H2020-INFRAEOSC-2018-3

NI4OS-Europe

National Initiatives for Open Science in Europe

Deliverable D6.3

User Communities' test-cases

Lead beneficiary(s): CYI

Author(s): Chrysovalantis Constantinou, Andreas Athenodorou, all partners

Status –Version: Final – I

Date: July 29th, 2020

Dissemination Level: Public

Abstract: Deliverable D6.3 – User communities' test-cases – This deliverable provides a detailed description of cross-border test-cases including research groups involved and their mapping to generic, thematic and repository services which will be tested. It describes cross-border use cases for the target scientific areas of Life Science, Climate Science and Digital Cultural Heritage as well as for the area of Computational Physics.

The NI4OS-Europe project is funded by the European Commission under the Horizon 2020 European research infrastructures grant agreement no. 857645.

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
June 10, 2020	a	Chrysovalantis Constantinou	Initial ToC
June 20, 2020	b	Hrachya Astsatryan	Comments on ToC
June 27, 2020	c	Bojana Koteska	Chapter contribution
July 7, 2020	d	Theodoros Christoudias	Climate Scientific Community contribution
July 11, 2020	e	Andreas Athenodorou	Editing
July 14, 2020	f	Chrysovalantis Constantinou, Bojana Koteska	Editing, Computational Physics contribution
July 14, 2020	g	Georgios Artopoulos	Digital Cultural Heritage Scientific Community contribution
July 15, 2020	h	Zoe Cournia	Life Science contribution
July 16, 2020	i	Andreas Athenodorou, Chrysovalantis Constantinou, Bojana Koteska	Editing, Adding information
July 21, 2020	j	Andreas Athenodorou, Chrysovalantis Constantinou	Editing
July 25, 2020	k	Milutin Radonjic	Internal quality check
July 27, 2020	l	Chrysovalantis Constantinou, Ognjen Prnjat	Final editing

Table of contents

References

[1] Project NI4OS-Europe-857645 - Annex I - Description of the Action
[2] https://marketplace.eosc-portal.eu/
[3] https://www.usability.gov
[4] https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/
[5] https://chembioserver.vi-seem.eu/
[6] https://afmm.vi-seem.eu/
[7] https://subtract.vi-seem.eu/
[8] https://feprepare.vi-seem.eu/

List of Figures

List of Tables

List of Acronyms

AFMM	Automated Frequency Matching Method
AI	Artificial Intelligence
DICOM	Digital Imaging and Communications in Imaging
DNA	Deoxyribonucleic Acid
EOSC	European Open Science Cloud
EOSC Pillar	Coordination and Harmonization of National Initiatives, Infrastructures and Data services in Central and Western Europe
EOSC Synergy	European Open Science Cloud - Expanding Capacities by building Capabilities
ERA	European Research Area
EXPANDS	EOSC Photon and Neutron Data Services
FAIR	Findable Accessible Interoperable Reusable
FAIRsFAIR	Fostering Fair Data Practices in Europe
FEP	Free Energy Perturbation
GENR	Generic Service
GPU	Graphics Processing Unit
HPC	High Performance Computing
IAAS	Infrastructure as a unit
NAMD	Nano Scale Molecular Dynamics
NWCHEM	North West Computational Chemistry
OpenAIRE	Open Access Infrastructure for Research in Europe
ORDM	Open Research Data Management
OS	Open Science
PID	Persistent Identifier
PMB	Partner Member Board
REPO	Repository
RoP	Rules of Participation
SERV	Thematic Service
WP	Work Package
DREAM	Dust Regional Atmospheric Model
ARIS	Advanced Research Information System
DCH	Digital Cultural Heritage

Executive summary

What is the focus of this Deliverable?

The focus of this deliverable is to provide the details of the construction and design of use-cases, which are set up in order to perform a first test on a number of selected services, which will be on-boarded to the European Open Science Cloud (EOSC). These services include thematic, generic as well as repository services.

Most of the mature thematic services which participate in the use cases belong to the consortium of the three NI4OS-Europe pre-identified fields, namely the scientific fields of Digital Cultural Heritage, Climate Science and Life Science.

Thus, the focus of this deliverable is to describe how the three scientific community leaders, those of the Digital Cultural Heritage, Climate Science, and Life Science fields in cooperation with all the related stakeholders of NI4OS-Europe will work towards the realization of use cases for the testing of the on-boarding of NI4OS-Europe thematic, generic, and repository services, which have been embedded into EOSC during the first round of on-boarding.

Based on the maturity of the thematic services, NI4OS-Europe has identified one additional scientific community, that of Computational Physics. A use case focusing on this field was also designed and will serve as an additional test of the proposed NI4OS-Europe scientific solutions.

The use cases will also have a highly cross-border character in order to test whether services across the target area can work well together.

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

This deliverable has the purpose of describing the use cases as these are conceived and defined by the scientific teams in collaboration with the stakeholders of the NI4OS-Europe region. Next in line is the execution of demonstrators, which is the actual running of the use cases which will be performed in Task 6.2 and will be described in deliverable D6.4 "*Service evaluation by user communities*" which is due by month M18.

What are the deliverable contents?

The deliverable describes in detail the actions that will be taken by the stakeholders as well as the scientific leaders of the Digital Cultural Heritage, Life Science, and Climate Science fields to test how NI4OS-Europe on-boarded services work and what the possible recommendations are. Specifically, they will test how thematic services can run on generic services (such as computational resources) and output their results on repositories. The generic services need not be found in the same country as the repositories. Ideally, they should be in different countries (thus being truly cross-border use cases, as proof of principle).

Conclusions and recommendations

The present deliverable provides guidelines to be used once the services are on-boarded. It is expected that for the actual execution of demonstrators a questionnaire will be put together which will evaluate the user experience.

1. Introduction

In this section we provide the basic information the reader needs to know in order to understand what the purpose of this deliverable is.
NI4OS-Europe aims to build capacity in the region and to support the national open science cloud initiatives. Hence an important aspect of this project is to be able to provide concrete solutions to scientists (working in any scientific field of research) for producing and curating scientific results within the platform of the European Open Science Cloud (EOSC). Hence well-designed testing should take place in order to verify that this is the case. This testing will be performed by the use cases.

1.1. Purpose

After the first set of generic, thematic and repository services get on-boarded, NI4OS-Europe will allow the first set of pre-defined users to test the service portfolio in WP6, and provide feedback through which WP5 can fine-tune the portfolio. Recall that WP5 is responsible for the on-boarding of services while WP6 is responsible for user engagement, training, and demonstrators. A detailed description with a well-designed timeline on the on-boarding of the services has been presented in deliverable D5.1. Based on this deliverable, by filtrating the outcome of the collected service description templates by the services providers, and by setting up a list of criteria the user communities should fulfill, NI4OS-Europe consortium came up with a number of scientific communities which will participate in the first part of the testing of the on-boarding of services.

Here we provide detailed cross-border test-cases that are candidates to test the whole spectrum of project solutions, focusing on the EOSC-embedded services from the first on-boarding phase. By cross-border we mean that thematic services from country A may run on computational resources (generic services) in country B, while the output may be stored in repositories located in country C. Moreover, the outcome of the use case is of interest to a broad geographic area and not just a single country. This reflects the highly cross-border character of the NI4OS-Europe project as well as EOSC itself, creating a truly pan-European research space.

The cross-border scientific test-cases may also include usage of services from the current EOSC [2], thus offering testing interoperability of project solutions not just between them, but also with the current EOSC service catalogue. Furthermore, the cross-border scientific use case may also include the usage of complementary services, which are already mature and are on-boarded on other infrastructures such as VI-SEEM (https://vi-seem.eu/).

The structure of this deliverable is as follows. First, we define the use cases. This involves the explanation of what a use case is, the requirements of use cases as well as the criteria of selecting the use cases in NI4OS-Europe. Subsequently, we describe the use cases in different scientific domains. These scientific domains include the highly inter-disciplinary fields of Digital Cultural Heritage, Climate Science and Life Science which were mentioned above. In addition, we included a use case from the cross-disciplinary field of Computational Physics. In the following section, we provide a simple description of how

the NI4OS-Europe stakeholders will participate in these cross-border use cases and finally we will conclude.

1.2. Background

This deliverable is written in the framework of Task 6.1 aiming to provide a complete list of use cases which will be used for testing the NI4OS-Europe services. As mentioned above these use cases are drawn from the three leading scientific communities, as well as computational physics, which was identified as an additional scientific community in D5.1.

Task 6.1 builds on the WP2 mapping, landscaping and analysis tasks, which performed a detailed probe of the national user communities. After analyzing the outcomes of T2.1 stakeholder analysis and mapping, the task defined the criteria for selecting user communities to be involved in the project activities. We now focus on the set up of the groups, which will carry out the test-cases required for testing the NI4OS-Europe solutions, including generic, thematic as well as repository services included in EOSC.

The demonstrators are expected to test the first set of services integrated into EOSC as performed by WP5, and the use-cases for these services are given in this deliverable. More specifically, Task 6.1 identified the scientific groups involved in the testing. Secondly, the thematic services and repositories undergoing testing were determined and so that a link to the generic services under testing will be established. A clear analysis of the required infrastructure capacity/resources will be provided. Subsequently, resources will be allocated in close coordination with the service providers. Finally, repositories where resulting data will be stored are identified.

2. Defining the use cases

2.1. Introduction

As it has already been explained, one of the goals of NI4OS-Europe is to populate the EOSC service catalogue with services originating from the geographic area covered by NI4OS-Europe. This includes state of the art thematic services developed by scientific teams and scientific communities within the aforementioned geographic area. These services have a strong cross-border character in the sense that they apply not only to a single country but to multiple countries. Of course, these services will undergo a procedure of harmonization with the EOSC principles as well as FAIRification. Hence, these services are expected to be accessible through the EOSC service catalogue.

Furthermore, FAIR-verified repositories will also populate the EOSC service catalogue through the route of OpenAIRE. A number of these repositories are focusing on just one scientific thematic region while others on multiple. Thus, a researcher would be able to access these repositories through the EOSC service catalogue and use them with a maximum efficiency.

Finally, a set of generic services will also be on-boarded to EOSC and will enable researchers to access tools and machinery, which will be crucial for the successful completion of research projects. Generic services do not belong to particular scientific communities but they have a rather cross-disciplinary character and can be used by researchers working on the long tail of science.

We can claim that with this effort NI4OS-Europe will provide COMPLETE SOLUTIONS to scientists, in order to run computationally intensive codes and produce fresh results in particular interdisciplinary scientific domains, store these results, and curate them by making use of well-designed repositories.

The combination of the three components of thematic, generic and repository services accessed throughout the EOSC portal should provide such complete solutions. Hence, after these services are on-boarded in EOSC, through all the necessary actions taken by NI4OS-Europe, the consortium of NI4OS-Europe should make sure that these services have successfully been on-bordered on EOSC, they reflect the right Maturity level (Technology Readiness Level), and they are fine tuned. To achieve that we need to run demonstration use cases in which scientific teams will participate, use and test such services. Their experience with their services will be reported in a questionnaire, which will be distributed to all the stakeholders which have participated in the use cases and collected once the use cases run.

2.2. What is a use case?

A use case is a written description of how users will perform tasks on a series of services, which are connected by a logical sequence [3]. Thus, in our case a description of how a user will use a combination of thematic, generic, and repository services. It outlines, from a user's point of view, a system's behavior as it responds to a request. Each use case is

represented by a sequence of simple steps, beginning with a user's goal and ending when that goal is fulfilled.

Use cases add value because they help explain how the system should behave and, in the process, they also help brainstorm what might go wrong and what is faulty. They provide a list of goals and this list can be used to establish the cost and complexity of the system. Project teams can then negotiate which functions become requirements and are built.

Use cases should include details about:

- Who is going to be using a service
- What the user wants to do with the service
- The user's goals
- The steps taken by the user in order to achieve a task
- How the service/website should respond to an action

Depending on how in depth we want a use-case to be, use-cases describe a combination of the following elements:

- **Actor** – anyone or anything that performs a behavior (the user of the system)
- **Stakeholder** – someone or something with vested interests in the behavior of the system under discussion
- **Primary Actor** – stakeholder who initiates an interaction with the combination of systems to achieve a goal
- **Preconditions** – what must be true or happen before and after the use case runs.
- **Triggers** – this is the event that causes the use case to be initiated.
- **Main success scenarios** [Basic Flow] – use case in which nothing goes wrong.
- **Alternative paths [**Alternative Flow] – these paths are a variation on the main theme. These exceptions are what happens when things go wrong at the system level.

2.3. Requirements to build a use case

Each use case should in general include the following steps:

1. Identify who is going to be using the service.
2. Pick one of those users.
3. Define what that user wants to do with the services. Each action a user takes on the services becomes a use case.
4. For each use case, decide on the normal course of events when that user is using the services.
5. Describe the basic course in the description of the use case. Describe it in terms of what the user does and what the system does in response that the user should be aware of.
6. When the basic course is described, consider alternate courses of events and add those to "extend" the use case.
7. Look for commonalities among the use cases. Extract these and note them as common course use cases.
8. Repeat the steps 2 through 7 for all other users.

2.4. Criteria to choose a scientific area to participate in the use case

The cross-border use cases which will be executed belong to particular scientific areas. Thus, as a first step NI4OS-Europe consortium has to select what these communities are.

NI4OS-Europe has worked to define the criteria for selecting user communities to be involved in the project activities. Namely, these criteria are quite simple and are based mostly on practical issues. We have identified four main criteria:

- **Critical capacity in the region** – The scientific community to be selected should have a strong representation within the scientific community in each different NI4OS-Europe country.
- **Have a cross-disciplinary character** – The scientific community should produce data or use data which are proven to be useful for other scientific areas. For instance, climate data is useful for researchers focusing on archaeology and paleontology.
- **Have a cross-border character** – Stakeholders in more than one NI4OS-Europe country should be involved in the use cases and the data produced by the use cases should be of interest to a broad geographic area and not focused on a single country.
- **Details from survey for scientists to be contacted** – It would be better if the scientific teams participating in the use cases have participated in the landscaping survey, provided input to NI4OS-Europe, and have shown interest in NI4OS-Europe actions. A number of scientists participating in the landscaping survey are contributing to NI4OS-Europe with thematic services, and thus could help us identify the scientific communities to participate in the demonstrations.

An advisory team with an inter-disciplinary research background have been set up in order to monitor and guide the choice of additional scientific fields to participate in the demonstrations. The advisory board team consists of the following scientists:

- Mihajlo Savic (BA)
- Bojana Koteska (MK)
- Hrachya Astsatryan (AM)

This team led by the task leader worked closely on the identification of the selection of scientific areas, which are participating in the use cases. Since most of the on-boarded thematic services with high TRL (TRL of 8 and 9) belong to the scientific communities of Climate Science, Life Science and Digital Cultural Heritage, we will include use cases focusing on these three scientific areas. In addition, the advisory board team has identified an additional scientific community, namely that of Computational Physics, which will be based on thematic services with TRL 6 and 7.

3. Application in scientific domains

3.1. Life Sciences

3.1.1. General description of Life Science in the NI4OS-Europe area

Advances in computational infrastructure during the last decade have facilitated the development of biological and medical data analysis for big data and computational biology as key research methodologies in both academia and industry. The use of computers in life sciences has enabled our better understanding of mechanistic aspects in health and disease and has accelerated the development of novel therapeutics, diagnostics and preventative measures.

The exponential growth and availability of data from patients and beyond have led to the "big data" era. Data sets became so large and complex that it becomes difficult to process them using traditional data processing applications. This points to a strong relationship between the life sciences community and the European Open Science Cloud. The associated data analysis needs and challenges include the capture, curation, analysis, search, sharing, storage, transfer, and visualization of the data. The upward trend of data sets is due to the additional information extracted from the analysis of a single large set of related data (meta-analyses), as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to prevent and understand diseases.

The Life Sciences (LS) research community within NI4OS-Europe deals with research topics that play a central role in achieving a higher quality of life in the region. For instance, the special genotypes/phenotypes of human sub-populations of NI4OS-Europe countries can help us describe the medical needs in this particular target area.

The Life Sciences Community thematic service portfolio will create and provide the necessary services capable to facilitate research for the understanding of disease mechanisms and associated biomedical questions. One example of this use is the analysis of dominant cancerous mutations or overexpression of oncogenic proteins in Eastern European patients that have been already genotyped with an eye on making these findings accessible to the wider scientific community through open science, fostering intra-discipline fertilization.

The Life Sciences Research community within NI4OS-Europe needs a variety of infrastructures. Large amounts of data need to be stored and made available to researchers for processing in the computer centers of the region. Therefore, apart from storage resources, fast and reliable networking infrastructure is important for moving large datasets from data archives to the computing centers and moving simulation results close to the researchers for further post processing and acquisition of results. In terms of computer infrastructure, the models and services to be used by the research groups require capacity and capability computing, as well as the provision of computing resources for the installation of user facing services. For example, codes such as NAMD and NWCHEM scale up to hundreds or thousands of cores and can benefit from scalable HPC clusters or supercomputers such as the IBMs BlueGene. Molecular dynamics applications are also known to perform well on GPU systems while they are also being ported to Intel's Phi

accelerator platform. On the other hand, parametric codes for human genome sequence analysis can benefit a lot from the Grid or Cloud IaaS computing model. Other scientific challenges and problems that require enabling, include the primary, secondary and tertiary analysis of Next Generation DNA sequencing data for the detection of aneuploidies, using DNA sequencing alignments of Next Generation Sequencing data, modeling and classification of pathologies by means of image processing. Finally, user-facing services can be also installed in the IaaS infrastructure that will be available in the project. It is evident that the Life Sciences Scientific Community requires a variety of infrastructure resources, all of which are going to be available through EOSC services.

NI4OS-Europe Life Sciences thematic portfolio, available in Table 1, will provide unique real-time access to biological datasets such as systematic meta-analyses of genetic polymorphisms in cutaneous melanoma and its in-depth molecular network visualization and exploration (MelGene). Also, simulated solitary pulmonary nodules in pairs of computed tomography and positron emission tomography images could be used by clinicians and experts to extract patient-centric correlations using artificial intelligence (SPNSim, EML, Dad_Mod). Another area, where NI4OS-Europe will play a pivotal role is the collection and re-use of medical data. "DICOM Network" collects, processes, and visualizes medical images online, while at the same time it provides a front-end user interface for patients, doctors, scientists, and can analyze data. "Image-Reconstruct" can reconstruct medical images from Single-photon emission computed tomography (SPECT), Positron emission tomography (PET) or X-ray computed tomography (X-ray CT), while "LFP" is a software harnessing signal processing and machine learning methods to interpret electrophysiological signals generated by the brain. At the same time, it will address the challenge of the lengthy and costly drug discovery process by integrating a fully-automated platform for computer-aided drug design including services such as "ChemBioServer" [5], "AFMM" [6], "Substract" [7], "FEPrepare" [8], which will enhance small molecule lead generation and optimization to be used by academic researchers, as well as the pharmaceutical industry. Such use cases utilize computing generic services to accelerate drug discovery. Digital health and healthcare big data analytics when accessed in real-time from a Cloud-based environment, may be converted into relevant patient insights to be utilized for the tailoring of medical treatment and personalizing care. With NI4OS-Europe services, open medical and biological data accessed from within the Cloud are expected to be effectively used in healthcare and wellbeing, as personalized medicine, while biological big data analytics will be incorporated in standard patient care through prevention schemes, disease profiling, and the discovery of novel therapeutics. The NI4OS-Europe Life Sciences service portfolio provides new opportunities in the field and a rich environment for relevant stakeholders to foster innovation.

The cross border/cross region nature of human disease will greatly benefit from data and local knowledge over a wider scientific base and will allow the de-duplication of efforts, which allow for efficient interlab comparisons.

In the following table we provide all the thematic services focusing on Life Sciences:

Name of service	TRL	Scientific Domain	Institution	Country	Description
Dicom Network	9	Medicine	IMSP Institutului de Medicină Urgentă	MD	"DICOM Network" provides a full set of functionalities for data collection, storage, distribution and exchange of DICOM medical investigations. It implements all the standard PACS interfaces as well as integrated security features. The service works with various imaging investigations such as tomography, Roentgen, ultrasound, angiography, etc. Familiarization and working experience accumulation by medical specialists in using such systems offers obvious advantages in imaging investigations and forming treatment decisions, allow collaborative work, and appealing for support from the best local and foreign specialists who have extensive experience in the field.
OpenBio Maps	9	Biology	OpenBioMaps Consortium	HU	OpenBioMaps is a web-based, open-access database framework project maintained by the OpenBioMaps Consortium. The databases involved are at least partially open-access or contain free-content. OpenBioMaps provides an open-access web application designed to create and use open-content biological databases, for scientists and conservationists, and its customizable toolset allows for the easy access and management of data.
REVIGO	9	Bioinformatics	Ruđer Bošković Institute	HR	REVIGO is a Web server that summarizes long, unintelligible lists of 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 interpretating: 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.
ProTraits	9	Bioinformatics	Ruđer Bošković Institute	HR	ProTraits atlas is a resource containing ~545 000 novel phenotype inferences, spanning 424 traits assigned to 3046 bacterial and archaeal species. These annotations were assigned by a computational pipeline that associates microbes with phenotypes by text-mining the scientific literature and the broader World Wide Web, while also being able to define novel concepts from unstructured text.
Nano Crystal	9	Chemistry, Biology, Biochemistry, Biophysics, Medicine	Biomedical Research Foundation, Academy of Athens	GR	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.
Chem BioServer	9	Chemistry, Biology, Biochemistry, Biophysics, Medicine	Biomedical Research Foundation, Academy of Athens	GR	ChemBioServer is a web-server for filtering, clustering and networking 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.

Image Reconstruct	9	Medical Imaging	The Cyprus Institute	CY	Image Reconstruct will offer a comprehensive solution intended to reconstruct medical images from single-photon emission computed tomography (SPECT), positron emission tomography (PET) or X-ray computed tomography (X-ray CT), using multiple methodologies, including but not limited to filtered back-projection (FBP), algebraic reconstruction techniques (ART) and maximum likelihood expectation maximization (MLEM). Furthermore, Bio-connect provides a set of tools used to assist with data validation, management, viewing and sharing of reconstructed data as images.
BioConnect	9	Bioinformatics	The Cyprus Institute	CY	Bio-connect will offer a comprehensive data discovery solution intended to provide a general-purpose, web-based tool that can be used by any omics data owner. Bio-connect will allow omics data to be appropriately discoverable in a manner that data owners maintain control of the data, where they can set multiple levels of access. Furthermore, Bio-connect provides a set of tools used to assist with omics data validation, management, viewing and sharing of complex omics data.
Subtract	8	Chemistry, Biology, Biochemistry, Biophysics, Medicine	Biomedical Research Foundation, Academy of Athens	GR	Subtract accurately calculates the volume of protein binding sites, and works both for crystal structures downloaded from the Protein Data Bank and for protein structures arising from Molecular Dynamics simulations trajectories. Subtract accepts an atom selection in the form of a PDB file and computes the three-dimensional convex hull of the atoms points with the help of the SciPy library. The next step of the algorithm is to compute the volume of the convex hull and the volume of the atoms that are included in the solid based on their van der Waals radii. The subtraction of those two volumes yields the volume of the investigated cavity. The algorithm computes cavity volumes of trajectory frames in parallel for maximum efficiency and speed. It requires minimal usage of memory due to the fact that it follows a buffering strategy of reading file chunks and therefore there is no need to load the entire file into memory. There is a wide support of trajectory formats like Gromacs trajectory files and multi-model PDB files due to its dependency to the MDTraj library.
FEPrepare	8	Chemistry, Biology, Biochemistry, Biophysics, Medicine	Biomedical Research Foundation, Academy of Athens	GR	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.
AFMM	8	Chemistry, Biology, Biochemistry, Biophysics, Medicine	Biomedical Research Foundation, Academy of Athens	GR	AFMM (Automated Frequency Matching Method) is a program package for molecular mechanics force field parametrization. The method used fits the molecular mechanics potential function to both vibrational frequencies and eigenvector projections derived from quantum chemical calculations. The program optimizes an initial parameter set (either pre-existing or using chemically-reasonable estimation) by iteratively changing them until the optimal fit with the reference set is obtained. By implementing a Monte Carlo-like algorithm to vary the parameters, the tedious task of manual parametrization is replaced by an efficient automated procedure. The program is best suited for optimization of small rigid molecules in a well-defined energy minimum, for which the harmonic approximation to the energy surface is appropriate for describing the intra-molecular degrees of freedom. It can be used for small organic molecules.
Analysis of Microscopy data for cells and embryos	5	Software for biological image analysis	University of Cyprus	CY	A method that detects mitotic events as well as the orientation of the mitotic axis. It also follows the development of embryos.

EEGHUB.GE	4	Neuroscience, psychophysiology, clinical neuroscience, medicine, psychology,neurophysiology, Cognitive and Social Science	Caucasus University	GE	The service contains time series of recorded brain electrical activity of healthy subjects and subjects with different disorder of Central Nervous system, with various gender and age group. Service has convenient search engine, which allows to find any recording which corresponds to specific requirements. The recordings are easily accessible and can be quickly downloaded for further exploration/exploitation. Possible customers are open-source groups of researchers/practitioners, lecturer/students, scientific institutions, Hospitals, Universities.
Toxicological evaluation	3	Environmental Toxicology	The Cyprus Institute	CY	Service which computes effective concentrations or Toxic Units that can be used to compare results among existing results in the database.
Effect-based toxicological assessment	3	Environmental Toxicology	The Cyprus Institute	CY	Service which computes Inhibition percentage that can be used to compare results among existing results in the database.

Table 1: Thematic services for Life Science and their descriptions

3.1.2. Life Science cross-border case: Extracting correlations for patient stratification using machine learning

Cloud-based technologies can enable prevention, diagnosis, and the discovery of new therapeutics in the framework of personalized medicine, and can also aid in classifying subgroups of patients according to their disease mechanism in order to provide them with the right drug at the right time and the right dose. Thus, if patients can be grouped and selected for treatment according to the mechanism of their disease before enrolment begins, efficacy and response rates should improve.

AI platforms can be trained to analyze biological information in remarkable breadths - from omics data to scientific journal findings - to construct or verify clinically-relevant hypotheses. The virtual infrastructure itself enables the sheer capacity and throughput to collect, process, screen, and target alongside the input resources. Research teams working with identifying melanoma biomarkers for patient stratification, for example the Klinakis lab at the Biomedical Research Foundation, Academy of Athens, will make use of the pool of NI4OS-Europe services to capitalize on existing datasets for harnessing the power of machine learning ("***BioConnect"*** thematic service) to exploit meta-analyses of genetic polymorphisms in cutaneous melanoma by making use of the MelGene thematic service provided by the Cyprus Institute of Neurology and Genetics. If machine learning applications require high performance computing then GPU cards from the Serbian generic service "***PARADOX-IV cluster***" will be utilized. The produced data will be stored on the generic services "***Archival Service***", which is mainly used for the archiving of scientific data as well as in the "***Simple Storage Service***" in order to share the data in a collaborative environment. For novel emerging biomarkers, the mature computer-aided drug design pipeline of NI4OS-Europe, namely the thematic services "***ChemBioServer***", "***AFMM***", "***Subtract***" and "***FEPrepare***", will be executed on one of the generic cloud service installations, and will be used to augment lab-based research so as to accelerate drug discovery and minimize expenses, by using an integrated service deploying an in-house protocol that has delivered high success rates in discovering new hits for established biomarkers, generating leads, and optimizing these leads to successful drug candidates for in vivo preclinical experiments. Finally, all data will be stored on the "**Generic Cloud Storage**" services of NI4OS-Europe and resulting publications will be submitted in publication repositories specializing in life sciences, such as the repository service "***Institutional Repository in Medical Sciences – Nicolae Testemitanu SUMP***".

Services aiming to test:

- ☐ [SERV] "BioConnect"
- ☐ [SERV] "ChemBioServer"
- ☐ [SERV] "AFMM"
- ☐ [SERV] "Subtract"
- ☐ [SERV] "FEPrepare"
- ☐ [GENR] "PARADOX-IV cluster"
- ☐ [GENR] "Archival Service"
- ☐ [GENR] "Simple Storage Service"
- ☐ [REPO] "Institutional Repository in Medical Sciences – Nicolae Testemitanu SUMP"

3.2. Climate Science

3.2.1. General description of the Climate Science in the NI4OS-Europe area

Traditionally, the climate and atmospheric composition modelling, and weather and air quality forecasting community has very strong needs related to the European Open Science Cloud. These needs are related to enabling scientific end-users to perform data analysis experiments on large volumes of research data from multiple disciplines. In the context of climate research, EOSC can foster cross-discipline fertilization, providing an open science environment able to deal with many new challenges.

The Climate and Earth system science community focuses on regional climate modelling and weather forecasting, complemented by global climate modelling. The results are crucial in predicting extreme weather and understanding the future trends and impacts. Climate impact studies provide the analysis of the upcoming change on humans, the environment, economy and society which is so crucial for policy makers. Another strong field of related research is the study of air pollution that includes the influence on climate human health, and well-being.

The interface between climate science and open science research, apart from the established part of simulation and data management, represents a great shift for the Climate scientific community and, at the same time, a significant opportunity to evolve towards the next generation of open (data-driven) science and innovation frontiers.

This in turn provides new opportunities for both the climate science research, influencing policy- and decision-makers in the government and private sectors, and directly affecting society in general. It also enables the advent of new and unexplored market opportunities, across different sectors, for climate services.

The NI4OS-Europe thematic service portfolio in climate science, which you can see in Table 2, provides a rich environment for scientists and can foster intra-discipline fertilization. The regional community's focus is on the study of air pollution that includes the influence on climate and human health, as well as regional climate modelling and weather forecasting, complemented by global climate modelling. The simulation output can now be made openly available as a policy demonstrator to all interested stakeholders.

The community is enabled through NI4OS-Europe as part of EOSC to guide policy makers to understand and manage air quality by identifying the sources of pollutants, their emission, transport and behavior in the atmosphere and their effects on human health. National emission inventories and atmospheric observations currently rely on disparate specifications and standards, hindering the seamless access to data and creating inconsistencies in the creation of regional to continental scale datasets.

NI4OS-Europe provides the community with the tools to collect and homogenize the products, by defining the relevant technical policies and adopting common, open standards. The relevant thematic services (AirQuality, DREAM) and generic services (NI4OS-Europe Data Repository, storage, high-end computing, B2SHARE) provide the necessary common platform to host the data and stage air quality model simulations to study real-world conditions and different policy scenarios.

Climate modelling and weather forecasting have very strong computational demands. In particular, the integration of various computational resources with data repositories jointly supports research and operational activity. The results are crucial to predicting extreme weather and abrupt changes and understanding the future trends of the climate system. The activities pursued based on the provided tools exhibit strong synergies, both geographically and thematically, and all require a seamless integration of data and computing resources. Climate modelling and numerical weather prediction traditionally rely on high-end computing generic services. Beyond the computation of climate models and weather forecasts, the community also relies on data from national, regional and international networks of measuring stations. These data are used as initial conditions and for validation purposes, hosted on cloud storage services.

Finally, the generic services support the climate user community by providing storage so that ensemble model output can be stored and used by more than one research group. In time, this prevents expensive re-runs and also facilitates multi- and super-ensemble prediction products. Thus, simulation outputs can be inter-compared and the model codes improved.

In the following table, we provide all the thematic services related to Climate Sciences:

Name of service	TRL	Scientific Domain	Institution	Country	Description
DREAM	9	Climate Modelling	Institute of Physics Belgrade	RS	The DREAM service simulates and predicts the atmospheric cycle of mineral dust aerosols. It is tuned for usage on high-performance computing infrastructures available today. A typical use case is the production of a dataset with the aerosol optical thickness and surface dust concentration for a particular period and for a particular geographical region. The service supports different horizontal and vertical resolutions. The results produced by the DREAM service have been applied, using the human health impact function and calculated global fine particular matter concentrations, for estimation of the premature mortality caused by the long-term exposure to airborne desert dust. The results have high sensitivity on the threshold concentration, which is a significant parameter of relevance to public health.
Cyprus Weather Data	9	Earth Sciences	The Cyprus Institute	CY	Service which provides model forecast products and real-time observations from Cyprus. This service is of the interest to European researchers in climatology, interested in the Eastern Mediterranean climate.
Airpolution prediction	9	Natural Sciences	Ss. Cyril and Methodius University in Skopje	MK	Simulation system for generation of prediction of air pollution levels based on WRF-Chem software. The outputs are hourly levels of air pollution for PM10, PM2.5, NO and SO2.

Open Mapping Application (OMApp)	5	Interdisciplinary	University of Montenegro	MN	OMApp is cloud application for automatic image mosaicking and georeferencing. The application is designed to support several users, whereby every user is able to 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 are able to perform multi-core parallel processing, which provides a better usage of the cloud resources.
RS2C	4	Engineering	University of Banja Luka Faculty of Electrical Engineering	BH	A web-based service for classification of remote sensing scenes into land use/land cover classes. The service is based on convolutional neural networks trained on publicly available datasets of high-resolution remote sensing images and implemented using TensorFlow, TensorFlow Serving, and Docker. The service enables faster analysis of large quantities of remote sensing images available from various sensors ranging from satellites to UAVs. These images are used for land cover and land use classification, monitoring urban growth, and forecasting climate changes, to name a few. Recently, using remote sensing for the monitoring of ecosystems, insects and animals gains are significant. The system in development aims to assist in making sense of the data acquired in these application domains. Main user communities envisaged to benefit from this service are in the areas of agriculture, food production, urban planning, and environment protection. However, we expect to identify new users groups both during the project, as well as after it is finished.
EML	4	Environmental science	Institute of Physics Belgrade	RS	The Explainable Machine Learning (EML) is a service for highly sophisticated machine learning (ML) and interpretation frameworks deployment aimed at delivering personalized and explainable predictive analytics solution to a research or commercial customer. Understanding and correctly interpreting models for predicting natural and social phenomena, parameterized with a large number of hyperparameters, such as random forests, deep neural networks, or an extreme gradient boosting, can be challenging. The EML is designed to implement the most accurate ML methods and to produce a posteriori explanation using consistent, locally accurate, individualized feature-attribution methods, thus shedding light on problems where human intuition and domain knowledge are often limited.
ClimCoSt	3	Climate Change	National Institute of Geophysics, Geodesy and Geography	BL	The service will produce reliable, comprehensive and detailed evaluations of possible regional/local climate changes and their consequences for different global change scenarios. Metrics and tools for evaluating some of the climate change impacts on environment and quality of will be available. Making use of vast computing resources, the service will enable scientist to perform in-depth assessment of the climate change impacts that cannot be achieved with the desired accuracy using local computing resources. Users are climate scientists and national as well as municipal policymakers.
ClInHealth	2	Geophysical Sciences	National Institute of Geophysics, Geodesy and Geography	BL	The service will generate reliable, comprehensive and detailed studies of the impact of lower atmosphere parameters and characteristics on the quality of life and health risks for the population in our country. It uses a synergetic application of extensive computer simulations on the supercomputer Avitohol, combined with sophisticate analysis of the parameters and characteristics of near surface atmosphere. In this way the impact of the atmosphere on the human health and quality of life can be thoroughly investigated. The users are cross-disciplinary, from the domains of atmospheric physics and environmental science, allergology and epidemiology, as well as national and municipal policymakers.

Table 2: Thematic services for Climate Science and their descriptions

3.2.2. Cross-border use cases in Climate Science: Atmospheric Composition Modelling and Air Quality Forecasting

Scientists from the National Institute of Geophysics, Geodesy and Geography in the Bulgarian Academy of Sciences predict the urban air-quality for cities in the South East Europe (SEE) and Eastern Mediterranean (EM) regions. The scientific groups perform meteorological forecasting using the "WRF-ARW" model and in conjunction with the data from the Serbian thematic service "***DREAM***" dust models, fed into the "***Airpollution Prediction***" service from the Republic of North Macedonia (**Figure 1**) produce significant indicators on the possible future increased pollution levels.

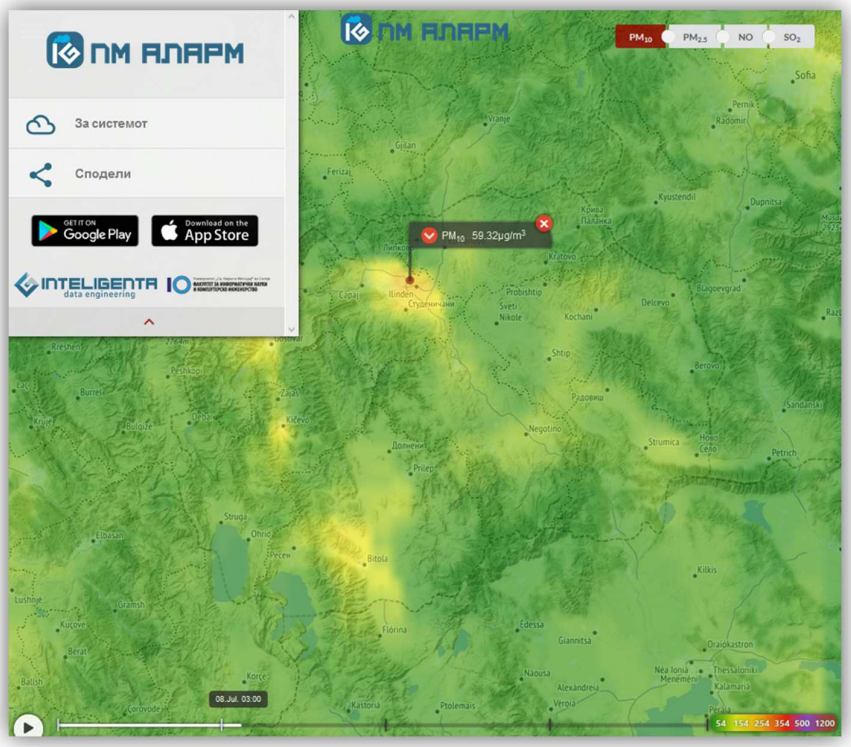

Figure 1: Air Quality Forecast Service over Republic of North Macedonia

The "WRF-ARW" model will be running on CPUs and GPUs of the Greek generic service "***ARIS***" allowing the prediction of atmospheric conditions including air temperature, wind speed and direction, radiation for photolytic reactions, etc. that can influence the increase or decrease of pollution levels.

The Aeolian dust particulate matter model, running simulation on CPUs of the Serbian generic service "***PARADOX-IV cluster***", also complement the weather prediction data, incorporating the dust content in the atmosphere, predicting atmospheric dust episodes and events. Finally, software sourced from the "***Airpollution Prediction***" thematic service is submitted for production on the Croatian generic service "***Isabella***".

All the above mentioned tools require significant storage space. Current EOSC services like "***B2DROP"*** and "***B2STAGE***" allow the input data to be transferred and staged throughout the workflow, while the outputs are to be published as open data on the NI4OS-Europe generic "***Simple storage service***" service and the published manuscripts into the repository service "***Greek Repository service***". This forecast is a great tool in the hands of the authorities who can take targeted measures against high levels of pollution to protect human health, the environment and economy and is available to the public through online and mobile platforms (**Figure 2**).

Figure 2: Mobile application enabled by data sourced from AirForecast service

Air pollution, especially in the urban areas, represents one of the biggest challenges societies are facing. Various measures are proposed by authorities to mitigate this problem. Unfortunately, the measures are usually applied once the pollution hits the critical levels and they only tend to address the current state and only lower the pollution temporarily. A more advanced approach would be to use proactive measures when there is a chance of increased air pollution, so that high pollution levels can be avoided.

Services aiming to test:

☐ [SERV] "Airpollution Prediction"
☐ [SERV] "WRF-ARW"

- ☐ [SERV] "DREAM"
- ☐ [GENR] "ARIS"
- ☐ [GENR] "PARADOX-IV cluster"
- ☐ [GENR] "Isabella"
- ☐ [GENR] "Simple Storage Service"
- ☐ [REPO] "Greek Repository service"
- ☐ [EOSC] "B2STAGE"
- ☐ [EOSC] "B2DROP"

3.3. Digital Cultural Heritage

3.3.1. General description of Digital Cultural Heritage in the NI4OS-Europe area

The archaeology and digital cultural heritage communities have only recently integrated computationally-based tools in their methodologies and practices, and thus started relying on the Open Science Cloud to conduct their research. In the process, their computational needs expanded as the positive impact of computation on the results of these communities' research became clear, enabling scientists, scholars and students to perform comparative studies and analytical methods at unprecedented levels of precision, breadth, resolution and scale, combining big, unstructured data from various disciplines to pursue innovative cross-disciplinary research.

At a time of fractured national and cultural identities, political volatility exacerbated by global economic and political crises, intensifying migrations of people and consequent hostile tensions and violence, archaeology and cultural heritage are facing unprecedented threats and are challenged by neglect, rapid development of built environment and war. Advanced modelling techniques, simulation and data analytics can significantly contribute to not only the study but more importantly the preservation and safeguarding of heritage, as well as of threatened and inaccessible monuments and sites. Bringing together archaeology with open science research, in addition to advanced visualization opportunities for the interpretation of point clouds, 3D models, scanned documents and biological data; modelling and simulation of the complex interactions of natural and cultural landscapes; and data management will offer a new paradigm for the archaeology community towards the generation of open data science.

Data-driven science will encourage intercultural dialogue in Europe, allowing archaeologists to develop new, more holistic understanding of complex historical socio-cultural phenomena influencing policy- and decision-makers, as well as the local authorities. By making heritage accessible and available to wider groups of our society, the Open Science Cloud for archaeology will promote intercultural and intergenerational dialogue by enabling new understandings of the past and more accurate interpretations of historical interactions between human actors, agency and the environment. These breakthrough contributions that are fundamentally interdisciplinary will have a great impact not only on the culture and policies of the country, but will also benefit tourism, urban planning, and the environmental and societal sustainability of the European urban and rural environments.

Applications and services that will enable *safeguarding Cultural Heritage* activities in the region include:

1. Enabling, facilitating and managing massive documentation and digitization processes.
2. Curating big data by means of semantic organisation of unstructured data.
3. Online interaction and visualization interfaces for the interpretation and study of archaeology and cultural heritage data such as point clouds, images, photographs, 3D models, scanned documents, chemical and biological data.
4. Modelling and simulating the complexity of historic built environments, archaeological excavation sites and monuments, as well as the production of cultural artefacts.
5. Digital collections access, data interoperability and services.

A collection of thematic services which belong to the field of DCH are shown in table below.

Name of service	TRL	Scientific Domain	Institution	Country	Description
EpHEMERA	8	Digital Cultural Heritage	The Cyprus Institute	CY	The World Heritage Convention, drawn by various international bodies in 1972, was designed to protect cultural or natural places of outstanding universal value so that future generations may be able to enjoy them. Responding to these principles as well as to the Charter on the Preservation of Digital heritage (Vancouver, 2003), this multidisciplinary project, which involves archaeologists, art historians, conservators and computer scientists, aims to create an open access, 3D interactive online geo-database of endangered architectural and archaeological heritage in the South Eastern Mediterranean basin; a region of tremendous cultural importance whose rich heritage is unfortunately threatened by both natural and human factors. A wide range of 3D modelling and topographic techniques has been applied to create accurate reconstructions of heritage sites, enriched by a extensive array of metadata. The Online 3D Database System for Endangered architectural and archaeological Heritage in the south Eastern MEediterRAnea area (EpHEMERA) is intended to serve as an infrastructure where it is possible to: Visualize online and through standard web browser 3D architectural and archaeological models classified according to a specific type of risk; Query the database system and retrieve metadata attached to each single virtual object; Extract geometric and morphological information.
Clowder 4DCH	8	Digital Humanities	The Cyprus Institute	CY	Clowder4DCH, a highly extensible active curation-based research data management platform. It contains three major extension points: preprocessing, processing and previewing. When new data is added to the system, preprocessing is off-loaded to extraction services which extract appropriate data and metadata. The extraction services attempt to extract information and run preprocessing steps based on the type of the data, e.g., to create previews. This raw metadata is presented to the user via a web interface. Users can upload, download, search, visualize research datasets and explore information linked to data. Users can link and organise datasets in online collections following the provided workflows for creating semantically structured data repositories specialized for digital cultural heritage. It enables users to form an online collaboration environment to support research communities and activities, and disseminate results.
Online Virtual Reality Environments Toolkit (OVRET)	5	Cultural Heritage	The Cyprus Institute	CY	Users will be able to use the provided software tools, workflows and plugins (scripts of code) to create immersive environments for virtual museums, interactive collections of cultural artefacts, visual interfaces for geolocated interaction with DCH assets in physical space through mobile devices, and virtual visits of inaccessible, or demolished, heritage, monuments and historic sites.

Table 3: Thematic Services for Digital Cultural Heritage and their descriptions

3.3.2. First cross-border use case in Digital Cultural Heritage: Virtual Sites and Museums

Digital Cultural Heritage**:** The EOSC-related needs in this field consist of creating, processing and accessing Cultural Heritage data; with data management and visualization - especially for 3D models – being of paramount importance. The NI4OS-Europe DCH thematic service portfolio, shown in Table 3, will enable not only interdisciplinary studies of archaeo-environments and past societies to key players, such as the researchers of the Cyprus American Archaeological Research Institute (http://caari.org), and the Cyprus Department of Antiquities, with background in archaeology, anthropology, history, geography, biology and urban studies, but also the creation of Virtual Museums and the advancement of Digital Management of Archaeological and Historical Sites.

For example, conservators and museum curators in the region and Cyprus (e.g., Leventis Municipal Museum, Nicosia) want to create Virtual Environment and Virtual Museum experiences, respectively, for the study and representation of sites collected from archaeological fieldwork in appropriate form to ensure interoperability of digital assets for data management in repositories which is of paramount importance. Exemplar cases will rely on generic services to accelerate the semi-supervised automated creation of a large number of 3D models of complex objects of cultural heritage, the "***CHERE***" service for more simplified objects, and the "***OMApp***" service for the documentation, analysis and reconstruction of whole sites and historical settlements. The next step can be exemplified by the exploitation of the tools offered by the "***OVRET***" service to leverage the "VI-SEEM RIVEEL3D" workflows to enable interaction with DCH assets in physical space through immersive experiences, offer virtual visits to inaccessible, or demolished, heritage, monuments and historic sites, and create Virtual Museums and Sites with interactive collections of cultural artefacts, respectively. Relevant datasets will be organized respecting the FAIR policies and will become open to the public via "**Clowder4DCH**" service. The generated data and metadata will be integrated in the "NI4OS-Europe Simple Data Repository", and existing EOSC "***B2SHARE***", "***B2STAGE***" services.

Specifically, the service users will be provided with software tools, workflows and plugins (scripts of code) to create immersive environments for virtual museums, interactive collections of cultural artefacts, visual interfaces for geolocated interaction with DCH artefacts (cultural heritage assets) in physical space, and virtual visits of inaccessible, or demolished, heritage, monuments and historic sites. Target user community includes the GLAM industry, scholars, researchers and students in digital cultural heritage. The value of the service is high for user communities both for research and education and dissemination purposes. The direct engagement of the user of Virtual Reality applications, and their capacity for data visualization and interpretation is testified by its exponential adoption by the entertainment, visual arts, architecture, social media and marketing industries internationally. The service will be free for European (and international) researchers. Regarding access policy, users will be able to install and run the platform locally, to access their own data and data other users share with them, as well as to control access to their data. User support for this service will be provided, including training material and existing user fora. Below we provide a list of data sets, which will be produced by the service:

a. Customized workflows for the 3D documentation and reconstruction of artefacts, structures (e.g., monument buildings) and sites, for the creation of 3D models. Separation between asset creation with photogrammetry and building or open-air heritage site, where a mix of laser scanner or/and drone aerial mapping will be required.

b. Unity applications for visualizing and interacting with virtual environments using multiple interface modes, such as mouse and keyboard, XR, Oculus Rift, and HTC Vive.

c. Interface for locomotion within the virtual environment.

d. Basic application features such as console commands, GUI interfaces, render quality settings.

e. Virtual scenes loading and toggling multiple time phases using an in-application GUI; object observation tools.

f. Samples of DCH assets (3D reconstructions of cultural artefacts, monuments, buildings or sites). Access rights to be set by copyright owners, if needed.

Services aiming to integrate:

☐ [SERV] "CHERE": https://services.vi-seem.eu/ui/catalogue/services/
☐ [SERV] "Clowder4DCH"
☐ [SERV] "OVRET"
☐ [SERV] "Open Mapping Application (OMApp)"
☐ [EOSC] "B2STAGE"
☐ [EOSC] "B2DROP"

3.3.3. Second cross-border use case in Digital Cultural Heritage: Clowder4DCH

The scope of the "***Clowder4DCH***" service is to offer a highly extensible active curation-based research data management platform. It contains three major extension points: preprocessing, processing and previewing. When new data is added to the system, preprocessing is off-loaded to extraction services for extracting appropriate data and metadata. The extraction services attempt to extract information and run preprocessing steps based on the type of the data, e.g., to create previews. This raw metadata is presented to the user via a web interface. Users can upload, download, search, visualize research datasets and explore information linked to data. The target user community for this service includes the GLAM industry, scholars, researchers and students in digital cultural heritage. Users can link and organize datasets in online collections following the provided workflows for creating semantically structured data repositories specialized for digital cultural heritage. It enables users to form an online collaborative environment to support research communities and activities, and disseminate results. The service will be free for European (and international) researchers. Regarding access policies of the service, when this is installed and becomes available to relevant community, users will be able to register on the platform using a form. The responsible Clowder application administrator will check user information and then manually accept users. Following Password Authentication process, users can see their own data and data other users share with them. Users can control access to their data in various ways and at different granularities (datasets, collections and spaces). There is user support for this service, including training material and a user forum.

With the Clowder4DCH service, users can upload massively (zipped datasets) or individual files of:

a. 3D Models: integrated extractors prepare files for interactive, online visualization on the platform itself featuring geometry analysis tools (e.g., measuring dimensions) and rendering (e.g., control lighting).

b. Scanned books and their metadata: OCR algorithms will be used to extract the text in the documents so that users can find books using both metadata information and the book's contents.

c. Image, text, video and sound files and their metadata.

In Clowder4DCH service users can search, tag, annotate data at various granularities. Datasets will respect H2020 FAIR policies. Metadata can follow the CIDOC-CRM RDF, or ARC2 triple store, ISBD–M, and UNIMARC standards. Metadata standardization, e.g., Dublin Core, can be applied.
Regarding EU GDPR principles, users' personal data are not shared or processed by anyone. Users can process their personal data and information, and if requested all of their personal information can be removed from the server.

Services & Datasets integrated:

☐ [SERV] "Clowder4DCH"
☐ [DATA] "University of Maribor Library Digital Repository"
☐ [DATA] "Digital Archive for Ethnological and Anthropological Resources"

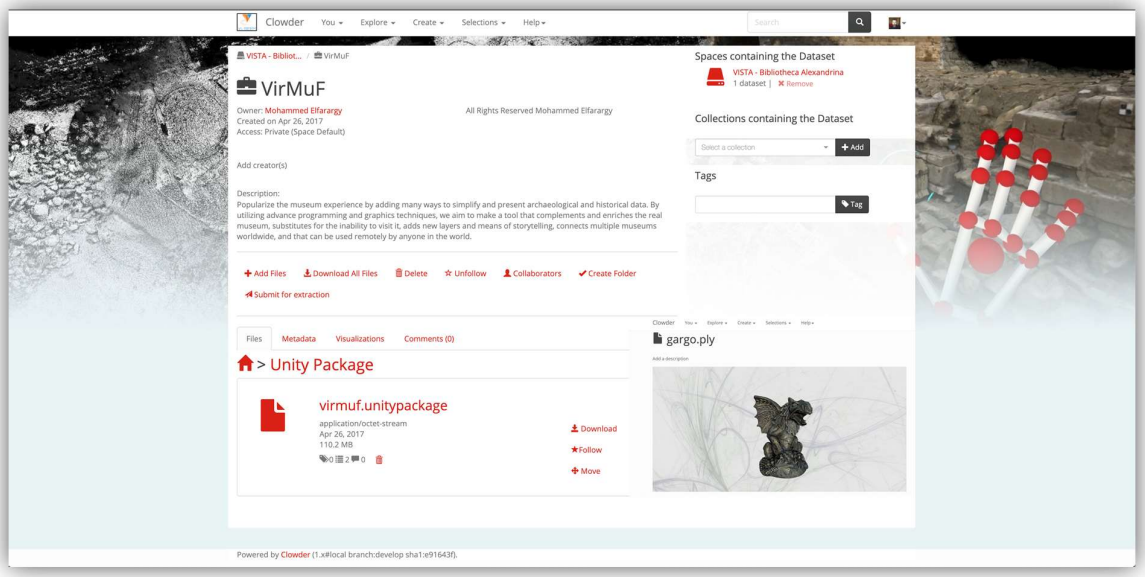

Figure 3: VI-SEEM online digital cultural content management system based on Clowder. This version allows users to store, manage, share and interact with content, including cultural heritage digital assets, such as 3D models of artefacts and their metadata, of museum collections.

3.4. Additional Scientific Communities

3.4.1. Identification of additional Scientific Communities

As it has been explained, in addition to the three pre-defined scientific communities of Life Sciences, Climate Science and Digital Cultural Heritage, an additional scientific community has been identified. By looking at the Table 4 which presents the thematic services which are expected to be on-boarded and do not belong to the three pre-defined communities we realize that four of them belong to the broad scientific community of Computational Physics/Chemistry. Although they are at the lower TRL level of 7, the service developers have ensured NI4OS-Europe consortium that the services will be ready for on-boarding soon.

We, therefore, chose the interdisciplinary scientific community of computational physics in addition to the other three scientific areas. An example of a use case on computational physics is provided in the next section.

Name of service	TRL	Scientific Domain	Institution	Country	Description
VideoLectures.NET	9	multidisciplinary	Jožef Stefan Institute, Ljubljana, Slovenia	SL	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 contents 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.
CPMot	7	Computational Physics	Institute of Physics Belgrade	RS	Density functional theory (DFT) is currently the most useful method for calculation of electronic structure of large molecules and materials. The basic quantity within DFT is electronic charge density. In typical DFT calculations one has to self-consistently solve the Kohn-Sham equations for single particle wave functions and the equation that expresses electronic charge density in terms of wave functions of occupied electronic states. Charge patching method (CPM) is the method based on DFT which avoids self-consistent calculations of electronic charge density. It is based on the idea that electronic charge density in the neighborhood of a certain atom depends mainly on its local environment and that it is the same in a large system that one wants to calculate and in some small system where that atom has the same local environment. These contributions are called charge density motifs. These motifs are essential input for electronic structure calculations using CPM. The CPMot service provides motifs for a variety of semiconducting and insulating materials and nanostructures.

GQL	7	Computational chemistry	Institute of Physics Belgrade	RS	Guanine(G)-quadruplex is considered relevant for various cellular processes. Recent research indicates that it represents an active site in antitumor treatment, since the formation of a G-quadruplex-ligand complex is expected to prevent further cell division. GQL service identifies ligands that efficiently bind to G-quadruplex. The main advantage of GQL service is that it is based on analysis of recently (in 2017 and 2018) synthesized ligands (squaraine-based compounds, benzimidazole-carbazole molecules, and complexes with transition metals) which have not been thoroughly investigated. The service monitors interaction between a drug (ligand) and selected G-quadruplex structures from the PDB database using intensive molecular dynamics simulations. As a result, GQL produces datasets with the equilibrated molecular structures, which provide additional information to the available experimental NMR and X-ray diffraction structures, which are used as initial structures for further studies.
LMDB	7	Computational physics	Institute of Physics Belgrade	RS	LMDB is a distributed database of numerical results and codes pertaining to a large class of condensed matter theoretical models, namely the correlated lattice models. Correlated lattice models are used for the study of crystalline materials exhibiting high-temperature superconductivity, antiferromagnetism, Mott insulating states and various other emergent phenomena. These models abstract the details of the ionic lattice, but take account of the Coulomb interaction between the electrons. Therefore, they pose a quantum many-body problem which is intractable in general, but can be solved using a variety of numerical methods, to different levels of approximation. Typical many-body calculations take thousands of CPU hours, so sharing existing results is essential for the efficiency and making further progress in the field. LMDB provides a natural and high-level systematization of the models and methods which allows people to effortlessly share and view other people's results and reuse codes. Even more importantly, the service allows for automatized analyses of large datasets which is unprecedented in the field. The service consists of a central server which tracks numerical data on the data servers, hosted on HPC facilities where the data is produced and stored. The central server provides a web API for the communication with the database. Additionally, the central server provides a flexible web-based GUI for browsing and plotting results. A typical small-scale usage scenario involves no programming, yet allows the user to combine results from multiple remote sources into new analyses.
SciRoHub	7	Earth Observation	National Institute for Research and Development in Informatics (ICI)	RO	The Service provides Copernicus's Sentinel1 and Sentinel2 products that cover the Romanian territory. Sentinel1 mission provides all-weather, day-and-night radar imaging for land and sea. Sentinel2 provides high resolution optical image data, including monitoring of vegetation, soil and water cover, as well as observation of inland waterways and coastal areas. It targets researchers from Romanian National Research Institutes and Universities that are interested in using satellite images for their research purposes. The Service is free to use for Romanian academic and research community. Registration is mandatory in order to access the data.

Schrodinger	7	Computational Physics	Faculty of Computer Science and Engineering and Faculty of Natural Science and Mathematics, Ss. Cyril and Methodius University in Skopje	MK	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 rigorous quantum description of the underlying processes. This, on the other hand, requires solving of either the stationary (time-independent) or the time-dependent Schrödinger equation. The effort required to complete such a task 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 solution of stationary and time-dependent Schrödinger equation, implementing several algorithms. The main intention would be to rely on series of datasets generated by quantum mechanical computations, while the actually implemented algorithms will base on either sequential interpolation – variational calculation or on some variant of discrete variable representation technique (DVR), Fourier-grid Hamiltonian approach etc.
Gaussian	7	Computational chemistry, computational physics, computational biology and life sciences, computational materials science.	Faculty of Computer Science and Engineering and Faculty of Natural Science and Mathematics, Ss. Cyril and Methodius University in Skopje	MK	Gaussian regression, 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, these approaches have been shown to be rather useful for 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. Therefore, the proposed service would provide a user-friendly (as much as possible) environment for development and application of the Gaussian regression technique, along with certain other machine learning techniques, to predict physiological activity, various molecular and materials' properties, phase diagrams of complex materials, and also to compute their basic structural and spectroscopic properties, primarily on the basis of the results from molecular dynamics simulations. We strongly believe that such service could provide a useful platform that could be also used for standardizing the most contemporary machine learning techniques accounting for specificities in various areas.
AstroMatch	4	Astronomy	Institute of Physics Belgrade	RS	The AstroMatch service provides astrometric calibration of astronomical images taken with any telescope which contain celestial coordinates of a central (or any other) pixel and an approximate pixel scale. The service uses the Digitalized Sky Survey (http://archive.stsci.edu/cgi-bin/dss_form) to access archived astronomical images from the First and the Second Palomar Sky Survey and the Sloan Digital Sky Survey (https://www.sdss.org/dr14/imaging/imaging_access/) when possible. Astronomical images are matched to archived images of the same field of view through a set of translations and rotations until the objects overlap, which provides an initial astrometric solution. Next, all the objects from the input (uncalibrated) images are selected and searched for in many existing star catalogs to refine their celestial coordinates. Finally, the improved astrometric calibration is calculated and the WCS (World Coordinate System) solution is provided, and may optionally be written to the header of the images.

IoT Cloud Platform	4	Interdisciplinary	University of Montenegro	ME	IoT Cloud Platform will enable users to collect data from internet-connected hardware, and visualise them in near real-time. Collected data will be able to be analysed online in Octave or R programming language (discovering relationships, patterns and trends in data), whereby users will be able to use a pre-written algorithm or to develop a new code directly in the web browser. IoT platform will support a variety of hardware devices, such as Arduino and Raspberry Pi platforms, mobile devices, PCs, etc. Beside web interface, API will be provided, so that data can be uploaded, stored, and accessed from the third party.
Cyber Security Incidents Analytics	4	Security	Computer Emergency Response Team of Moldova	MD	Cyber Security Incidents Analytics project, that in fact will be implemented as cyber security platform, should collect all security incidents data of a network, process it, enhance the data and make it available for analyze using one big data mechanisms. Main Idea of this platform is to give access to the related incidents data analysis for large number of users based on their permissions, that means that both small ISP or organizations will have access to their incidents as well as governmental or international CERT's.

Table 4: The additional thematic services which do not belong to the three pre-defined scientific communities of Climate, Life and Digital Cultural Heritage and have TRL higher than 4

3.4.2. Thematic use case 4: Service for advanced methods for solving of multidimensional stationary and time-dependent Schrödinger equation

In-depth understanding of numerous phenomena in different scientific disciplines, such as chemistry, physics, materials science etc., require exact quantum mechanical analysis of the underlying processes. In such cases, therefore, it is often necessary to solve the stationary or time-dependent Schrödinger equation in one, two or three spatial dimensions. For example, many problems in the aforementioned areas may be solved by analyzing an oscillatory motion in one, two or three dimensions. In a general case, such motions are often highly anharmonic and the resulting vibrational Schrödinger equation is not analytically solvable. Treatment of the problems starting with the simplest one-dimensional case and extensions to two- and three-dimensional cases is often required in order to consider a range of e.g. inter- and intramolecular interactions relevant to the phenomenon in question in a systematic and sequentially improving manner. It is therefore helpful to have a series of codes designed for such purpose, based on the same algorithm, so that a particular problem can be treated at several levels of increasing complexity with the same numerical approach, enabling the researchers to focus on the physical problem instead of the algorithmic and accuracy issues.

The team led by Professor Ljupco Pejov from the Ss. Cyril and Methodius University in Skopje, North Macedonia will be running a use case based on the "***Schrödinger"*** thematic service designed to provide methods for solving the Schrödinger equation in one, two or three dimensions. The Discrete Variable Representation method (DVR) is implemented to solve the quantum eigenvalue problem for all cases. As the realistically encountered physical problems may require different types of potential energy surfaces (potential energy parts of the vibrational Hamiltonians), the used service offers several standard and widely used types of potentials, with a certain flexibility to adjust these to the particular needs of the user. The user team can also define their own potential for each particular purpose of the study. For a particular research-grade study, this often means that the user

should carry out a series of computations to generate the potential energy of the considered problem, most frequently followed by an appropriate numerical fitting.

The results extracted from this service are of the interest of researchers working in various different fields such as chemistry, physics, materials science, etc. The service will be free for European (and international) researchers. Regarding access policy users will be able to consume it as RESTful API service which means they can integrate it in their own application or use it as a URL link. Users can provide their own input data.

The service will be hosted on the "***FINKI-Cloud"*** generic service, one of the quick-on-boarding generic services. The service is Openstack based cloud, providing resources to the thematic services to be on-boarded to ESOC through NI4OS-Europe project. The produced datasets will be published at the "***FINKI-Archive"***, a service based on ArchiveSpace that will be used to support multiple scientific communities to share scientific data.

Services integrated:

☐ [SERV] "Schrödinger"
☐ [GENR] "FINKI-Cloud"
☐ [REPO] "FINKI-Archive"

4. How stakeholders are involved in cross-border cases

One of the aims of the cross-border cases is to strengthen the involvement of the NI4OS-Europe stakeholders in the NI4OS-Europe several tasks. Hence, it is expected that the stakeholders of NI4OS-Europe will be actively participating in the use cases. This is crucial since it will indicate at what level the NI4OS-Europe partners can work together in order to support open science in the region.

Firstly, the NI4OS-Europe PMB members of each participating country will be responsible for bringing in contact the scientific team executing a cross-border case with all the service providers whose services are participating. All the relevant contact details have been gathered by the PMB members during the survey procedure, as well as during the collection of the service template descriptions in the framework of the deliverable D5.1 "*Provider landscape analysis and provider categorization*". This will enable the scientific teams to learn more details about the services and their implementation and, on the other hand, it will enable the service providers to fine-tune their services to the needs of the users. If the communication between the scientists and the service providers cannot be achieved, the PMB members should report to the WP6 and WP5 leaders in order for them to take further actions. It is important that a good communication level between the service providers and the on-boarding team is maintained, ensuring, thus, the success of the on-boarding of services.

Furthermore, a number of NI4OS-Europe stakeholders are providing the generic services. These partners are actually responsible for providing all the user teams participating in the cross-border cases access to the aforementioned services. This involves access to HPC machines, cloud machines and general services. Their involvement will be important for the successful completion of the use cases. Namely, the PMB members should contact the administrators of the generic services and make sure that all the support is given to the users in order to successfully access and facilitate the generic services. An example of such an involvement is the support given, and the computational resources which have already been allocated to the Bioinformatics European Research Era Chair and the Bioinformatics Group at the Cyprus Institute of Neurology and Genetics. The research team has put their efforts in the multi-omics analysis and network-based integration towards a highly-informed decision regarding a short list of repurposed drugs and related to them natural products against COVID-19. Throughout the Fast Track Access to NI4OS-Europe resources towards the fight against covid-19 (https://ni4os.eu/ni4os-europe-vs-covid19/), NI4OS-Europe partners in IPB, Serbia provided computational resources and all the required support for optimizing the codes used to achieve the completion of this investigation. This example serves as a first test of a cross-border case applied on NI4OS-Europe services, and appears to be rather successful.

Finally, the PMB members of a country with a scientific team participating in the use cases will be responsible for reporting the progress of the cases to the WP6 leader. This will be a continuous procedure.

5. Conclusions

The on-boarding of thematic services, generic services and repositories on the European Open Science Cloud is something that needs to be tested properly. To this purpose NI4OS-Europe will run use cases which will involve mainly the set of thematic, generic and repository services which will be on-boarded on the first round of on-boarding. This procedure will enable NI4OS-Europe consortium to test the on-boarding of the services and in addition to fine tune the services.

To this purpose, NI4OS-Europe have set up a number of cross-border cases focussing on four highly cross-disciplinary fields of research. These scientific areas include the pre-defined communities of Climate Science, Life Science and Digital Cultural Heritage and in addition Computational Physics. Life Science is contributing with one use case, Climate science is contributing with one use case, Digital Cultural Heritage is contributing with two use cases and Computational Physics is contributing with one use case. The description of the above use cases is provided in this deliverable.

It is expected that by M15 of the project, the first set of the three pre-defined communities as well as of Computational Physics are provided access to the aforementioned services for testing.

