraries, where the most computationally demanding among them can perform multi-core parallel processing, which provides better usage of the cloud resources.

3.3. Repositories

Short name	Full name	Description
CHERRY	CHEmistry RepositoRY	CHERRY provides open access to the publications, as well as to other outputs of the research projects implemented in this institution.
CeR	Central Repository of the Institute of Chemistry, Technology and Metallurgy	The aim of the repository is to provide open access to publications and other research outputs resulting from the projects implemented by the Institute of Chemistry, Technology and Metallurgy.

DAIS	Digital Archive of the Serbian Academy of Sciences and Arts	The aim of the repository is to provide open access to publications and other research outputs resulting from the projects and other activities implemented by the SASA and its institutes, as well as to ensure their long-term preservation. DAIS responds to the requirements of national and international funding bodies to share the outputs of publicly funded research and is compliant with the Open Science Platform of the Ministry of Education, Science and Technological Development of the Republic of Serbia.
Digital Library	Digital Library of University of Maribor	It supports open access to scientific, research and professional works, and research data, which are results of research and education at the University. DKUM includes works from all University of Maribor members. Next to diplomas, master's degrees, doctorates and other works by students it also includes reviewed publications from funded projects, electronic academic textbooks and materials, and other works whose authors are the University of Maribor staff or if they are published by the University of Maribor.
HELIX Data	Hellenic Data Service	A data catalogue and repository, with a dual role in storing and preserving data that are self-deposited by researchers and in harvesting data records from other national data sources and catalogues, such as HARDMIN.
NI4OS-Europe repository service		NI4OS Repository Service (NRS) is the main storage service of a community that holds "Regional Community Datasets". The NRS is also the platform to host all kinds of additional data such as publications (and their associated data), software (or references to software), workflow descriptions (e.g. how to generate research data) or even materials targeting the general public (e.g. images, videos, etc.). NRS is integrated with a persistent identifier service, as an assigned PID is required for each digital object (item, collection, community).
Repository of Faculty of Science, University of Zagreb		The repository provides access to publications and research data produced by the employees and the students of the faculty. Content of the repository is open to all users for searching and downloading with clearly stated usage rights.

Repository of the Institute of Public Finance, Zagreb		The repository provides access to publications and research data produced by the employees of the institute, papers published in scientific journals, conference proceedings, dissertations, books, manuals, guides and complete documentation related to the Institute's activities. Content of the repository is open to all users for searching and downloading with clearly stated usage rights.
SZTE repository of publications		The SZTE Repository of Publications makes the full text of publications created as a result of scientific and artistic activities at the University available for the widest possible academic audience. Depositing works at the repository secures their long term archiving, and can also increase their visibility and number of citations. This is because uploaded documents are indexed by general search engines (e.g. Google, Google Scholar) and professional databases (e.g. BASE, MTA OAI).
Social Scientific Research Documentation Centre Repository		The Research Documentation Centre of the Centre for Social Sciences provides information on and access to research conducted at the Centre. The metadata and some of the documents of the Research Documentation Centre (RDC) are available to all visitors, but many are restricted to registered users. The interface can be set to many languages and contains RSS feeds to alert users of new content.
University of Zadar Institutional Repository of evaluation works		The repository provides access to publications and research data produced by the employees and the students of the faculty. Content of the repository is open to all users for searching and downloading with clearly stated usage rights.
VideoLectures.Net		VideoLectures.NET is an award-winning free and open access educational video lectures repository. The lectures are given by distinguished scholars and scientists at the most important and prominent events like conferences, summer schools, workshops and science promotional events from many fields of Science. The portal is aimed at promoting science, exchanging ideas and fostering knowledge sharing by providing high-quality didactic content not only to the scientific community but also to the general public. All lectures, accompanying documents, information and links are systematically selected and classified through the

		editorial process taking into account also users' comments.
Georgian Integrated Library Information System Consortium 2017	National repository of scientific works	Georgian Integrated Library Information System Consortium GILISC was founded in 2017 to assist educational institutions in their development. GILISC gives them access to the electronic databases of scientific journals, helps them to use modern electronic catalogues/integrated library systems, store their scientific works in the digital repository (openscience.ge), publishes scientific articles of their students and staff in electronic journals (openjournals.ge), helps them to increase the quality of their student's scientific works. GILISC has been cooperating with the National Science Library of Georgia since its establishment and one of the results of their interaction is the repository openscience.ge
Meteorological and Hydrological Service of Croatia Repository	Institutional repository	The repository provides access to publications and research data produced by the employees of the institute. Content of the repository is open to all users for searching and downloading with clearly stated usage rights.
NaRDuS – National Repository of Dissertations in Serbia	National repository of PhD theses	NaRDuS (National Repository of Dissertations in Serbia) is a common portal of PhD dissertations and thesis evaluation reports from all Serbian universities. It is based on the Law on Higher Education (Amendments, Sept. 2014). All universities are obliged to deposit basic information about the dissertation – together with the thesis evaluation report and the dissertation itself to NaRDuS within three months period starting from the date of PhD dissertation defence.
Veterinar – Repository of the Faculty of Veterinary Medicine	Institutional digital repository of the University of Belgrade, Faculty of Veterinary Medicine.	Veterinar – Repository of the Faculty of Veterinary Medicine is the institutional digital repository of the University of Belgrade, Faculty of Veterinary Medicine. The aim of the repository is to provide open access to publications and other research outputs resulting from the projects implemented by the Faculty of Veterinary Medicine.

Table 3: List of on-boarded repositories

4. Conclusions and next steps

This document provides a detailed description of the open call procedures and requirements, the available services for the call, and the established user support. WP6 oversees the engagement of real user communities with EOSC and specifically with the on-boarded providers. Within task T6.5 "Open call support", WP6 supports the open access call in order to provide the project solutions to the user communities in the wider pan-European area.

This call provides a framework for scientists to perform their work, extract and publish results based on Open Science FAIR principles. The call is designed and implemented to use and test the NI4OS-Europe services, which are part of the European Open Science Cloud. These services are thematic services, generic services, as well as repositories.

This deliverable reports on the preparatory activities conducted within task T6.5 "Open call support" to ensure the smooth progress of the open call. Early promotion of the call is planned to ensure a solid response by scientists. Establishing an efficient collaboration between the users and the support services in NI4OS-Europe is also necessary for high-quality results. Finally, at the end of the lifecycle of the supported projects, T6.4 will ensure that all relevant information is collected and made available, imposing Open Science principles on the final product (manuscripts, data, software, workflows).

Final report on the results of the review process of the submitted applications, as well as the outcome of the actual projects, which have been given access to generic services and support to thematic services, as well as repositories, will be presented in D6.8 "Report on the open call".