H2020-INFRAEOSC-2018-3

NI4OS-Europe

National Initiatives for Open Science in Europe

Deliverable D6.4

Service evaluation by user communities

Lead beneficiary(s): CYI

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

Status –Version: Final – k

Date: May 18, 2021

Dissemination Level: Public

Abstract: Deliverable D6.4 – Service evaluation by user communities - deals with the execution of demonstrators by the user communities. Specifically, we report the evaluation of the first set of on-boarded services by the user communities. Thematic services belonging to the Life Science, Climate Science, Digital Cultural Heritage, as well as Computational Physics communities, together with generic and repository services were used as use cases by researchers from all user communities. These cross-border use cases were described and reported in Deliverable 6.3 - User communities' use cases. The researchers provided brief reports detailing their experience using these services, with an emphasis on the ease of access and usage. In this deliverable, we also discuss briefly the status of generic, thematic, and repository services which are or will be on-boarded to NI4OS-Europe.

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

The NI4OS-Europe project is co-funded by the European Union Horizon 2020 programme under grant number 857645.

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
January 26, 2021	a	Chrysovalantis Constantinou	Initial ToC
March 30, 2021	b	Chrysovalantis Constantinou, Alexis Chatzigoulas, Aspasia Vozi, Michail Papadourakis, Dimitris Papakonstantinou, Zoe Cournia, Ljupco Pejov, Anastas Mishev, Bojana Kotetska	Adding the life science and computational physics use cases
April 19, 2021	c	Chrysovalantis Constantinou	Editing
May, 9, 2021	d	Theodoros Christoudias, George Mikuchadze, Hrachya Astsatryan, Boro Jakimovski	Adding the climate science community input
May 10, 2021	e	Chrysovalantis Contantinou, Andreas Athenodorou	Editing
May 11, 2021	f	Chrysovalantis Constantinou, Iason Giraud, George Artopoulos	Adding the DCH community's input
May 13, 2021	g	Chrysovalantis Constantinou, Andreas Athenodorou, George Artopoulos	Editing
May 14, 2021	h	Chrysovaslantis Constantinou, Andreas Athenodorou, George Artopoulos, Andreas Gavrielides	Editing, add information on Computational Physics use case, add information on RePol use case
May 14, 2021	i	Chrysovaslantis Constantinou, Andreas Athenodorou	Final editing
May 16, 2021	j	Milutin Radonjić	Internal quality check
May 18, 2021	k	Chrysovalantis Constantinou	Addressing quality check remarks

Table of contents

References

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

[2] H. Astsatryan, H. Grigoryan, A. Poghosyan, R. Abrahamyan, et al., Air temperature forecasting using artificial neural network for Ararat valley, Springer Earth Science Informatics, DOI: 10.1007/s12145-021-00583-9, 2021.

[3] Sh. Asmaryan, V. Muradyan, et al., Paving the way towards an Armenian Data Cube, MDPI Data, Special Issue "Earth Observation Data Cubes", 4(3), 2019 DOI: 10.3390/data4030117.

[4] N. C. Polfer, Chem. Soc. Rev. 40, 2211 (2011).

[5] M. Okimura, L. I. Yeh, J. D. Myers, Y. T. Lee, J. Chem. Phys. 85, 2328 (1986).

[6] M. Okimura, L. I. Yeh, J. D. Myers, Y. T. Lee, J. Chem. Phys. 94, 3416 (1990).

[7] V. Kocevski, Lj. Pejov, J. Phys. Chem. A, 116, 1939 (2012).

[8] V. Kocevski, Lj. Pejov, J. Phys. Chem. A, 114, 4354 (2010).

[9] V. Barone, J. Chem, Phys., 122, 014108 (2005).

[10] A. Masson, E. R. Williams, T. R. Rizzo, J. Chem. Phys., 143, 104313 (2015).

[11] A. D. Becke, J. Chem. Phys., 98, 5648 (1993).

[12] C. Lee, W. Yang, R. G. Parr, Phys. Rev. B, 37, 785 (1988).

[13] T. Yanai, D. P. Tew, N. C. Handy, Chem. Phys. Lett., 393, 51 (2004).

[14] G. Zheng, H. Witek, P. Bobadova-Parvanova, S. Irle, D. G. Musaev, R. Prabhakar, K. Morokuma, M. Lundberg, M. Elstner, C. Kohler, T. Frauenheim, J. Chem. Theory and Comput., 3, 149 (2007).

[15] C. Peng, P. Y. Ayala, H. B. Schlegel, M. J. Frisch, J. Comp. Chem., 17, 49 (1996).

[16] H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, G. E. Scuseria, A. D. Daniels, M. J. Frisch, J. Chem. Phys., 114, 9758 (2001).

[17] S. S. Iyengar, H. B. Schlegel, J. M. Millam, G. A. Voth, G. E. Scuseria, M. J. Frisch, J. Chem. Phys., 115, 10291 (2001).

[18] S. Grimme, S. Ehrlich, L. Goerigk, J. Comp. Chem., 32, 1456 (2011).

[19] F. A. Hamprecht, A. Cohen, D. J. Tozer, N. C. Handy, J. Chem. Phys., 109, 6264 (1998).

[20] A. D. Boese, N. L. Doltsinis, N. C. Handy, M. Sprik, J. Chem. Phys., 112, 1670 (2000).

[21] Effect of Dust Aerosols in Forming the Regional Climate of Georgia, May 2018, Scalable Computing 19(2):87-96, DOI: 10.12694/scpe.v19i2.1398

List of Figures

List of Tables

List of Acronyms

AFMM	Automated Frequency Matching Method
AI	Artificial Intelligence
ARIS	Advanced Research Information System
DCH	Digital Cultural Heritage
DICOM	Digital Imaging and Communications in Imaging
DNA	Deoxyribonucleic Acid
DREAM	Dust Regional Atmospheric Model
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 Management Board
REPO	Repository
RoP	Rules of Participation
SERV	Thematic Service
WP	Work Package

Executive summary

What is the focus of this Deliverable?

Deliverable 6.4 – "Service evaluation by user communities" – discusses use cases run by the Life Science, Climate Science, Digital Cultural Heritage, and Computational Physics communities. Each scientific community assigned a small group of researchers with the task of requesting access and using thematic services related to the scientific area under question, but they also tested generic services, as well as repositories. The access was requested via the NI4OS-Europe helpdesk or by directly conducting the service providers. Some scientific communities produced results which will be used in their scientific research pipeline.

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

The service evaluation performed at this stage of the project will help inform service providers about loopholes in their helpdesk systems, problems with the services themselves, etc. The researchers involved in the demonstrations reported not only the ease of usage of the services but also their relevance and usefulness in the scientific sense. Thus, future service providers can study this deliverable and reach conclusions on how to improve their services. Finally, after this deliverable is submitted the on-boarding of services will continue and an open call will be issued, which will provide guidelines on how European researchers will be able to access and use the on-boarded services.

What are the deliverable contents?

The deliverable provides a brief introduction detailing the status of the on-boarded generic, thematic, and repository services and provides use case reports from the Life Science, Climate Science, Digital Cultural Heritage, and Computational Physics communities. Following this brief introduction, we provide summary of the reports as written and delivered to us by the researchers belonging to the four user communities mentioned above. Furthermore, we provide results of surveys tailored in a way that can reveal critical loopholes in the implementation of the on-boarding as well as in technical functionalities of the service.

Conclusions and recommendations

This deliverable demonstrates that the first set of services which have been fully on-boarded on NI4OS-Europe meet the standards to belong in this category. Also, most of the candidate services expected to be on-boarded in the near future have demonstrated no weaknesses, however some fine-tuning needs to be done before these moves to the next stage of on-boarding. Furthermore, services which will soon become candidates for on-boarding have been tested without demonstrating any serious faults which would prohibit their inclusion in the service catalogue. All critical problems encountered throughout this evaluation will be treated carefully. This means that within a time frame of approximately twelve months, before the open call sets in, NI4OS-Europe should work closely with service providers in order to resolve any issues.

1. Introduction

In Deliverable 6.3, a set of use-cases designed by the scientific community leaders of the life science, climate science, digital cultural heritage, and computational physics communities are described. The use cases aim to test as many thematic, generic, and repository on-boarded services, hosted in as many different NI4OS-Europe countries as possible, and involve as many researchers from different countries as possible. At the same time the use cases are designed and executed in a way that that they will resemble a pipeline used in actual research projects by the aforementioned research communities. The idea behind the deliverable is to simulate a realistic environment of service usage and test whether scientists as well as different kinds of stakeholders can use the services in the way these are on-boarded in NI4OS-Europe. This will enable a realistic testing of the on-boarding of services and aims to investigate whether the on-boarding has been performed properly, the services can produce the expected results, and if scientists can get the needed support whenever this is necessary.

Each scientific community leader worked closely with a team of researchers that were either assigned with the task of testing specific services and report their findings in brief reports that they made available to us, or provide reports on their experience using services of the NI4OS-Europe consortium. Since NI4OS-Europe has services on-boarded in the service management system and scientific teams have already made use of such services, this approach would have maximized the feedback from actual users.

The Scientific community leaders which participate in the use cases are the following:

- Life Sciences: Professor Zoe Cournia (BRFAA)
- Climate Sciences: Professor Theodoros Christoudias (CYI)
- Digital Cultural Heritage: Professor Georgios Artopoulos (CYI)
- Computational Physics: Professor Bojana Koteska (UKIM)

The use-cases have a strong cross-disciplinary as well as cross-border character and are designed so that they involve as many partners from the NI4OS-Europe consortium as possible. Through this effort we managed to cover a large geographic area of service providers as well as users, and mobilize a significant sector of NI4OS-Europe partners.

A large set of thematic services, generic services, and repositories was tested. In addition, an ORDM tool (RePol) has been tested. In a few cases (Bioconnect service) the services were not fully on-boarded thus the service provider worked closely with the researchers to assist them in using the service. In other cases, such as the NI4OS-Europe repository service or the FINKI-cloud service, access was given after the researchers conducted the helpdesk. Still in other cases, the researchers were able to simply login and work with the services with no intervention by the helpdesk, the task leader Dr Chrysovalantis Constantinou, or the WP6 leader Dr Andreas Athenodorou.

To aid with the identification of services which are potentially useful in a use case scenario, a survey was contacted amongst the members of the climate science community, which belongs to the NI4OS-Europe consortium, in order to identify current and past users of services which are currently on-boarded on NI4OS-Europe. The users who completed the survey were then asked to provide a brief report, providing details about the services they used, the issues they faced while using the services, the usefulness of those services in

the scientific sense, etc. The members of the Digital Cultural Heritage community conducted a similar survey. The computational physics community conducted a different kind of survey. Specifically, in the computational physics community survey, users were asked to test the thematic service Schrodinger and the generic service FINKI-cloud and fill in a survey with questions related to the two services. Suggestions were also made regarding the improvement of the two services. For the creation of different surveys, the WP3, WP4, WP5 as well as WP6 leaders have been strongly involved.

Each scientific community leader adopted a different approach in order to carry out a use-case. For instance, life science community leader has set up a team of young researchers who evaluated in detail each different service. This resulted to a very detailed report which is presented in Section 2. Climate science use case (Section 3) was based on the users' engagement initiated by a survey conducted amongst the NI4OS-Europe partners working on this topic. Digital Cultural Heritage scientific community leader approached a large consortium of DCH stakeholders and asked to use the services under evaluation and assess them by filling out a questionnaire; this is presented in Section 4. Computational physics use case (Section 5) adopted a hybrid approach, namely they used a similar perspective as that of DCH and in addition they assigned a research project, making use of the services under testing, to an established research team in the field. Finally, the use case on the RePol, presented in Section 6, resulted by the needs of an institution to create a repository as well as by the intensive training efforts of NI4OS-Europe to inform repository administrators on ORDM tools and repository policies.

Overall, an effort was made to test as many services as possible and report all the problems that arose. This will enable the NI4OS-Europe consortium to understand the weaknesses of the on-boarding, the difficulties scientific teams face when using the services, the level of support a user can get and, hence, to fine-tune the solutions provided by NI4OS-Europe and consequently by EOSC. Below we provide the reports researchers who belong to the four scientific communities provided us with.

2. Life Sciences use case

With this section we begin presenting the different use cases, starting from that of life sciences. For each different use case we provide a list of all the services tested and the evaluation of these services by the teams. After we finish with life sciences, we proceed with the climate science community, the Digital Cultural Heritage Community, with the computational physics community and we close with a use case focusing on the ORDM tool RePol which is expected to become on-boarded soon.

2.1. Use case description in Life Sciences

In D6.3, a description of a use case for the Life Science community titled cross-border case: "Extracting correlations for patient stratification using machine learning" was described.

The idea behind the use case is that 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 requires 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**". A

pictorial perspective of the pipeline one needs to follow in order to execute the aforementioned use case is provided in Figure 1.

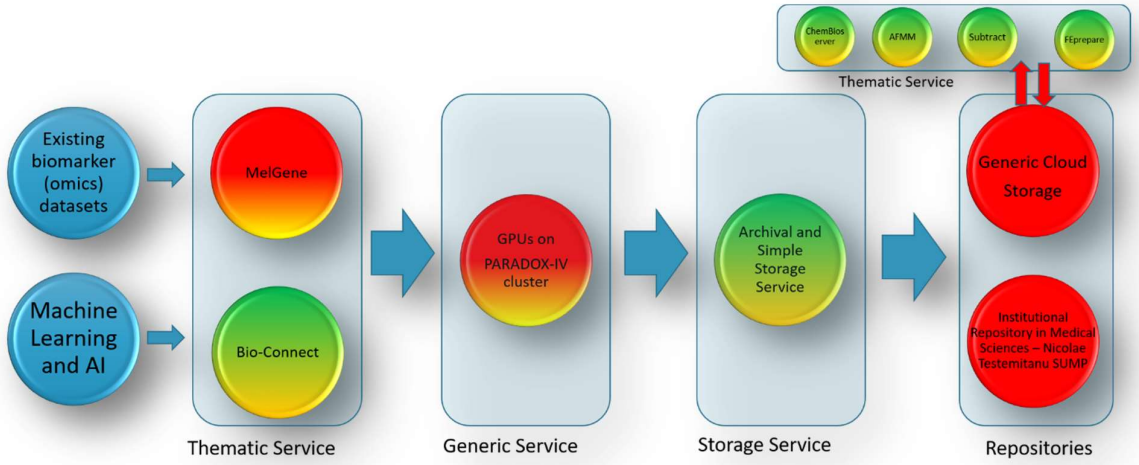

Figure 1: A picture representing the pipeline one has to follow to execute the use-case in life sciences

The scientific community leader of Life Sciences set up an inter-disciplinary team of four researchers with diverse backgrounds. Some of these researchers have more experience than others. These researchers were called to test whether the proposed scenario for the use case could have become real by testing all the NI4OS-Europe services participating in the demonstration. All the services tested are relevant to the **Life Science community** and are part of the NI4OS-Europe platform. The services in Table 1 were expected to be evaluated with respect to their **accessibility**, **quality** and **efficiency**:

Number	Name of service	Providing Country
1.	Bio-Connect	Cyprus
2.	MelGene	Cyprus
3.	PARADOX-IV cluster	Serbia
4.	ChemBioServer	Greece
5.	FEPrepare	Greece
6.	AFMM	Greece
7.	Subtract	Greece
8.	Archival Service	Serbia
9.	Simple Storage Service	Greece
10.	Institutional Repository in Medical Sciences – Nicolae Testemitanu SUMP	Moldova

Table 1: The services participating in the Life Sciences use case

The organizers of the test-cases, Prof. Zoe Cournia, Dr. Chrysovalantis Constantinou and Dr. Andreas Athenodorou along with the four evaluators conducted team meetings to plan and reevaluate the steps for each specific test case as the use case was running. As explained before, the evaluators are people with scientific orientation in the life sciences but with completely different backgrounds and research experience. Namely, the team consists of a medical doctor, a chemist and two bioinformaticians and their research experience ranges from being a master student to a PhD student as well as a Post-doc. All the details of the evaluating team can be found on Table 2.

Name	email	Background	Affiliation
Aspasia Vozi	aspav@protonmail.com	Medical Degree, NKUA	master's student, NKUA, BRFAA
Michail Papadourakis	mpapadour@bioacademy.gr	PhD in Chemistry, University of Edinburgh	Postdoc, BRFAA
Dimitris Papakonstantinou	dimitrispapak@protonmail.com	MCs Bioinformatics	Research Assistant, BRFAA
Alexios Chatzigoulas	chatzig@di.uoa.gr	PhD candidate in Bioinformatics, Department of Informatics and Telecommunications	Doctoral student, BRFAA, NKUA

Table 2: Use Case participants, contact details, background information and affiliations

The initial goal was the implementation of four use cases as described in detail in the aforementioned deliverable, one by each researcher. According to the first team meeting, it was proposed that each researcher would have to use all the services in a pipeline, using the results obtained from one service as input to another. In the end of the workflow, the users of the services were to complete a survey which can be found on the following link [https://ni4os-demonstrations.limequery.com/542299?lang=en] as well as provide the organizers with a detailed report on the use case.

The first step was to acquire access to all the services and confirm that the services were functional. Due to the fact that not all the aforementioned services were fully-on-boarded on NI4OS-Europe (at the time of the testing) and thus, the initial plan would not have been possible, it was decided that the testing of the services in a pipeline was not feasible. Thus, each service was independently tested and evaluated. In Table 3, the status of each service on the NI4OS-Europe platform (fully on-boarded, candidate, not on-boarded), as well as the type of the service (Thematic, Generic, Repository), is listed.

Service	NI4OS Status			Type of Service
	Fully on-boarded	Candidate	Not on-boarded	
Melgene		✓		Thematic
Bioconnect			✓	Thematic
PARADOX IV Cluster	✓			Generic
ChemBioServer	✓			Thematic
AFMM			✓	Thematic
Subtract			✓	Thematic
FEPrepare	✓			Thematic
Archival Service		✓		Generic
Simple Storage Service		✓		Generic
Institutional Repository in Medical Science-Nicolae Testemitanu SUMP		✓		Repository
*FINKI Cloud	✓			Generic
*NI4OS-Europe Repository Service	✓			Repository
*Data Analysis Service - PARADOX Hadood	✓			Generic

Table 3: Type of the tested services and their status with respect to NI4OS-Europe. Services listed with an asterisk (*) were added to replace the respective candidate services

Additionally, during the second team meeting it was proposed that for each service, the following properties should be evaluated:

- Accessibility: How easy is it to access the service?
- Functionality of the service:
  - User interface experience and documentation.
  - Service credibility, data sources and methods.
- Technical problems encountered
- Overall experience.

It was also agreed that recommendations regarding changes and improvements of the services should also be reported.

As shown in Table 3 there were four services which were not on-boarded on NI4OS-Europe at the time of the testing. However, MelGene which is now a candidate thematic service in NI4OS-Europe service management system was freely accessible via its website [http://bioinformatics.cing.ac.cy/MelGene/Database.php] thus it was possible to test its functionality. Bioconnect was accessible via a VPN connection to the Cyprus Institute's network, and access was provided to the evaluators by the organizers. The plan is to have Bio-connect on-boarded by the end of the project. Specifically, the team consisted of the members listed in Table 2, was provided with temporary CYI email addresses in order to be able to login to the Cyprus Institute's network via VPN. Finally, AFMM and Subtract were not tested at all since they are under functional updates.

As far as the three candidate services ("Archival Service", "Simple Storage Service" and "Institutional Repository in Medical Sciences – Nicolae Testemitanu SUMP") were concerned, it was also not possible to test them, since none of them was on-boarded in the NI4OS-Europe service catalogue. In order to be able to complete the use case three different repository services were chosen as replacements of the originally proposed repositories, in order for the use case to be completed.

More specifically, instead of the three aforementioned repositories, the repositories annotated with star (*) in Table 3, which are fully on-boarded on NI4OS, were used for storage purposes. Furthermore, it was proposed that the final step, which was to complete the survey [lhttps://ni4os-demonstrations.limequery.com/542299?lang=en] was obsolete, since the users' reports for the services covered all the topics addressed by the survey.

Finally, **eight reports** were prepared by each researcher. The reports addressed the following issues:

1. Accessibility

   Regarding the fully on-boarded services, for those services for which access could not established through AAI, all requests for access were completed using the NI4OS-Europe Helpdesk, in order to simulate a real-life scenario.

2. Functionalities

   As far as the functionality of the services is concerned, the tutorial of a service, if provided, was followed by each researcher in order to test whether one can easily use the service without prior experience with it. Otherwise, if a tutorial was not available (common for non onborarder services), a simpler use case was orchestrated to test the various functions of the service, using toy data.

   For example, in the case of Bioconnect, a team meeting with the developer of the service Dr. Charalambos Chrysostomou was organized by the task leader, in order for the team listed in Table 1 to be given instructions on how to use the service since a tutorial was not available. Moreover, Dr. Charalambos Chrysostomou provided the evaluators with two test files that were used for the review of the service.

3. Technical Problems

   All technical problems discovered throughout the testing phase of each service were issued to the providers via the NI4OS-Europe Helpdesk, as well as reported in the respective service report.

4. Recommendations

   The overall experience of using each service along with recommendations for improvements were mentioned in the final reports. Each researcher, according to her/his expertise, was able to propose specific changes to the services that would enable their improvement.

Below we give a brief description of each service tested in this use case, along with a description of what was tested.

MelGene: is a database with a publicly available collection of published genetic association studies performed on Cutaneous Melanoma (CM). It consists of two main services. The first one aims to provide meta-analysis studies conducted by the authors, related to gene polymorphisms and the incidence of Cutaneous Melanoma (CM). The second one provides information about the cooperative role of genes using network graphs. The service provides access to published genetic association studies performed on CM, which have been, compiled on a database. According to MelGene Statistics there are currently 192 studies included in the database and 1126 reported polymorphisms in 280 different genes. Finally, the results from 79 meta-analysis studies are also available. The MelGene can be accessed by clicking on the link [http://bioinformatics.cing.ac.cy/MelGene/]; the layout of the front page is presented in Figure 2.

In order to review the service, the Life Sciences community leader assigned two genes to each researcher for testing the search engine of the database for their polymorphisms, the results from the conducted meta-analysis studies and create their network graphs.

Bioconnect: is an online tool that provides two services, a repository service (Repository) that gives the user the ability to store and share genomic data and a tool for visualizing the results of genomic experiments (Genome Browser).

Four temporary emails and credentials for accessing the Cyprus Institute's network were provided to the researchers by the organizers of the test-cases. This step was necessary, since the service is, at this stage, only available via a VPN connection to the Cyprus Institute. In the future this will become fully on-boarded in NI4OS-Europe and accessing the service would be achieved through NI4OS-Europe AAI.

Moreover, due to the fact that the service is still under development, there is no manual/tutorial available and thus, navigating through the service and evaluating its functionalities was difficult. Therefore, direct contact with the developer of the service, Dr. Charalambos Chrysostomou was arranged. During such meetings an overview of the service and its functionalities was presented by Dr. Chrysostomou. Furthermore, he provided the evaluators with two test files that were used to test the service. More specifically, a test gff file, was uploaded in the repository section of the service, to further evaluate its capabilities and a compressed bed file (.bb), containing known repeating sequences of DNA from Chromosome 1, was used to test the visualization tool (Genome Browser).

Figure 2: The layout of MelGene's frontpage

ChemBioServer: is a publicly available web application for filtering and clustering chemical compounds. It allows the user to upload and browse compounds in 2D and 3D, filter compounds, create a hierarchical clustering, it allows for affinity propagation clustering by providing exemplars for each cluster, and it allows for creating structural similarity network visualization and analysis.

The service provides a well-documented and detailed tutorial and manual for its functionalities as well as many sample datasets one can use for testing. These datasets were used as use cases. The multiple services that are provided are broken into three major categories, Filtering, Clustering and Networking, and were tested as follows:

- In the Filtering section, the researchers using the aforementioned test files, reviewed the subsections of: Browse Compounds, Predefined, Queries, Combined Search, Structure, Van der Waals, Toxicity, re-ranking for Ensemble Docking and graphical representations of molecules.
- In the Clustering section, the researchers reviewed the provided services of clustering molecules according to affinity propagation and creating hierarchical clustering trees based on multiple distance metrics using the resulted files of the previous filtering section, in order to simulate a real-life scenario.
- In the Networking section, using the respective provided test files, the researchers used the tools of Structural Network Visualization and Analysis, that provide additional ways to assess the similarity of batches of molecules, while giving the capability to pick certain nodes to examine. The service can be accessed using the service catalogue of NI4OS-Europe which links to https://chembioserver.vi-seem.eu/. The layout of the ChemBioServer interface is presented in Figure 3.

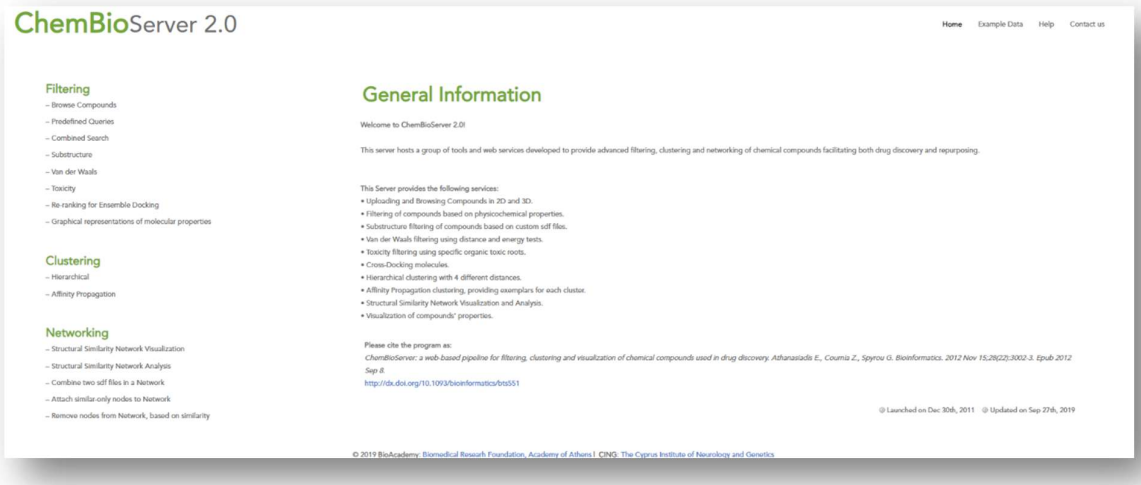

Figure 3: The ChemBioServer interface

FEPprepare: is a web server, which automates the set-up procedure for performing NAMD/FEP simulations. Users can provide the structure files obtained from the LigParGen server to the FEPrepare web-server. As a result, the algorithm provides the user with everything needed to run a simulation with NAMD using the OPLS force-field. FEPrepare provides manual and tutorial files, as well as a video demonstration, that guides the user through the service.

The aforementioned example provided by FEPrepare was used to review the service. The specific example contains, the input files, Arp2/3 protein, the reference ligand CK666 and AI003 mutant ligand, that should be used by the algorithm in order to produce the files needed to run a NAMD simulation using the OPLS force-field. There was no further testing regarding the quality of the produced files. FEPrepare can be accessed through the NI4OS-Europe service catalogue: https://catalogue.ni4os.eu/?_=/resource/onboarded which directs the user to the actual service at https://feprepare.vi-seem.eu/. A screenshot of the service layout is provided in Figure 4.

PARADOX-IV is a cluster at the Scientific Computing Laboratory of Institute of Physics Belgrade, consisting of 106 compute nodes (1696 CPU cores and 106 GPUs) interconnected by the QDR InfiniBand network. It provides various modules such as Applications, Environments, Compilers, Libraries, Parallel programming modules and other various tools.

The service was checked in terms of its accessibility and provided functionalities. Regarding the accessibility of the service the known procedure via NI4OS-Europe Helpdesk was followed. As far as the evaluation of the service functionality, some examples of the service's tutorial were followed by each researcher. More specifically, the basic operations of file transferring, job submission (MPI, OpenMP, CUDA) and compiling (OpenMP, CUDA and CUDA-MPI programs) were tested. The results of the service review, along with recommendations and comments were reported by each researcher.

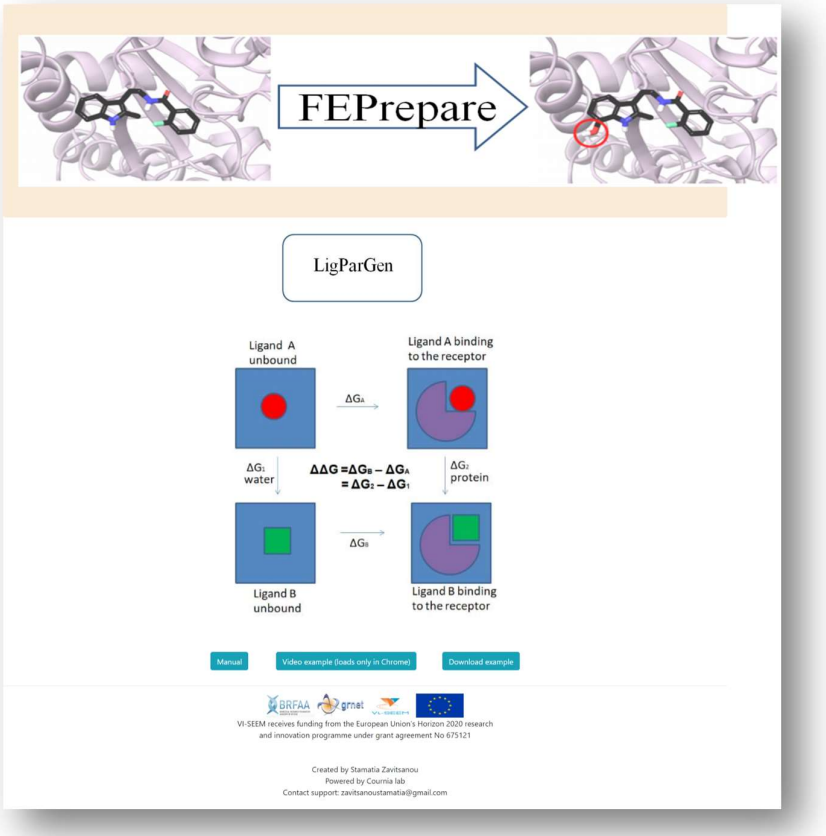

Figure 4: The layout of the FEPrepare web page

In the case of the repository/generic services FINKI Cloud, PARADOX Hadoop, NI4OS-Europe Repository Service, the services were reviewed regarding their accessibility, as well as, the ease in uploading toy data.

More specifically:

- **NI4OS-Europe Repository Service:** The creation of a NI4OS-Europe account was attempted as well as the upload of a test file in the respective TEST API, provided by the service. Moreover, accessing uploaded experimental/scientific data was evaluated. NI4OS-Europe Repository Service can be accessed through the link: https://repo.ni4os.eu/
- **PARADOX Hadoop:** The evaluation procedure was similar to that of PARADOX IV. The researchers aimed to access the service as well as test its functionalities using the examples available in the service's manual.
- **FINKI Cloud:** The evaluation procedure was limited to accessing the service, navigate through its interface as well as report on its basic functionalities using some test files.

2.2. Results

In this section we provide excerpts from the reports of the services described in the previous section which were provided to us by the researchers shown in Table 2**.**

MelGene: As mentioned above, MelGene is a database with a publicly available collection of published genetic association studies performed on Cutaneous Melanoma (CM). It consists of two main services. The first one aims to provide meta-analysis studies, conducted by the authors, related to gene polymorphisms and the incidence of CM. The second one provides information about the cooperative role of genes using network graphs.

For the purpose of the report, two genes (TYRP1 & APEX1) were used for testing the search engine and specifically its database for their polymorphisms, obtaining results from the conducted meta-analysis studies, and create their network graphs.

Using the Melgene database search engine, 13 results were obtained and 1 result for the TYRP1 and APEX1 respectively. According to the service, information for all polymorphisms found by the search can be viewed.

More specifically, by selecting a reviewed polymorphism, the results of the meta-analysis of the log odd ratios using random-effects models and the funnel plot of the log odd rations vs standard error can be seen. Further information regarding the meta-analysis study is also provided.

For example, one can visualize the location of each gene and its polymorphisms on the chromosomes, as shown in Figure 5 for one of the polymorphisms of the TYPR1 gene. There was no information available, regarding the found polymorphism of APEX1.

As far as **molecular networks** go, there are four tools that provide them, the first is SNAP which is related to the networking of the SNP polymorphisms and the GeneMania, String, as well as Gene Network Central tools that provide networks of genes. Using APEX1 and TYRP1 genes and their obtained polymorphisms, the evaluating team reviewed the functionality of the **"Molecular Networks"** service.

For the **SNP Networks**, a feature can be selected based on which the network of the selected polymorphism will be created and displayed. After selecting one feature, a network of all the available polymorphisms is created. The desirable polymorphism was selected (rs1408799 and rs3136820) from a drop-down list, and further adjustment of the parameters and visualization filters that refer to the network layout was attempted. Furthermore, the values of the parameters of the networks was obtained.

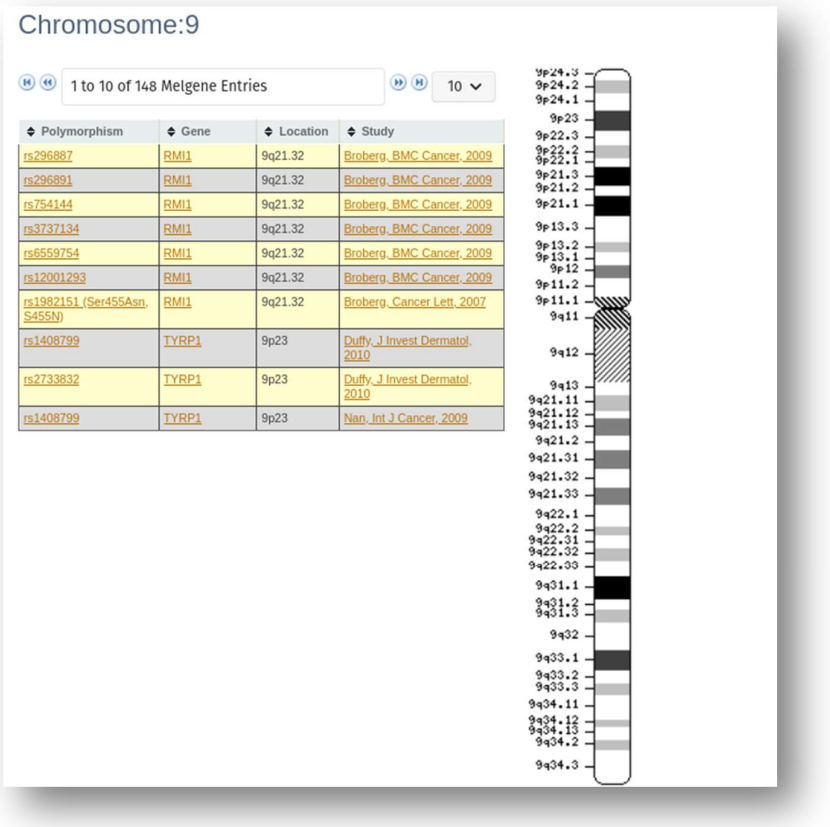

Figure 5: Locations of genes and their polymorphisms on chromosome 9

For the **Gene Network**, the same procedure was followed for the genes under examination using the three available options for the network generation. Six networks, 3 for each gene were obtained.

The service was not on-boarded on NI4OS-Europe, at the time of testing. Access was achieved after contacting directly the provider of the service. This service now appears as a candidate for on-boarding in the service catalogue of NI4OS-Europe.

As far as the **user interface and documentation** go the following conclusions were drawn:

- Regarding the search functionality and the user interface in general, the researchers who run the service concluded that both should undergo an upgrade, before on-boarded to NI4OS-Europe.
- It is a major drawback that there is currently no tutorial to provide the user with information about the search procedure as well as the interpretation of the results for both meta-analysis and molecular networks.

As far as the **Melgene database Search Engine** go the following conclusions were drawn:

- There was a great number of polymorphisms that contain no information.

- Most of the related bibliography with respect to each polymorphism is rather old, since there were no papers before 2013.
- The link to the related publication is a very helpful feature of the service.

As far as the **meta-analysis** results go the general conclusions were that:

- The forest plots provided by the service, are not well documented and furthermore they lack in quality, which makes them unreadable.
- The same stands for the funnel plots as well.
- As far as the meta-analysis details, apart from the specific studies that were included in the studies further information is not provided.

As far as **molecular networks** go the general conclusions were the following:

- The lack of sufficient documentation makes it rather difficult for the user to adjust the features and the parameters for the creation of the networks, as well as the interpretation of the results.
- It is easier to select genes in order to create their networks, since they are stored alphabetically. However, it is rather difficult to select the desirable polymorphisms, because of the lack of a prompt searching box, the large number of polymorphisms and the fact that they only differ in the final digits of their name.
- It can be very helpful to compare simultaneously the results from the different services in the ”Gene Network” section. However, this option is not provided.

The **overall conclusion** is that in general, MelGene could be a useful tool for researchers, since it provides a useful synopsis and meta-analysis of peer-reviewed genetic association studies performed on CM. Although it is functional, enrichment of the database is required, since its data seems to not be updated since 2013. Moreover, a user manual that explains the various parameters needed for the interpretation of the results would be very handy. A few more changes regarding the presentation of the search results as well as the meta-analysis data would also be useful. The same goes for the Molecular Networks section of the service, where the interpretation of the results is very difficult for a user that has no experience with the field. Another drawback is that the user cannot compare simultaneously the results of this section.

Bioconnect: As mentioned above, Bioconnect provides two services, a repository service (Repository) that gives the user the ability to store and share genomic data and a tool for visualizing the results of genomic experiments (Genome Browser).

The Bioconnect Repository service provides the user the ability to not only store online genomic data but also share them with other users worldwide. Specifically, the user is able to share a link that can be used from another user. In this way, every owner of the link can visualize results of genomic experiments, without owing the raw data of the experiment.

The Bioconnect Genome Browser service provides the user with the ability to use locally stored, as well as online files in order to visualize genomic data. Moreover, the user can simultaneously visualize data from different sources, with respect to known annotated genomes.

Two test files (Homo sapiens.GRCh37.71 gene1.gff and repeats.bb) were used, in order to test the two services provided by Bioconnect.

In order to use the Bioconnect Repository service, service accounts were created using an email account. The service consists of three functionalities, **add**, **view**, and **edit**.

Using the Add operation, the file Homo sapiens.GRCh37.71 gene1.gff was uploaded, where one can define the following parameters: name, description and passphrase. The evaluation team choose the coordinate system of the data, whether it is going to be private or public as well as the gene representation.

By clicking add one can see, that the file is uploaded along with its related information, and a DAS link is generated, that can be shared. The owner of the link can add it in the Genome Browser and visualize the results. The View and Edit operations, can be used by the user to view uploaded files and edit them, respectively. The Edit operation is useful for downloading the data, renaming or deleting the file, changing its privileges from public to private etc.

As far as the **genome browser** goes, in order to use the service, one has to create a different account using an email. Genome Browser provides the user with the following options for adding new data files and visualize genomic data:

- The user can choose in the Defaults section among the known annotated human genomes, provided from GENCODE (versions 19 and 37) as well as Genome.
- The user can add custom DAS data, using the link of external DAS sources, or specifically the DAS link generated by the Bioconnect Repository service.
- The user can also add indexed binary data from a web server that supports CORS, as well as locally stored files. This option is currently supported for bigwig, bigbed and indexed BAM formats.
- Finally, the user can select and visualize specific known DNA sequencing peaks of the human genome.

Moreover, the user can configure parameters that have to do with the interface such as the display and highlight of the current genome location, reverse trackpad scrolling option and others.

Subsequently, the local file repeats.bb was uploaded. This is a compressed bed file containing known repeating sequences of DNA from Chromosome 1. The file is stored locally and visualized using the different coordinates available (GRCH38/hg38 and GRCH37/hg19) as shown in Figure 6.

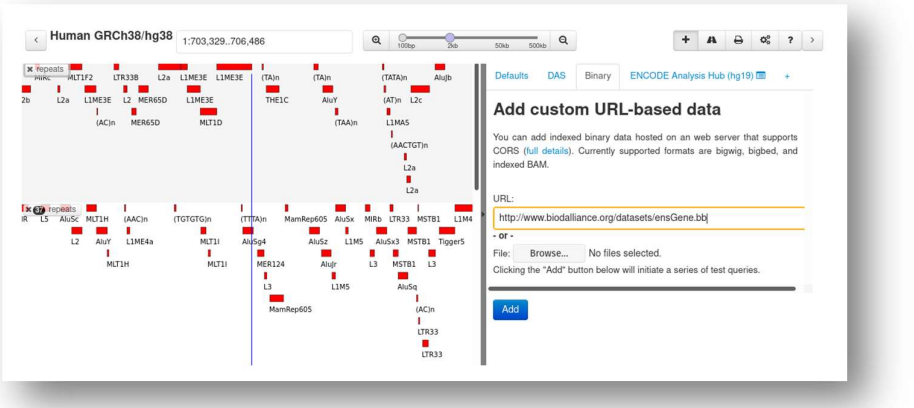

Figure 6: Test file visualization using different annotation within Bioconnect

The evaluators then used the link generated from the upload of their file in the Bioconnect Repository service that contains the data of the Homo sapiens.GRCh37.71 gene1.gff file, in order to visualize it. The evaluators used the GRCh38/hg38 as coordinate system. Finally, they used the GENCODE as the default genome, and visualized the results as shown in Figure 7.

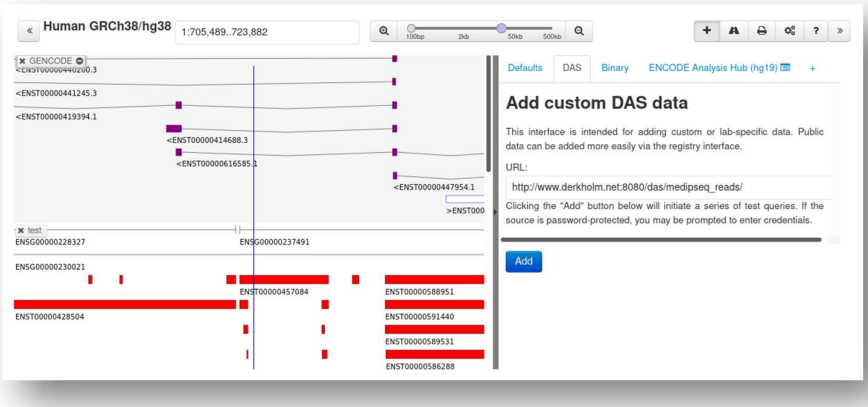

Figure 7: Visualization of DAS link data in Bioconnect thematic service

As far as the **user interface and documentation** goes, currently there is no documentation file and/or tutorial available for either of these services. There is only a help page for keyboard shortcuts, which provides information regarding the navigation in the browser and data. This makes it difficult for the user to navigate through the services and use them.

Another inconvenience with respect to the user interface lies on the fact that the user has to create two separate accounts, one for the Repository and one for the Genome Browser

Services. This is expected to change in the future, since it undermines one of the basic purposes of the services.

With regards to the repository, one of the limitations that need to be addressed, is that the service only supports the upload of GFF or GFF3 formatted files.

As far as the genome browser goes, currently the user can select as default only human genomes. It would be useful if the user was able to select rat, mouse and other genomes as defaults as well, since a great amount of genomic experiments are conducted using these organisms.

As a visualization tool, the Bioconnect Genome Browser does not provide any additional advantage in comparison to the available online visualization tools. However, this service provides the ability to share the results of genomic experiments without sharing the raw data, which is a very useful tool.

The **overall conclusion** is that in general Bioconnect is a very useful tool, because it provides researchers with the opportunity to share and visualize the results from their genomic experiments, without sharing their raw data. This specific trait is valuable and almost none of the services currently available offers a similar service. However, a big drawback, is that there is a lack of manual files and tutorials that could help an inexperienced user navigate throughout the Genome Web browser and the Repository service. Another inconvenience with respect to the user interface lies in the fact that the user has to create two separate accounts, one for the Repository and one for the Genome Browser Services. In addition, there is no physical connection between the two services. It does not offer any particular advantage in terms of visual inspection of the genomic data compared to other visualization tools, for instance in comparison to Genome Data Viewer, however it provides the ability of sharing genomic data throughout the scientific community. Taking everything into consideration and with the proper documentation and improvement in the webserver, Bioconnect could be a handy tool for the NI4OS-Europe project.

ChemBioServer: ChemBioServer is a publicly available web application for filtering and clustering chemical compounds. It provides tools for uploading and browsing, filtering of compounds, hierarchical clustering, affinity propagation clustering providing exemplars for each cluster, and structural similarity network visualization and analysis. Unlike the two previous services presented above this is fully on-boarded.

In order to review the three services included within ChemBioServer, the evaluating team used the test samples provided by the service tutorial.

To test the **filtering services**, the molecules included in the test files, can be visualized and filtered according to their chemical properties. Specifically, the team worked on the following:

- To **browse compounds**, they used the test.sdf file containing 10 compounds, which can be visualized.
- For **predefined queries**, they could perform mass filtering with respect to predefined criteria that are commonly applied to filtering compound for a virtual screening procedure.
- For **combined search**, they could adjust the various filtering parameters for mass filtering.

- For **substructure**, they could examine whether two files contain any number of same compounds.
- In the **Van de Waals** section, they could filter the molecules with respect to the existence of bad van der Waals interactions.
- In the **toxicity** section, they could filter the molecules with respect to the existence of toxic moieties.
- In the **Re-ranking for Ensemble Docking** section, they could find those molecules that have a low energy of binding for a target protein whilst having a high energy of binding for others.
- In the **Graphical representations of molecular properties section,** the following graphical representations are presented: PCA2 vs PCA1 Principal Component Analysis (PCA) first component (PCA1) against the second component (PCA2), based on the tanimoto coefficient (distance). PSA vs logP Logarithm of the calculated Partition coefficient (logP) against the Polar Surface Area (PSA). PSA vs MW Molecular Weight (MW) against the Polar Surface Area (PSA). logP vs MWMolecular Weight (MW) against Logarithm of the calculated Partition coefficient (logP). Plots were created by using the Raphael JavaScript library.

To test the **Clustering** service, the team worked with the following features of the service:

- In the **Hierarchical** section, they uploaded a file (or a txt file with custom binary fingerprints) and clustered compounds with similar characteristics (tanimoto, dice, cosine, simple matching coefficient).
- In the **Affinity Propagation** section, they uploaded a file (or a txt file with custom binary finger prints) and grouped compounds with similar characteristics (tanimoto, dice, cosine, simple matching coefficient). Most representative compounds (Ex emplars) for each cluster are provided for further investigation.

To test the **networking** service, the evaluating team worked with the following:

- In the **Structural Similarity Network Visualization** section, they uploaded an sdf file and after choosing a similarity metric and a value cutoff threshold for the edges (similarity values), they visualized the generated network.
- In the **Structural Similarity Network Analysis** section, they uploaded an sdf file and after choosing a similarity metric and a value cutoff threshold for the edges (similarity values), they ran the previously generated network analysis.
- To **Combine two sdf files in a Network** they used one sdf file, that acts as the reference network, and a second sdf file that is composed of the compounds, that they wanted to query against the reference set in order to find the closest structural neighbors.
- The **Attach similar-only nodes to Network** section one can infer relations between a main Network of compounds and a secondary list of compounds.
- Finally, they could use the **Remove nodes from Network, based on similarity** section to remove nodes from a main Network of compounds based on a secondary list of compounds.

As far as the **User Interface and Documentation** go:

- ChemBioServer is fully on-boarded on NI4OS-Europe and information and material about the service can be found on the NI4OS-Europe webpage.
- The service is well documented. There is a step by step tutorial provided for all the subsections of the service.
- Apart from the tutorial, there are various sample data provided by the developers of the service in order to help the user interpret the results obtained as well as the functionality of the service.
- Although the Custom Pipeline section exists in the tutorial, the team members were not able to find it in the main menu of the service. Immediate action should be taken by the NI4OS-Europe in order to resolve this subtlety.

As far as the **filtering** part of the service goes:

- They used some of the test files, provided by the ChemBioServer tutorial, to review the service. More specifically, the 2D representation of the compounds was viewed in the Browse Compounds section, using the test.sdf file that contains 10 compounds. Then, the team used the test4.sdf file containing 557 molecules, to review the various filtering tools. An example of visualization of compounds, using the Browse Compounds section is provided in Figure 8.

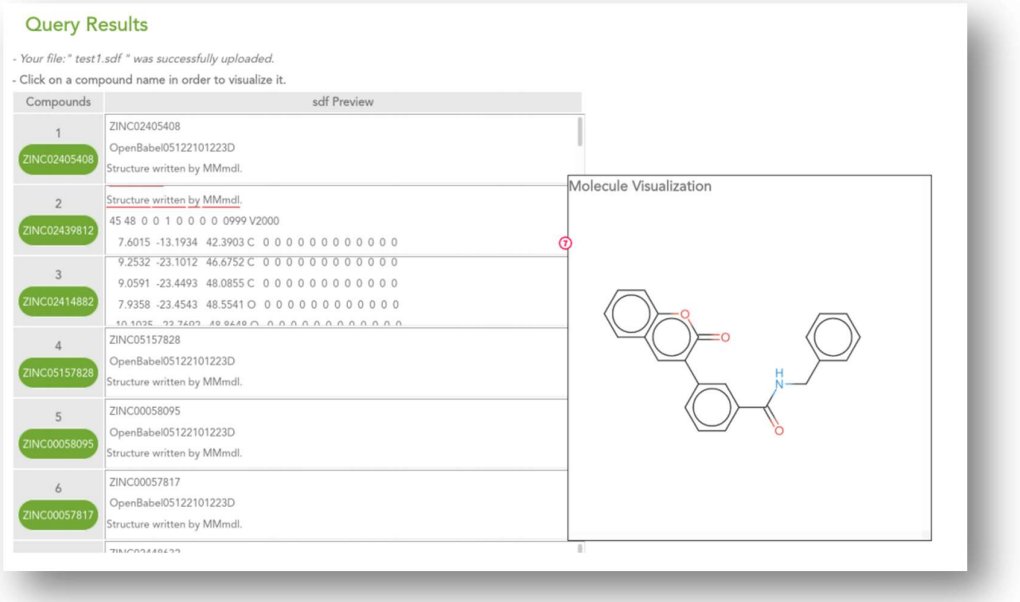

Figure 8: Visualization of compounds, using the Browse Compounds section

- First, the compounds were filtered in the Predefined Queries section according to the Lipinski's rules of 5, the Veber's rules, and the Ghose filters resulting in 313 compounds successfully passing the filtering.
- In the Combined Search section, the user also has the ability to filter the compounds with custom criteria values for various physicochemical properties such as molecular weight, charge, number of atoms, number of aromatic rings, etc.

Testing for the number of Carbon atoms, 10 successfully filtered the molecules, resulting in 401 molecules.

- In the Substructure Filtering section, taking as input the output of the Predefined Queries section, the team checked for five common unwanted fragments with undesired functional properties from an sdf file provided by ChemBioServer common unwanted.sdf, resulting in 296 compounds. These compounds were further filtered for bad Van der Waals interactions and toxic moieties, resulting in 197 and finally 192 compounds. The remaining compounds were visualized in the Graphical representations of molecular properties section, to further evaluate their chemical and structural properties.
- All the results can be downloaded in sdf and txt format, and in some cases can be visualized on the server as well.

As far as the **clustering** service goes:

- Using the compounds that have passed all the previous tests and in order to further reduce them, the team used the Hierarchical Clustering section of ChemBioServer. They adjusted the clustering parameters as follows: Distance Selection: Soergel (Tanimoto Coefficient), Clustering Linkage Selection: complete, Clustering Threshold: 0.94, and obtained five clusters.
- In the Affinity Propagation section, using as input the results extracted from the previous step for each cluster, and adjusting the distance parameter to Soergel (Tanimoto Coefficient), the team obtained the exemplars of each cluster: Cluster 1: 17 exemplars, Cluster 2: 24 exemplars, Cluster 3: 3 exemplars, Cluster 4: 3 exemplars, Cluster 5: 2 exemplars.

For the **networking** service they did the following:

- In the Structural Similarity Network Visualization and Structural Similarity Network Analysis sections of ChemBioServer, they used the sdf-refernce.sdf file provided by the manual to visualize and run the network analysis. Both of these procedures were successfully run.
- In the Combine two sdf files in a Network section, the team members used the sdf-refernce.sdf and sdf-test.sdf files as reference and test respectively. The procedure was successfully run as well. An example of the Networking section results, regarding the combined analysis of two file structures is demonstrated in Figure 9.
- The same holds for the last two sections of the Networking service.

The overall **conclusion** was that in general, ChemBioServer is a very useful and practical tool for filtering, clustering and analyzing compounds resulting from virtual screening. It is well documented with many example test sets and tutorials available. Finally, it accelerates a time-consuming procedure, by automating the filtering of the compounds, facilitating drug discovery. This in full alignment with the decision of NI4OS-Europe to fully on-board this very useful service.

FEPprepare: As mentioned above, FEP prepare is a webserver, which automates the set-up procedure for performing NAMD/FEP simulations. Users can provide the structure files obtained from the LigParGen server to the FEPrepare web-server. As a result, the algorithm provides the user with everything needed to run a simulation with NAMD using the OPLS force-field.

FEPrepare provides manual and tutorial files, as well as a video demonstration, that guides the user through the service. The example provided by FEPrepare was tested, which contains the Arp2/3 protein, the reference ligan CK666 and the AI003 mutant ligand.

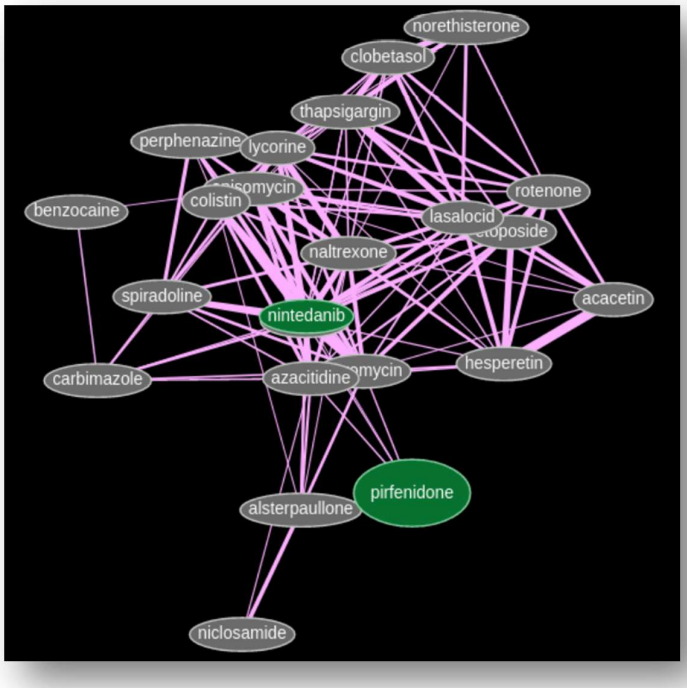

Figure 9: Example of the Networking section results, regarding the combined analysis of two file structures

In order to use FEPrepare, there are initially the coordinate and the stream files for both ligands (reference and mutant) given as pdb and str file as well as the pdb file of the prepared protein. The files required as inputs, the pdb of the protein and the pdb, rtf, and prm files of the reference and mutant ligands are all provided by the tutorial example.

The evaluation team fed the files into the service and waited a few minutes for the process to be completed. The team downloaded the file.zip file that is generated and obtain its content:

- two subfolders, complex and solvent,
- six renamed files, pdb, rtf, and prm, for both ligands,
- two hybrid files named as ligand.pdb and ligand.rtf,
- the renamed prm file after the creation of the two hybrid files,
- the updated.prm file,
- the complex.pdb file,
- the fep.tcl script, that is needed in order NAMD to run the simula- tion,
- a file with the OPLS-AA parameters of proteins, and
- a file with the OPLS-AA topology of proteins.

The complex folder contained the pdb files for each of the protein's chains and the ligand, the ionized files ionized.pdb, ionized.fep, ionized.psf and the updated ionized files ionized new.pdb and ionized new.fep, two text files, min-max center and vmd log, the files used to run VMD, psfgen, and VMD prepare complex after gui autopsf, and the files that VMD gives as an output, complex wb.log, complex wb.pdb, complex wb.psf, psf-complex.psf, psf-complex.pdb. The solvent folder contained similar outputs, but for the solvent system.

The general comments of the evaluating team are the following:

- The service is fully on-boarded on NI4OS and all the information and material regarding the service are available on the NI4OS-Europe page.
- FEPrepare is functional at the moment, since all the necessary, according to the service manual, files were produced. However, there was no further testing regarding the quality of the produced files.
- It is very useful that such a service is available, since the preparation of the files is a time-consuming process. It is very helpful for such processes to be automated.

The general conclusion is that FEPrepare is a very useful and practical tool, since it helps automating the setup of all the necessary input files for relative FEP/MD calculations using the dual topology approach of NAMD and the OPLS-AA force-field parameters. Overall, the service is functional and well-documented. A manual and a tutorial along with a video demonstration is provided, which makes it very easy and straightforward to use the service. It is very handy that such a service is available via NI4OS, since the preparation of the files needed to run a NAMD simulation is a time-consuming process. Once more, this is in full alignment with the decision of NI4OS-Europe to fully on-board this very useful service.

PARADOX-IV: PARADOX-IV is a cluster at the Scientific Computing Laboratory of Institute of Physics Belgrade, consisting of 106 compute nodes (1696 CPU cores and 106 GPUs) interconnected by the QDR InfiniBand network.

In order to access the PARADOX-IV cluster, which is fully on-boarded on NI4OS-Europe, the evaluators used the NI4OS-Europe helpdesk. After a document of agreement was signed, access to the service was granted. Some problems regarding the ssh connection to the service were resolved with the use of NI4OS-Europe helpdesk.

To test the functionalities of the service, the toy tasks recommended by the manual that PARADOX-IV provides were used as described below.

The service provides various modules such as Applications, Environments, Compilers, Libraries, Parallel programming modules and other various tools that can be viewed with the use of the command module avail.

In the user guide of the service there are the following examples that were all executed without any problem:

- Transfer files to the cluster.
- Job Submission Tutorials:
    - Batch and sequential job submissions.
    - MPI job submission.
    - OpenMP job submission.
    - CUDA single node job submission.

- Compiling Tutorials: The loading of the required libraries and modules was tested. The compilers were only tested on toy programs.
    - Compiling OpenMP programs.
    - Compiling CUDA programs.
    - Compiling CUDA-MPI programs.
- The researchers used vim to edit the code directly from the console, however PARADOX provides other options as well. Furthermore, there is also a tutorial on how to use different IDEs (Atom & Eclipse).

After testing the service, the evaluating team reached the following conclusions regarding the accessibility, technical problems, and documentation:

- The NI4OS-Europe helpdesk response was prompt and immediate. All communication was mediated by the helpdesk of PARADOX via NI4OS-Europe with the use of tickets and every technical problem addressed, was resolved in a few hours.
- Moreover, the responses to every technical issue were complete and updated.
- A very well documented manual is provided, that was helpful and makes it easy for every user, regardless its background to use the service to all its extent.
- There were some unexpected power outages at IPB facilities. However, the providers kept us informed and resolved the problems as quickly as possible. Since, at this time, there was no running job, we cannot evaluate the response adequately.

Moreover:

- Following the examples in the tutorial, no problem was addressed throughout the whole testing procedure.

- All the modules tested, are currently functional and running if requested.

The overall **conclusion** is that, the PARADOX-IV cluster is a quite functional, useful and well-organized service. The documentation is sufficient and the accessibility is easy using the NI4OS-Europe helpdesk. This justifies the decision of NI4OS-Europe to fully on-board the PARADOX-IV in the service management system and subsequently to EOSC.

NI4OS-Europe Repository Service: The NI4OS-Europe 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).

The repository service allows NI4OS-Europe users to deposit and share data via a user-friendly web interface. It can host publications and their associated data or software. It automatically generates a Persistent Identifier for each shared item. Access to shared items can be public or limited to selected repository users.

For the purpose of this use case the team evaluated the accessibility to the service, its functionalities as well as its interface. They used the NI4OS-Europe helpdesk to contact the providers of the service and request access to it and a test file for uploading.

In order to access the service, they submitted a request to the NI4OS-Europe Helpdesk, and the response was immediate, describing the steps that should be followed. During the first step of the process, that is to login with an academic ID, there was a problem with the academic organization (NKUA). The problem was reported to the providers and their response was prompt. However, the problem was not resolved in a period of 14 days and there was no follow-up on its progress.

To that end, we requested an alternative option to access the service using the NI4OS-Europe Helpdesk, but it was suggested to wait for the problem with the academic login to be resolved. Since there was no progress for 14 days, we were advised by the task leader to login to the NI4OS service using a Gmail account and the login procedure was completed without any problems.

Regarding the uploading procedure, after the access to the service was granted, the evaluating team were, promptly and without any problem, given the appropriate access rights to upload the test file. Moreover, the service provides the user with various options regarding the description of the uploaded file, the authors of the file and other useful information. Most importantly, the evaluators could choose the license of their item/data and accept granting distribution license. The whole procedure was straightforward and no problems were addressed, the test file was uploaded and can be reviewed and downloaded using this link [https://repo.ni4os.eu/handle/21.15102/NI4OS-374].

Navigating in NRS can be done even if the user does not have the appropriate rights for uploading. Moreover, the user can access and download all the available data that are stored in NRS, as long as an account is generated.

There are five communities, namely "Climate Sciences", "Digital Cultural Heritage", "Life Sciences", "Project Community" and "TEST API", each one providing a collection of uploaded items. The evaluating team navigated through the collection of the "Life Sciences", where they could further search by author, subject, keywords and date issued. Furthermore, the team were able to download the data provided by Gihan Kamel and Matalagah Salman under the title "Sequence Prediction Algorithms vs. Infrared Spectroscopy for Proteins Secondary Structure Optimization".

The general **conclusion** is that the NI4OS-Europe Repository Service is a well-organized and potentially useful tool for sharing and accessing experimental data. Although similar services are already available, NRS has the advantage that it can be combined with other on-boarded NI4OS services and grow a strong research community. It is well constructed and maintained and seems to follow the FAIR Data principles. Although currently, there are only four communities available in NRS, it is expected that in the future it will grow bigger. Regarding its accessibility, the possibility of academic login is of great importance to our understanding, and the technical problems regarding the authentication should be resolved. However, the user can also login using her/his Gmail, GitHub or Facebook accounts, and furthermore registration is only required for uploading data. Although browsing and accessing the data does not require registration, it is of great importance that the authentication problem of the academic logins will be resolved.

PARADOX Hadoop: The 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 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 purpose of this use case, the research team tried to access PARADOX Hadoop cluster and test its functionalities. For the latter cause, they used the service's manual, similarly to the PARADOX IV Cluster use case. This user guide provides a tutorial to MapReduce parallel computing approach and how it can be done on SCL's HADOOP cluster.

The evaluating team implemented some examples provided by this tutorial, such as testing the HDFS filesystem of the cluster and its capabilities and they ran MapReduce in Hadoop in some of the configurations available.

Initially, they followed the same procedure, requesting access via the NI4OS-Europe Helpdesk. Unfortunately, there was a problem with the service's Helpdesk, but after contacting directly with the provider of the service, the problem was resolved and access to service was granted. Since they have already acquired access to test and use PARADOX IV Cluster, they used the same credentials to access this service as well and there was no need to sign an agreement this time. No problems were addressed throughout the testing procedure.

Similar to the PARADOX-IV cluster, PARADOX Hadoop is a functional, useful and well documented service. The problem with the NI4OS-Europe Helpdesk was resolved immediately after communicating with the providers of the service and permission was granted. The documentation of the service is sufficient and all the tests ran without any further problem.

FINKI Cloud: FINKI cloud is a cloud service provided by the University Ss. Cyril and Methodius, Faculty of Computer science and engineering in Skopje. The infrastructure is based on OpenStack cloud computing platform and is hosted on 15 Huawei servers, each with 128GB RAM and 20 HT CPU cores, totaling in 300 vCPU cores and 37TB SSD and 32TB SAS storage. The connectivity to the Internet is 1Gbit through MARNET provided link to GEANT. Currently the system hosts templates for all popular Linux distributions, and Windows variations. The primary target communities are researchers and research groups form all scientific areas, with special focus on the long tail of science, due to the modest capacity of the services. It is meant to provide those researchers and groups easy access to computing environment in the cloud, with images for the operating systems and solutions most frequently used in scientific cloud environments. The infrastructure is fully integrated with the eduGAIN / NI4OS AAI, enabling seamless access to virtualized computing, network and storage resources.

For the purpose of this Use Case the evaluators tested the accessibility and functionality of the service. To perform the evaluation, the tutorial of the service was used in order to guide us for the creation and testing of a virtual machine instance.

In order to get access to the service, the team communicated with the providers of the service using the NI4OS-Europe Helpdesk. Initially, a few problems were addressed, since there are no clear steps that should be followed and the providers only replied using an automated message that directs to the manual of the service, which does not provide any further information. An authentication procedure, is required in order to request permission to the service and this can be performed through the web-based UI using keystone credentials, or a NI4OS-Europe AAI account.

After the authentication step was completed, permission to access and use the service was granted to them immediately and without any further problems.

A new virtual machine instance was successfully created following the steps described in the section 3.1.1 of the tutorial. The evaluators chose Compute → Instances and they created an instance called test-VM and the respective description, test case for NI4OS-Europe, was given. In the next step, (Source Step) they selected the operating system for the created instance from an exhaustive list of operating systems (ubuntu-server-18.04). It is also available to add any additional volume besides the default one, where the operating system is installed. In the Flavor step, they selected the compute, memory, and storage capacity of the instance. More specifically, they selected t2.nano that contains 1 VCPU. 512 MB RAM and 20 GB storage capacity. In the Networks step, they selected the external-3000 option as a virtual network for the virtual machine in order to grant their instance a public access. There is a section where more details are provided. Finally in the key pair section, they could select an SSH public key that will be added as an authorized key inside the operating system. In the end, their VM was launched and in active state and using the default remote access protocol (SSH for GNU/Linux instances), they connected to the virtual machine by using its allocated IP address and tested some basic operations such as file transferring and manipulation.

Moreover, they could overview in detail their instance, upon selecting it, including the option to perform one of the available actions, from managing the network settings and attached interfaces, working with volumes and snapshots, to controlling the execution of the instance including pausing, suspending or shutting down the instance.

A new volume attachable to their virtual machine was created by selecting the Create Volume button in the top right-hand side of the Volumes → Volumes section. In the popup window appeared, they selected the desired configuration for the new volume, type of the volume (RBD for normal performance) and size of the new volume (1 GB). Finally, they were also able to create a snapshot called "test-snapshot" without any description.

In order to further check the service, the evaluators ran a small program. In the tutorial provided, there is an extensive description of the steps that need to be followed for accessing the service, which is very useful. However, there is no information regarding the authentication step that is essential in order for the user to be able to require access. The troubleshooting section of the tutorial also seems very handy, since it describes common problems that a VM instance user may address as well as their potential solutions.

Overall, FINKI Cloud seems to be a well-organized and documented service. Service's manual provides detailed information regarding its functionalities and thus, navigating

through the service is easy for the user. However, it is suggested that supplementary information should be included in the tutorial, regarding the authentication step. Apart from that, there was no problem addressed during the procedure of creating and interacting with the virtual machine instance. In general, FINKI Cloud is a very useful tool for the scientific community.

3. Climate Sciences

3.1. Use case description

Similar to the life science community, in D6.3 a description of a use case for the Climate Science community titled: "Atmospheric composition modelling and Air Quality Forecasting" was described. A number of researchers with different backgrounds in the field of Climate Sciences (see Table 4), coming from different partnering countries in the NI4OS-Europe consortium provided reports for us, which contained evaluations of their experiences using services which are on-boarded to the NI4OS-Europe platform, which are relevant to the **Climate Science community** and their research projects.

The organizers of the test-cases, Scientific Community Leader Professor Theodoros Christoudias, the task leader Dr. Chrysovalantis Constantinou and the work package leader Dr. Andreas Athenodorou along with the three evaluators conducted team meetings to plan the steps for each specific test case, and followed through as the use case was running. Two questionnaires were conducted, targeting the researchers involved in the use cases. Additional reports with more details were provided by the users in the Table 4: The names, affiliation, job title, scientific research are as well as the user community they represent of the scientist participated in the Climate Scientific Community use-cases Below we provided the survey used for contacting the scientific community of Climate Sciences within the NI4OS-Europe region.

Survey: The NI4OS-Europe climate community survey's purpose was to collect information about the research areas in which the NI4OS-Europe partners are involved with, and establish use cases based on the offered thematic, repository, and HPC/Cloud services. The survey included 12 questions which took only a few minutes to complete. The following information/questions were collected from the researchers:

- Email Address
- A brief description of their research (a couple of lines)
- The general domain in which their research belongs to with five available options (climate modelling, air quality, numerical weather prediction, early warning systems, or other followed by a brief catch phrase)
- The NI4OS-Europe repositories that are relevant and/or useful to their research with the options being DREAM (dust modelling), HESPERIA (solar weather forecasting tools), Airpolution prediction (airpolution prediction levels for the next 4 days), WRF-ARW (numerical weather prediction), Live Access Server (visualization of geo-referencing data), OMApp (automatic image geo-referencing), IoT Cloud Platform (data collection, processing and visualization).
- The NI4OS-Europe generic/core services that are relevant to their work with the options being Data Discovery Service, Simple Storage Service (file sharing), Archival service, NI4OS-Europe helpdesk, NI4OS-Europe training and webinar portal, Data Analysis Service.
- The researchers were then asked whether they used the following HPC/Cloud services (the only options were YES/NO): CYI Cloud facility, AVITOHOL cloud, Cyclone Supercomputer (Cyprus), AVITOHOL, FINKI cloud, PARADOX-IV cluster, ARIS.

- The researchers were then asked whether they are involved with Education, Research, Outreach Activities, Training Events, Policy making/advice, and Data Analysis with the options being YES or NO.
- The next question asked whether the researchers are satisfied with the NI4OS-Europe **Service Discoverability, Documentation, and Helpdesk** with five available options ranging from **Very Satisfied** to **not at all satisfied.**
- The next question asked whether the researcher's work requires any of the following: **HPC Services, Data Storage, Virtual Machines, Thematic Services Training, User Support,** and **Cloud Services.**
- Finally, the researchers were asked to provide a **list of publications which were enabled by the NI4OS-Europe project,** a list of any other **results such as datasets, talks, dissemination activities, or outreach events which were enabled by the NI4OS-Europe project,** and **a list of recommendations and/or suggestions for the improvement of available services.**

As outlined in Deliverable 6.3, the Climate and Earth system science community focuses on regional climate modelling and weather forecasting, complemented by atmospheric composition (air quality) modelling that includes the impacts on human health, and well-being. Climate modelling and weather forecasting have very strong computational demands as Figure 10 suggests. In particular, the integration of various computational resources with data repositories jointly support research and operational activity, and maximize the productivity in this particular field of research.

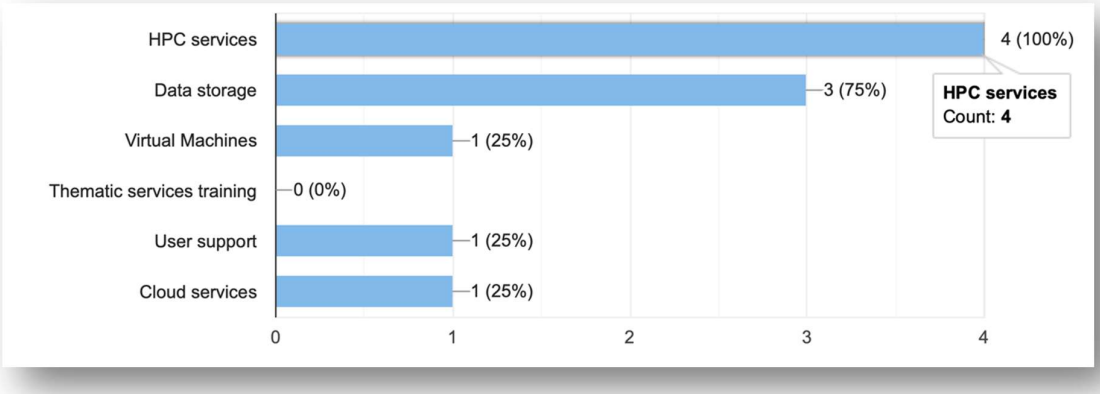

Figure 10: Climate User Community Needs for Use case assessment through the targeted survey

Air quality in particular is a critical environmental issue in many countries covered by NI4OS-Europe because of industrial and mining emissions, food processing, as well as vehicle emissions. High concentrations of these pollutants near the Earth's surface can cause serious health problems. The parameterizations of air quality models are pretty challenging to simulate aerosol properties accurately, such as emission sources, meteorological parameterizations, or aerosol chemistry.

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.

Name	Affiliation	Job Title	Scientific Research Area	User Community represented
George Mikuchadze (Georgia)	National Environmental Agency of Georgia (GNEA)	Main Specialist in Hydrometeorological Forecast Models Adaptation and Implementation Administration of Hydrometerological Depart ment.	High Performance Computing, Numerical Weather and Climate Prediction.	Weather and Climate Science
Hrachya Astsatryan (Armenia)	INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF ARMENIA	Head of Centre for Scientific Computing, Institute for Informatics and Automated Problems of NAS RA	High Performance Computing	Weather and Climate Science
Boro Jakimovski (N.Macedonia)	UKIM	Vice-dean for science at Faculty of Computer Science and Engineering	Meteorology and atmospheric composition	Air Quality

Table 4: The names, affiliation, job title, scientific research are as well as the user community they represent of the scientist participated in the Climate Scientific Community use-cases

According to the climate use case scenario as this was presented in Deliverable D6.3, a scientific team will 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 produce significant indicators on the possible future increased pollution levels. 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. A pipeline visualizing the above plan is provided in Figure 11.

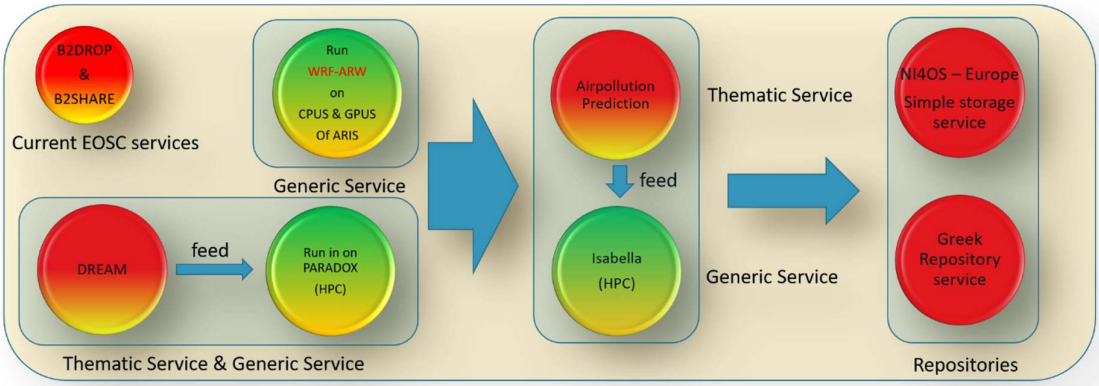

Figure 11: A pictorial representation of a climate pipeline designed so that it makes use mature thematic services, generic services, repositories as well as other EOSC services

Hence, the following services presented in Table 5 were proposed to be evaluated with respect to their **accessibility**, **quality** and **efficiency**:

Number	Name of Service	Providing Country
1.	Airpollution Prediction	North Macedonia
2.	WRF-ARW	-
3.	DREAM	Serbia
4.	ARIS	Greece
5.	PARADOX-IV cluster	Serbia
6.	Isabella – HPC	Croatia
7.	Simple Storage Service	Greece
8.	Greek Repository Service	Greece
9.	B2STAGE	EUDAT
10.	B2DROP	EUDAT

Table 5: The services participating in the Life Sciences use case

The status of the above services in NI4OS-Europe is shown in Table 6 below:

Service	NI4OS Status			Type of Service
	Fully on-boarded	Candidate	Not on-boarded	
DREAM		✓		Thematic
ARIS		✓		Generic
PARADOX IV Cluster	✓			Generic
Isabella			✓	Generic
Greek Repository Service			✓	Generic
WRF-ARW			✓	Thematic
Simple Storage Service		✓		Generic
B2STAGE	EOSC/EUDAT service			Repository
B2DROP	EOSC/EUDAT service			Repository

Table 6: All the services participating in the Climate Scientific Community use case, their status regarding the on-boarding in NI4OS-Europe as well as their type

This was the main target of service on-boarding and the services offered were evaluated by select user communities. At the end of the use case, all targeted services were tested, capturing all aspects of the scientific work enabled and supported by NI4OS-Europe (Figure 12). All services were used in at least one instance (Figure 13), and in most cases in two or more.

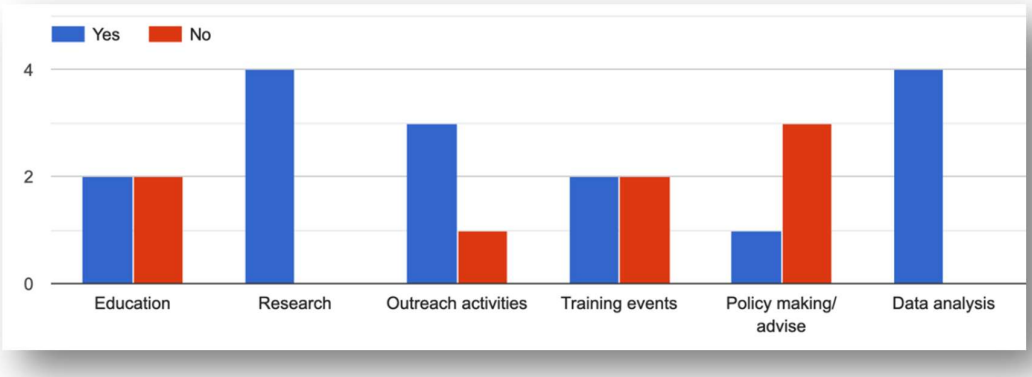

Figure 12: Scientific life cycle activities relevant of the Climate Science Community that were captured by the use cases

Overall, the outcomes were positive, with all use cases reporting satisfaction in terms of service discoverability, documentation and service manuals, and the support provided by the NI4OS-Europe Help Desk and overall support system (Figure 15).

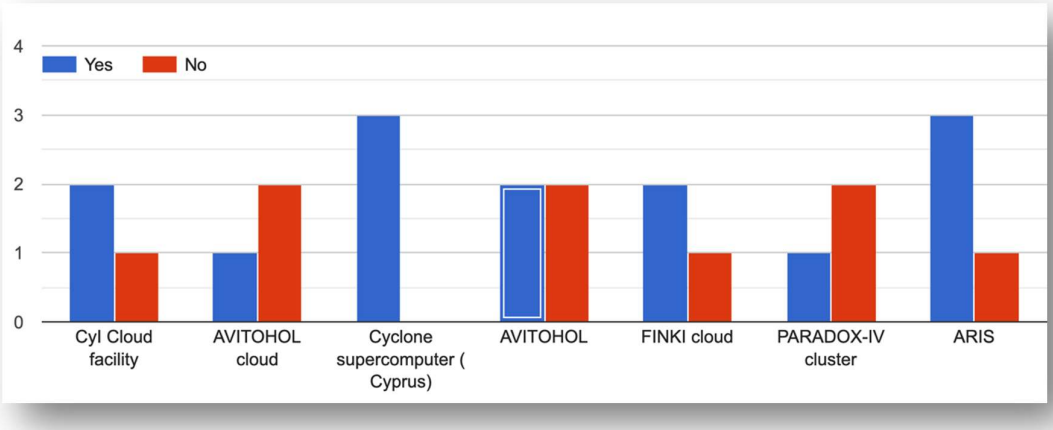

Figure 13: Use case service coverage. Blue indicated use of service individual user

3.2. Individual Use Cases

Air-pollution prediction is a web-application that enables both publicly available datasets as well as services for execution of air pollution predictions. The web application currently runs using a predefined dataset that is dynamically adapted for the region of North Macedonia. Further development is aimed to allow users to browse through the already processed simulation data, as well as uploading mobile emissions data and execution of simulations based on this input data. The input data is provided based on a specified format, as well as information about the simulation region and its resolution. Due to the sensitivity of configuration parameters of the **WRF-ARW** model, the system allows several predefined presets, which will be used for the simulations. The simulation process is very CPU intensive and produces the results in several hours up to few days. Upon competition, the output files are post-processed, and the data is stored in a database. Simulation output files are kept separately and are kept for up to a month due to the large volume of data. The web application of Air-Pollution prediction was originally located on the resources of **FINKI Cluster** and post-processing is executed on another VM on the **FINKI Cluster**. NI4OS-Europe allowed for the transformation of the system into a service. Previously, the system was only available to the researchers that developed the system. Now, it is transformed into a service, that will provide datasets and simulation services to many researchers and the general public (Figure 14). The simulation of the WRF Chem model was executed using the **ISABELLA** HPC cluster. Porting to NI4OS-Europe services, allows the current execution workflow to be further automated for different user entries.

The **FINKI Cloud** service was used as a flexible platform to test the scalability of the models, before the actual runs performed on the HPC systems such as **ISABELA**. Access

to **FINKI Cloud** is enabled through NI4OS-Europe AAI system, allowing very fast deployment of IaaS resources on demand. This is a very usefull service to test and verify the scalability of the codes before allocating time on the real HPC clusters. Once the scalability was reached and parameters tuned, the model was moved to the ISABELA cluster, for more complex computations. **ISABELA cluster** is very well documented (https://wiki.srce.hr/pages/viewpage.action?pageId=8487673), enabling the users' seamless access to this powerful HPC cluster.

Once the computation was done on the ISABELA cluster, the output was moved to the **B2DROP** services, using simple webdav client easywebdav. Using the **B2DROP** enables easy access to the results from various clients, as well as sharing the resulting datasets within larger teams.

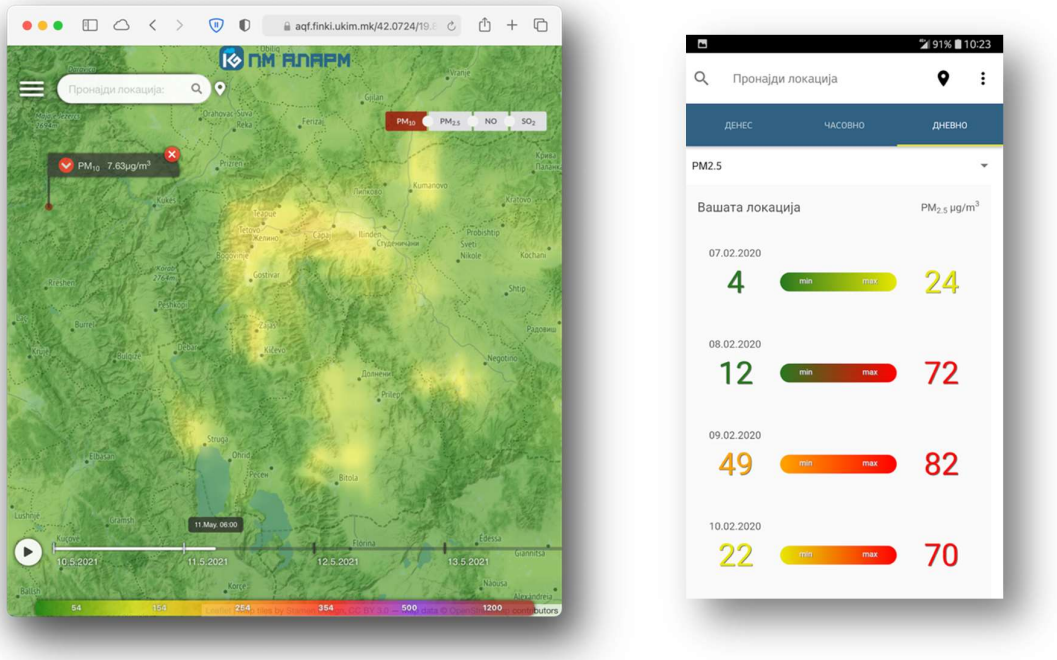

Figure 14: Examples of the Air Quality Prediction service running for North Macedonia as a web application (left) and Android application (right)

The use case from Georgia concerns the investigation of climate change for a long period of 30-100 years and simulation of mineral dust pollution over the Caucasus region. These tasks were fulfilled on the GRENA high-computing infrastructure. Simulations were carried out using the Climate model RegCM4 on the GRENA HPC infrastructure for the different initial data and land models over 30 (CLM4.5) and 100 (BATS) year periods. The model output included calculated daily and monthly means of Precipitation, Evaporation, and Temperature. Additional simulations of dust distribution over the Caucasus region by WRF-Chem were conducted to extract dust particles with 0.5, 2.4 and 8.0 μm effective radius fields from the WRF model output. The results of these investigations were uploaded in

netCDF format following climate and forecast conventions to the NI4OS-Europe **digital repository service** for testing.

In Armenia, the Weather Research and Forecasting model with online coupled chemistry (WRF-Chem) is applied to simulate air pollutants for the territory of Armenia during 2018-2019, causing severe diseases. The application relies on CPU intensive WRF-Chem version 3.9 digital model customized on the Bulgarian **AVITOHOL** supercomputer supporting the NetCDF and hdf5 formats for input and output data. The application studies the sulfur dioxide and nitrogen dioxide air pollutants for Armenia's territory. The simulation output files are analyzed with the Copernicus SENTINEL-5P atmospheric measurements with high spatio-temporal resolution and ground-based measurements. But since datasets being assembled are often of different spatial and temporal resolution for analyzing these data as coherent units, the data stored, aligned, and indexed using datacube methodologies on the NI4OS-Europe infrastructure. The use case underway aimed to study and develop machine learning (ML) algorithms to increase forecasting and monitoring accuracy considering the growing amount of simulation, in-situ, and GIS datasets that can now be staged on the **simple storage service** and are available for analysis on the **AVITOHOL service**. Thus, Earth observations and digital atmospheric models provided robust monitoring to address different environmental challenges using satellite and in-situ data. **AVITOHOL** team provided various tools, compilers, and libraries to port and optimize the Armenian application for the supercomputer. Currently, the two teams are working together to carry out simulations to use the resources of **AVITOHOL** effectively. The Armenian team has received all the needed support from the administration of the **AVITOHOL** cluster.

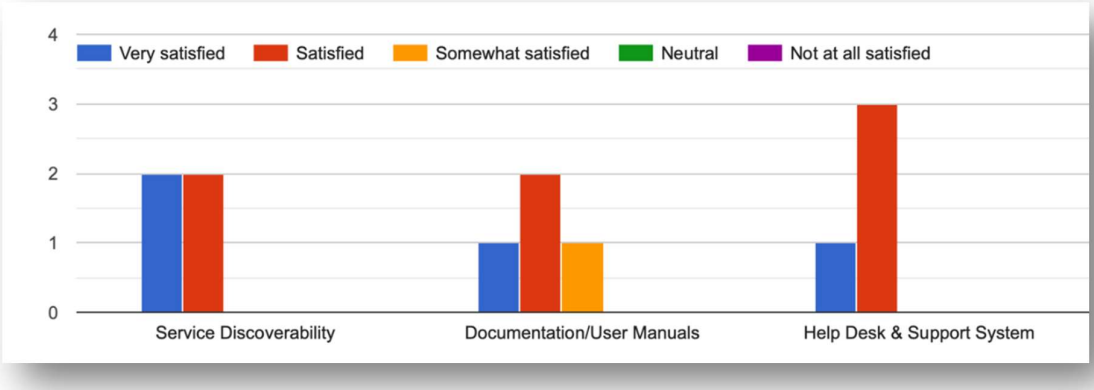

Figure 15: User satisfaction from NI4OS services. Outcomes of use cases

4. Digital Cultural Heritage

Similar to the other scientific communities, in Deliverable D6.3 a description of two use cases for the Digital Cultural Heritage community titled: "Virtual Sites and Museums" and "Clowder4DCH" were described. A number of researchers with different backgrounds in the field of Digital Cultural Heritage which appeared to be anonymous, this time, coming from different world countries and not just by those participating in the NI4OS-Europe provided reports for us, which contained evaluations of their experiences using services which are on-boarded to the NI4OS-Europe platform, which are relevant to the **Digital Cultural Heritage community** and their projects.

The organizers of the test-cases, Professor George Artopoulos, Dr. Chrysovalantis Constantinou and Dr. Andreas Athenodorou contacted via a survey (see Section 4.1) a consortium of DCH stakeholders located on countries all over the world. The evaluators provided their replies in a form of a questionnaire.

Originally, the following services were proposed to be evaluated with respect to their **accessibility**, **quality** and **efficiency**:

1. "CHERE"
2. "Clowder4DCH"
3. "OVRET"
4. "Open Mapping Application (OMApp)"
5. "B2STAGE"
6. "B2DROP"

However, the DCH community leader decided to focus on the evaluation of the DCH flagship service/repository and tools contained therein. Hence, the testing focused mainly on the services **Clowder4DCH** which is the DCH flagship service as well as the **OVRET** tool which appears as a tool in **Clowder4DCH**.

Clowder4DCH is a candidate service and will soon become fully on-boarded while **OVRET** is an active tool which has not yet reached a top TRL of 8 or 9, but will soon reach and would be considered fully on-boarded. Since OVRET is not a stand-alone service and is provided through DCH, this evaluation is expected to justify at some extend its matureness.

4.1. Questionnaire

The NI4OS-Europe digital cultural heritage community survey's purpose was to collect information about the research areas in which the NI4OS-Europe partners are involved with, and establish use cases based on the offered thematic, repository, and HPC/Cloud services. The survey included 13 questions which took only a few minutes to complete. The following information/questions were collected from the researchers:

- Email Address
- A brief description of their research (a couple of lines)
- The general domain in which their research belongs to with six available options: i) **digital humanities**, ii) **galleries, libraries, archives, and museum industry**

(GLAM), iii) **Historian, art history scholar**, iv) **built heritage conservation, preservation, and safeguarding**, v) **researcher, technician in digital practices, archiving, documentation**, and vi) **geo-humanities**.

- The NI4OS-Europe repositories that are relevant and/or useful to their research with the options being NI4OS-Europe repository service, UKIM repository, University of Maribor Library Digital Repository, Digital Archive for Ethnological and Anthropological Resources.
- The NI4OS-Europe generic/core services that are relevant to their work with the options being Data Discovery Service, Simple Storage Service (file sharing), Archival service, NI4OS-Europe helpdesk, NI4OS-Europe training and webinar portal, Data Analysis Service.
- The NI4OS-Europe thematic services which are relevant to their work with the options being Clowder4DCH online content management system for digital cultural heritage, Live Access Server (visualization of geo-referencing data), OMApp (automatic image geo-referencing), IoT cloud platform (data collection, processing and visualization)
- The researchers were then asked whether they used the following HPC/Cloud services (the only options were YES/NO): CYI Cloud facility, AVITOHOL cloud, Cyclone Supercomputer (Cyprus), AVITOHOL, FINKI cloud, PARADOX-IV cluster, ARIS.
- The researchers were then asked whether they are involved with Education, Research, Outreach Activities, Training Events, Policy making/advice, and Data Analysis with the options being YES or NO.
- The next question asked whether the researchers are satisfied with the NI4OS-Europe **Service Discoverability, Documentation, and Helpdesk** with five available options ranging from **Very Satisfied** to **not at all satisfied.**
- The next question asked whether the researcher's work requires any of the following: **HPC Services, Data Storage, Virtual Machines, Thematic Services Training, User Support,** and **Cloud Services.**
- Finally, the researchers were asked to provide a **list of publications which were enabled by the NI4OS-Europe project,** a list of any other **results such as datasets, talks, dissemination activities, or outreach events which were enabled by the NI4OS-Europe project,** and **a list of recommendations and/or suggestions for the improvement of available services.**

4.2. EOSC challenges in DCH

Digitization of our cultural heritage will promote dialogue by enabling new understandings of the past and more accurate interpretations of historical interactions between human-actors, agency and the environment. Computer vision, real-time data processing-for the interpretation of culturally valuable assets and sites, together with scalable modelling and simulation techniques, will impact significantly our effort of digitally studying tangible heritage, and applying data-driven methods in practice for the sustainable preservation of built heritage, and consequently in empowering our mission towards building resilient

futures for our societies. Datasets generated by European efforts in digitizing cultural content for preservation and online access, in response to European Commission's 2011 recommendations (https://ec.europa.eu/digital-single-market/en/digitisation-digital-preservation), have been growing and have started to include not only 2D information about cultural artefacts of museums and collections, but also 3D models and assets, as in EUROPEANA (https://pro.europeana.eu/project/3d-content-in-europeana), Horizon 2020 projects with relevant APIs, standards, metadata schemata in 3D linked datasets (https://share3d.eu/, https://www.inception-project.eu/en), as well as centralized online repositories for semantically enriched 3D representations of cultural assets (https://sketchfab.com/tags/europeana).

The EOSC-related needs of the field of Digital Cultural Heritage 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, enables interdisciplinary studies to be conducted by humanities scholars, museum curators and researchers, with background in archaeology, history, social sciences, architecture and urban studies, through the creation of online community driven dynamic repositories of datasets of cultural heritage collections and archives, and research generated semantically enriched datasets, as well as the creation of Virtual Museums for providing access to remote digital collections and inaccessible sites. Enabling remote access to knowledge and facilitating communication by means of these platforms becomes of paramount importance in our era, where digitization and remote collaboration are at the center of every societal transformation promoted by the European Commission.

In the use case assessment, representatives of the GLAM industry (Galleries, Libraries, Archives and Museums), and specifically humanities scholars, museum curators and researchers, based in Cyprus, Greece, Germany, Egypt and the US, with background in archaeology, history, social sciences, architecture and urban studies, were provided access to the NI4OS-Europe services developed for the DCH community and asked to test the available features that are relevant to their expertise and interests. After the completion of the assessment of the features available and integrated in the service packages, the users were asked to capture their experience and provide us with their feedback through an online survey (https://docs.google.com/forms/d/e/1FAIpQLSeASK2UI-RmHXq7kPACfKKR4zVCcjnwtwTfrudQ-w5Yo7TKvw/viewform), specifically created for the requirements of the DCH communities. The ongoing use case test allowed humanities scholars with expertise in 3D digitization and reconstruction techniques, as well as museum officers and curators in the region to create an online Virtual Research Environment as well as Virtual Museum experiences, respectively, for the study and representation of DCH assets from their collections, managed in appropriate form to ensure interoperability of digital assets. Description of the steps involved in the specific use cases:

4.3. First Cross-border use case in Digital Cultural Heritage: Virtual Sites and Museums

- The first step of the testing process involves the generation of 3D assets to be used in the virtual environment of the online museum to be created. This step includes the application of a photogrammetric process of using manual and unsupervised

computer vision techniques for the production of 3D models of CH sites and artefacts, from object scale to the building of a monument or an archaeological site. Two procedures are offered to the users: (a) a custom created document of instructions, including links to online resources, software and tutorials, for the step-by-step application of photogrammetry (https://clowder.hpcf.cyi.ac.cy/files/60993eb8e4b05a40fdb741ff?dataset=60891c06e4b05a40c4c90ace&space=60891bc7e4b05a40c4c90aba); (b) suggested use of NI4OS service OMApp for the generation of 3D terrain surfaces from sequences of geotiffs collected by aerial photogrammetric processes (e.g., by means of a drone survey).

Below in Figure 16 and Figure 17 we provide screen shots as well as the contents of the **Online Virtual Reality Environments Toolkit (OVRET)** produced document of photogrammetry instructions, linked above.

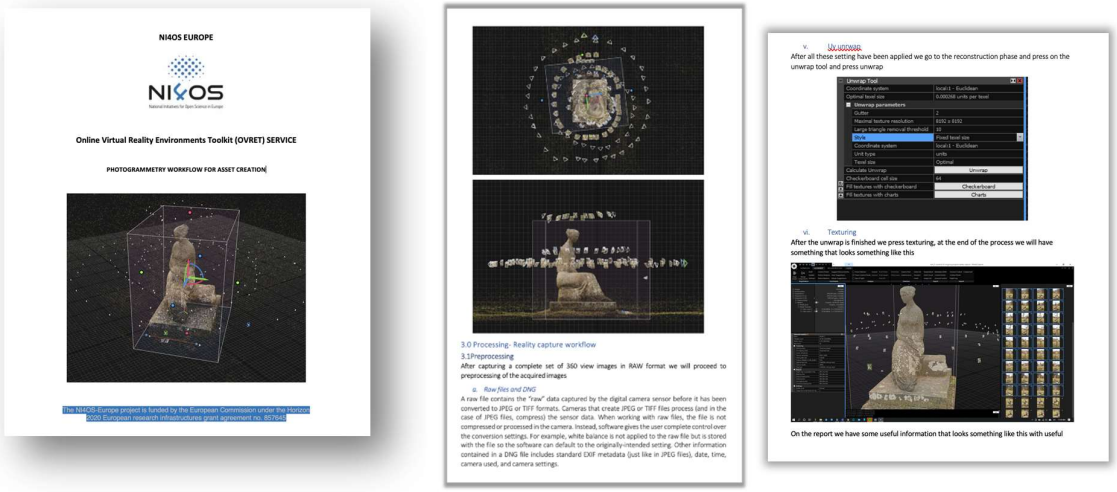

Figure 16: The front-page as well as examples taken from the instruction of Online Virtual Reality Environments Toolkit (OVRET)

- The next step involves the use of the tools created 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.

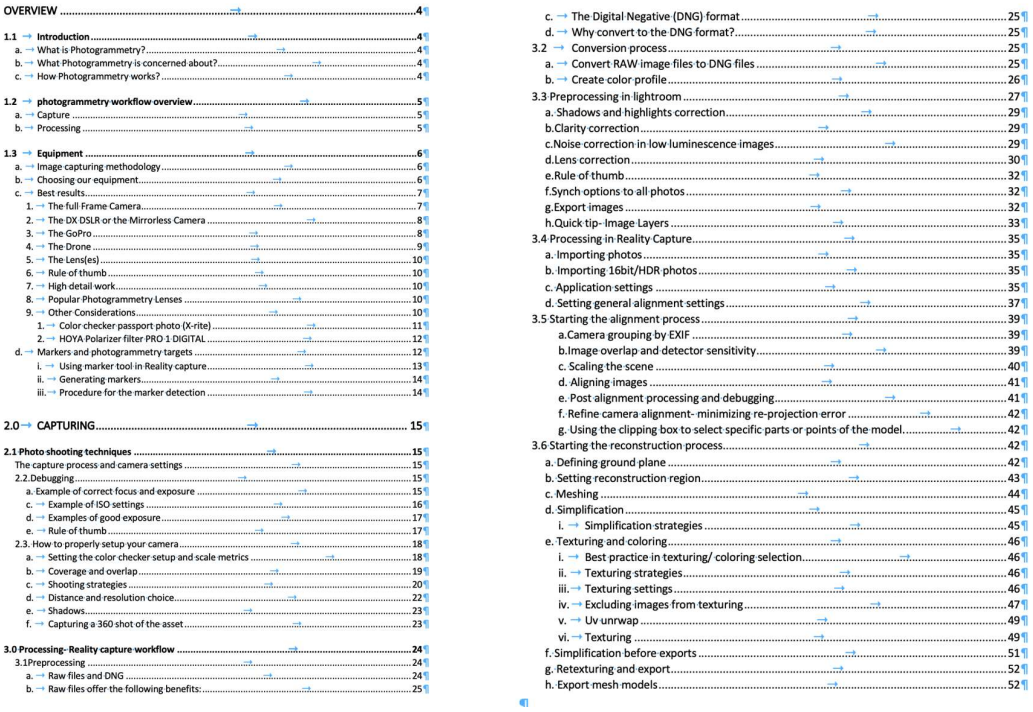

Figure 17: The content of the Online Virtual Reality Environments Toolkit (OVRET) produced document of photogrammetry instructions

The contents of the "**OVRET"** Document of instructions for the creation of virtual reality museum spaces, are provided in the following link: https://clowder.hpcf.cyi.ac.cy/files/60993e92e4b05a40fdb741f7?dataset=60891c06e4b05a40c4c90ace&space=60891bc7e4b05a40c4c90aba. For illustration purposes we provide photos of examples from the instructions in Figure 18.

- Generated datasets are be organized respecting the FAIR policies and will become open to the public via "**Clowder4DCH"** service: https://clowder.hpcf.cyi.ac.cy/. A view of the "**OVRET"** tool through "**Clowder4DCH"** (to be fully on-boarded service) is provided in Figure 19.
- Eventually, after the completion of the activity by the researchers engaged in the assessment procedure, they will be asked to upload the generated data and metadata on the "NI4OS-Europe Simple Data Repository", and existing EOSC "***B2SHARE***", "***B2STAGE***" services.

In the next subsection we provide a short description of the details of the service:

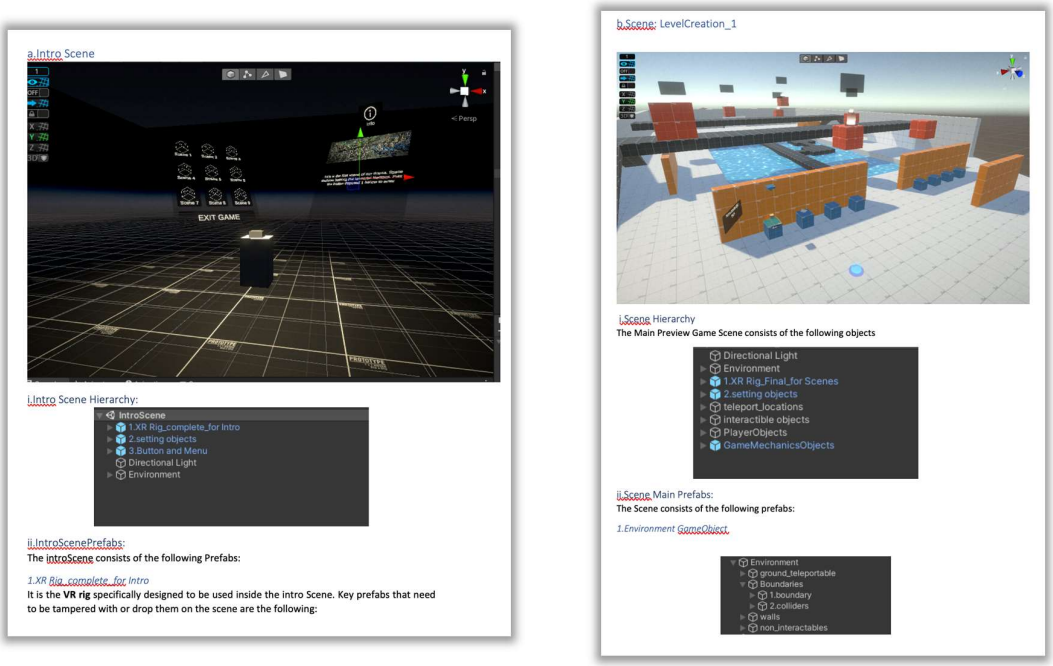

Figure 18: Examples from the instructions provided in the OVRET Document of Instructions for the creation of virtual reality museum spaces

4.3.1. Online Virtual Reality Environments Toolkit (OVRET)

In the following link we provide OVRET Space & Dataset:

https://clowder.hpcf.cyi.ac.cy/datasets/60953d20e4b0e3ffee3949b7?space=60891bc7e4b05a40c4c90aba

Overall description and scope of the service: 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 and visual interfaces for DCH assets and virtual visits of inaccessible, or demolished, heritage, monuments and historic sites.

Target user community and value provided for it: Museums, scholars, researchers and students in digital cultural heritage. Value of the service is high for user communities both for research and education and dissemination purposes.

1. *Service readiness*: Pilot.
2. Service operational details
    - *Describe access policies, conditions of access*: Creative Commons Attribution 4.0 International (CC BY 4.0).
    - The service is free for European (and international) researchers.
    - *User support for this service*: documentation / instructions & links to online tutorials. User support could be provided in the future by service provider.

Data sets which are used / produced by the service:

- This includes workflows and plugins (software scripts & code), as well as datasets of DCH assets.
- Unity application for visualizing and interacting with virtual environments using multiple interface modes, such as mouse and keyboard, Oculus Rift QUEST2.
- Interface for locomotion within the virtual environment.
- Basic application features such as console commands, GUI interfaces, render quality settings, and error logging.
- Loading virtual scenes.
- Viewing media based on user location.
- 3D object observation.
- Samples of DCH assets (3D reconstructions of cultural artefacts, monuments, buildings or sites).

Does this data already follow FAIR principles (findability, accessibility, interoperability and re-use): Datasets will respect H2020 FAIR policies. Users can search, tag, and annotate data at various granularities.

Does this data follow the EU GDPR principles: Yes. 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.

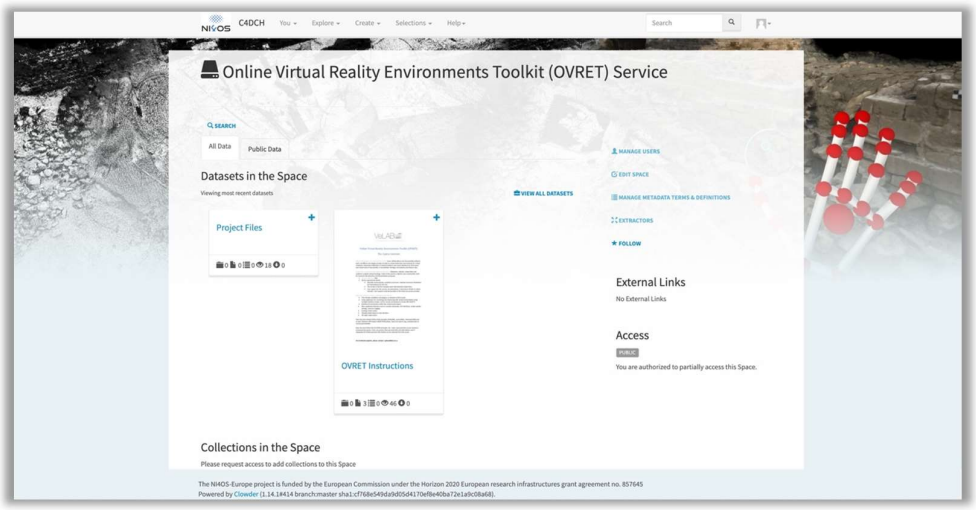

Figure 19: The OVRET service tool as this appears in Clowder4DCH service

4.4. Second Cross-border use case in 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 (see Figure 20). 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 to the user the chance to organize files in three different ways, datasets, spaces and collections. Regarding access policies of the service, users are able to register on the platform using a form. The responsible Clowder application administrator checks user information and then manually accepts 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).

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. File formats supported and extractors integrated in the NI4OS Clowder4DCH service are listed below, also to be found at: https://clowder.hpcf.cyi.ac.cy/extractors.

4.4.1. Description of Extractors and Previewers

One of the most interesting aspects of Clowder is the ability to extract metadata from any file. This ability is provided by using extractors. Available Extractors:

- ncsa.audio.preview : Creates image previews, thumbnails and previews of audio files;
- ncsa.file.digest: File digest extractor;
- ncsa.geoshp.preview: Geoshp extractor takes .zip input file and communicates with GeoServer to retrieve WMS metadata. Geoshp preview extractor can be used to display shapefiles;
- ncsa.geotiff.metadata: GeoTiff Metadata extractor, extracts metada of a GeoTiff image. This extractor can be used in combination with the ncsa.geotiff.preview extractor;
- ncsa.geotiff.preview: GeoTiff extractor communicates with GeoServer to get WMS metadata. Geotiff preview extractor can be used to display GeoTiff files. This extractor can be used in combination with the ncsa.image.preview extractor;
- ncsa.image.metadata: Extracts metadata from Image files;
- ncsa.image.ocr: Simple OCR (Optical Character Recognition) extractor to extract text from an image;
- ncsa.image.preview: Creates thumbnail and image previews of Image files;
- ncsa.nlp.simplelanguage: Detects the language of the content of a text file;
- ncsa.nlp.simplesummary: Finds the sentence of the text which contains the most relevant words based on the text's word frequency (ignoring stop words);
- ncsa.nlp.tika: Text extracted from the input and additional file metadata are associated with the original file;
- ncsa.pdf.preview: Creates image previews, thumbnails and pdf previews;
- ncsa.video.preview: Creates thumbnail and image previews of Video files, as well as downscaled video preview.

Previewers are used to visualize information about datasets and files. Previewer can work together with extractors and external services. List of available previewers:

- Audio previewer is a previewer for .mp3, .mpeg and mpeg3 files.
- Html previewer is a previewer for .html files.
- Plain text previewer is a previewer for .txt, .doc, .docx and .csv files.
- Viewer 3DHOP is a previewer for .ply and .nxz 3D object files.
- Threejs is a previewer for .fbx 3D object files.
- 3D pdf is a previewer for .u3d files.
- Pdf previewer is a previewer for .pdf files.
- Person tracking is a previewer for video files. This previewer shows you when different subjects enter the frame and leave the frame and even where in the video those subjects are.

With the C4DCH service, users upload massively (zipped datasets) or individual files of:

a. 3D Models (as 3D PDF; FBX, PLY and NXZ 3D files): 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). NI4OS C4DCH created a Previewer (based on CNR's 3D HOP API), as well as an FBX Previewer (Threejs) for the online visualisation of 3D models, at https://github.com/clowder-framework/clowder/tree/develop/public/javascripts/previewers/three_js, a contribution that was acknowledged by the CLOWDER community.

b. Scanned books and their metadata, as PDF files: 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.

d. GeoTiff files for georeferenced land surface (remote sensing) image collections.

In Clowder4DCH service users can search, tag, annotate data at various granularities.

Datasets 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.

Useful links that were provided to test users prior to the evaluation sessions below.

CLOWDER 4CH

https://catalogue.ni4os.eu/?_=/resource/id/d0c2baae-45dc-496f-922c-dcb051ce5fc0

NI4OS repo:

https://repo.ni4os.eu/; https://catalogue.ni4os.eu/?_=/resource/id/b35269cb-82f2-4aec-bc61-f2684069bff3

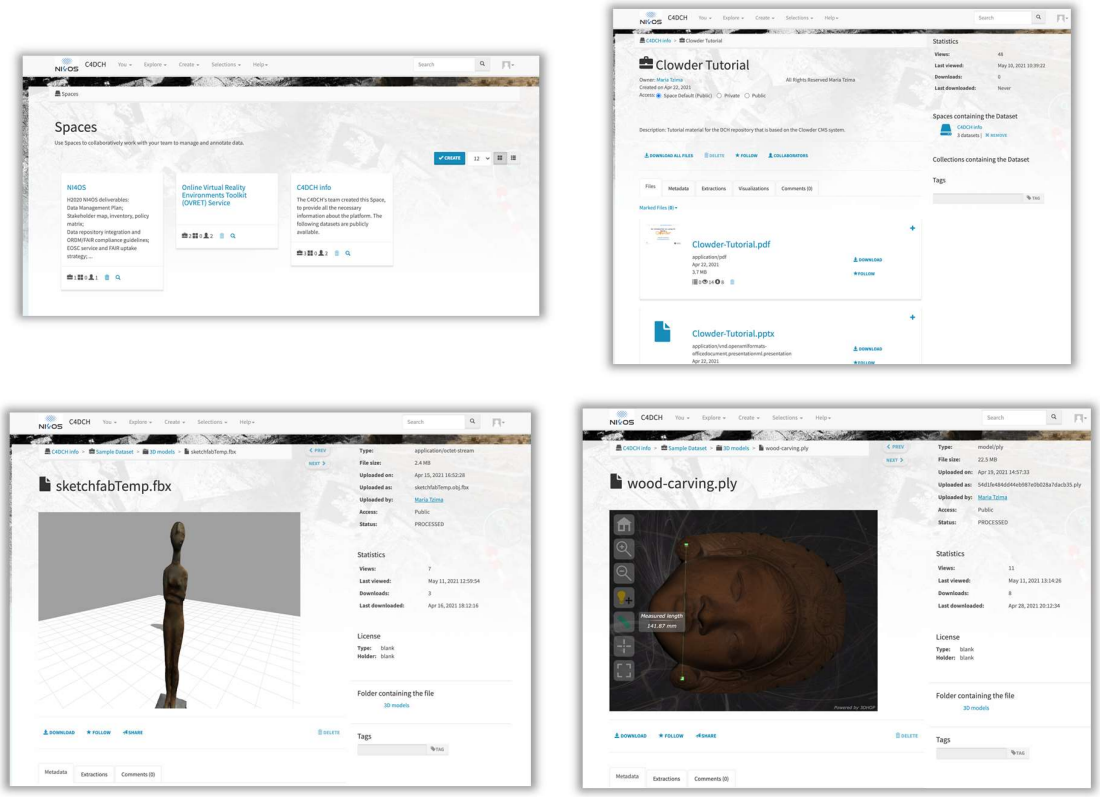

Figure 20: The clowder service's Spaces, service tutorial, and FBX/PLY file previewers

4.5. User feedback collected from the test sessions

Current version of OVRET service was developed by building on past experience of the H2020 Vi-SEEM Digital Cultural Heritage user community needs, and mostly on the results collected after the respective NI4OS-Europe webinar. Below the topics that were offered to the community members participating in the webinar.

1.Service description

- **Introduction- setup field**
  - Service name and description;
  - Goal of service;
  - How it relates to FAIR principles;
  - Target communities;
  - A methodology applicable in many scales and scenarios;
- **Past experience/ state of the art**
  - Past projects that we are involved:
    - Vi-seem;
    - Visumf;
  - State of the art examples from list of examples:
    - 360 platforms:

        - Examples;
        - Platforms;
    - Virtual reality platforms;
    - Augmented reality;
- **State the challenge for the project- launch your virtual exhibition space in less than a day**
    - What it involves?

2.Service scenarios and possible specific services

- **Workflows for virtual exhibitions**
    - **Option 1- start from scratch- assets and environment**
        - Photogrammetry for asset creation;
        - Photogrammetry for environment creation;
        - Add interaction with unity;
    - **Option 2- using existing digital assets or recreate only assets and use preset environments:**
        - Follow appropriate instructions to create 3d models of assets;
        - Upload models in a digital repository that acts as a medium;
        - Use existing template or virtual environment to create an appropriate virtual experience;
    - **Option 3- create a virtual tour using 360 photos of existing exhibitions:**
        - Create 360 imagery of place of implementation;
        - Use appropriate platform to create the transitions;

3.Tools to be used- proposed tutorials

- **Tools to be used**
    - Photogrammetry workflow;
    - Unity VR interaction workflow and possibilities;
        - Interactivity;
        - Gamification of the process;
    - Online platforms;
        - Mozilla hubs;
    - Future possibilities;

4.Interactive session Q & A

- **Inquiry about interface and style of service:**
    - Preference for step by step;
    - Preference for direct input of models and resources on an online platform like Mozilla hubs;
    - Preference for a standalone web-based museum approach + 360 virtual visits;
    - Preference for an interactive website with direct connection with Sketchfab?
- **Inquiry about possibilities:**
    - Would you be interested in applying gamification techniques into the VR processes in order for the simulation and the virtual tour to be more engaging?

- Would you prefer for the VR rooms to be used by a single user or to have the ability for multiplayer use with user to user interaction?
- Use artifacts and assets as puzzles and introduce cyberspace only mechanics and aesthetics or recreate the physical environment?

In addition, the use case presented was assessed by humanities scholars, museum curators and researchers, based in Netherlands, Cyprus, Greece, Germany, Berlin, Egypt and the US, with background in archaeology, history, social sciences, architecture and urban studies, who provided us with their feedback, below. Some general remarks from the survey:

- All responders consider an online CMS for collaboration as a very useful service for their activities.
- 100% are involved in research.
- 30% of the responders are also involved in the education sector.
- 30% of the responders are also involved in the outreach sector.
- 50% of the responders rely on the data analysis operations.
- Service discoverability of NI4OS: 66% were satisfied; 22% very satisfied; 22% somewhat satisfied.
- Documentation/User Manuals of NI4OS: 66% were very satisfied; 22% satisfied; 22% somewhat satisfied.
- Support provided by NI4OS: 66% were satisfied; 22% very satisfied; 22% somewhat satisfied.
- 50% of their work requires "Data storage", and 50% "Thematic services training".

The research interests of the responders were very topical and relevant to the scope of the services, and we provide an exemplar list of the descriptions below, grouped by topic of interest:

- Use of online CMS solutions for the creation of personalized and flexible project repositories and research collaboration spaces: 40% of responders.
- Use of ICT for digital heritage applications, such as smart applications in tourism, museums, etc.: 30% of responders.
- Use of digital 3D documentation of heritage for study, protection, safeguarding, preservation, conservation and education purposes: 30% of responders.

4.5.1. User feedback about OVRET service

User A: [female; age: 20-40] digital humanities scholar, with education background in architecture, based in Germany; higher degree: MPhil.

- ***Research interest in using the OVRET service****:* Applying digital documentation and analysis for the study and protection of built heritage.
- ***Overall experience***: It was a pleasing experience.
- ***Describe your experience***: I feel that it was designed for users with basic unity knowledge, that why it was a bit hard for me understand some sections of the draft. But after a couple of reads, I started to get the logic and everything was easy to follow.
- ***Comments***: I had two issues: One when I during building the game, because I version of Unity that I had was missing (IL2CPP) script was missing so I googled

how to fix it. Second: After the building was done, I was unable to navigate the scene but I think it was because I was not using a VR set.

- ***Suggestions for improvement***: it will be important to users if you include a troubleshooting section identifying the most common problems that they can face, and how to solve them.

User B: [male; age: 40-50] architect; Researcher, technician in digital practices, built heritage conservation, preservation and safeguarding, 3D documentation of built heritage; based in Egypt; higher degree: MSc.

- ***Research interest in using the OVRET service****:* apply immersion solutions for community communication and interpretation of Built Heritage to contribute to conservation and management processes.
- ***Overall experience***: satisfied.
- ***Comments***: The functionalities of the service require to be more elaborative and more specific towards their benefits in relation of building an immersive environment.
- ***Suggestions for improvement***: Regarding the practical testing of the service, the service included functionalities that need to be more informative and provide guidance to enable the users determine which tools to choose, as well as why and how to use them and their relevant tools. Regarding the service's instructions and documentation, these require to be less technical and more descriptive in relation to the scope of their purpose and functionality, the functionality of each tool, as well as have to better diagrams of the various steps that should be taken by the user and how to make them work.

4.5.2. User feedback about C4DCH service

User C: [male; age: 20-40] Digital Humanities; Researcher, technician in digital practices, archiving, 3D documentation of built heritage; based in USA; higher degree: MA.

- ***Research interest in using the C4DCH service****:* the creation of personalized and flexible project repositories and research collaboration spaces.
- ***Overall experience***: NI4OS offers an alternative content platform that meshes better with institutional imperatives and creative/intellectual rights management.
- NI4OS -Europe project enabled the responder to publish research work.
- NI4OS -Europe project provided the responder with cloud services for research datasets storing and dissemination activities.
- ***Suggestions for improvement***: FBX previewer is great but appears vertically stretched.

User D: [male; age: 20-40] Digital Humanities; Researcher, technician in digital practices, archiving, 3D documentation of built heritage; based in USA; higher degree: MA.

- ***Research interest in using the C4DCH service****:* developing smart technological interfaces for personalized tourism experience in post conflict sites to improve the socioeconomic condition in conflict affected countries.
- ***Overall experience***: very satisfied.

User E: [male; age: 20-40] Digital Humanities; Historian, Art History scholar; Researcher; based in USA; higher degree: MA.

- ***Research interest in using the C4DCH service****:* developing smart technological interfaces and online platforms for education purposes and the promotion of heritage sites, as well as for community building around monuments. I am interested in the politics of cultural heritage and the afterlives of public products of cultures.
- ***Overall experience***: satisfied.
- ***Comments***: Responder used C4DCH to conduct public engagement and outreach events on several occasions.

4.6. WIP Services promoted and disseminated through international networks of communities

C4DCH service, in a work-in-progress state, has been presented and promoted in international events and venues on several occasions. List below:

- (25-26/07/21) Artopoulos, G., 'H2020 NI4OS C4DCH service', in **DARIAH ERIC WG** workshop in Palermo, Italy (to take place).
- (15/07/21) Artopoulos, G., 'H2020 NI4OS C4DCH service', in **CAAD Futures 2021** workshop & round table, to be organized by the University of Southern California.
- (11/05/21) Artopoulos, G., 'H2020 NI4OS C4DCH service', in "Discovering discovery" ThatCamp organized by the **TRIPLE project** (Grant Agreement No. 863420), CNRS, Huma-Num.
- (01/2021) Artopoulos, G., 'H2020 NI4OS C4DCH service', in the Scientific Advisory Board of the **NSF-funded Cyberinfrastructure for Sustained Scientific Innovation** (CSSI), coordinated by NCSA, University of Illinois at Urbana-Champaign, http://www.ncsa.illinois.edu/news/story/can_you_imagine_supporting_data_infrastructures_with_software
- (02/12/20) Artopoulos, G. and M. Deligiorgi, 'Urban PERISCOPE', in Scholarly Primitives - **DARIAH Annual Event 2020** (https://www.dariah.eu/event/dariah-annual-event-2020/).
- (23 -24/11/2020) Artopoulos, G., 'Urban PERISCOPE', at the virtual Conference of **TIME MACHINE EU** (https://www.timemachine.eu/tm2020-poster-wall/).

4.7. Next steps of development and (tentative) conclusions

Concluding this mid-term reporting, a couple of overall remarks are listed:

- The scientific community leader has confirmed there is an important user community active in the GLAM industry and digital cultural heritage research field who are interested in the provided services, namely in user-driven virtual research environments for collaboration and immersive representations of cultural heritage – the online survey was not concluded, and we are receiving messages from users

who continue testing the services, and will keep providing us with further feedback for improvement in the next few weeks.

- Clowder4DCH service is mature enough to be fully on-boarded on NI4OS-Europe.
- Feedback received confirms the expectation that with the addition of support content (examples, tutorials and FAQs), the OVRET service can become a cutting edge, highly popular and reliable tool that can be provided to cultural heritage and digital humanities communities of the NI4OS-Europe both as a feature of C4DCH (as in its current state of development), as well as to wider audience via GitHub (a relevant account on GitHub has been already created by the authors).

The next steps in the development of the services presented involve the following actions:

- Fine tuning C4DCH service;
- Further elaborate and expanded instructions of installing a personalized C4DCH instance locally;
- Fine tuning FBX previewer;
- Experimenting with new extractors and integrating those which are requested by user communities, e.g., Building Information Modelling dataset viewer, or GIS data viewer;
- Expand and further elaborate instructions for the OVRET service;
- Inclusion of more tutorials and links to online resources for the OVRET service;
- Introduce FAQ section for the OVRET service.
- Provide OVRET to wider community via its dedicated GitHub space.

5. Computational Physics

As explained in the Deliverable D6.3, in addition to the three flagship communities of Life Sciences, Climate Sciences and Digital Cultural Heritage, the NI4OS-Europe consortium worked towards the establishment of an additional scientific community which was expected to carry out a use-case. A survey, demonstrated that a number of mature thematic services belonged to the scientific community of computational Physics. Hence, the consortium of NI4OS-Europe, assigned the duties of the scientific community leader to Professor Bojana Koteska (UKIM) who coordinated the demonstration in computational Physics. The task focused on testing the thematic service **Schrödinger API** using the generic Service **FINKI Cloud** which are fully on-boarded in NI4OS-Europe service catalogue. The Interface of Schrödinger API is presented in Figure 21.

As we have already mentioned in Section **Error! Reference source not found.** Professor Bojana Koteska compiled a survey aiming to test the service **Schrödinger API.** It should be made clear than since FINKI-Cloud is used to host the thematic service **Schrödinger API,** an external user can only test the thematic service. Thus, she adopted two approaches of testing the two services.

1. In the first approach she contacted a network of scientists working on computational physics, coming from all over the European region, and asked them to test **Schrödinger API** and then fill in the survey.
2. In the second approach, she formed a team consisting of Professor Ljupco Pejov who performed a scientific investigation from scratch based on the thematic service **Schrödinger API.**

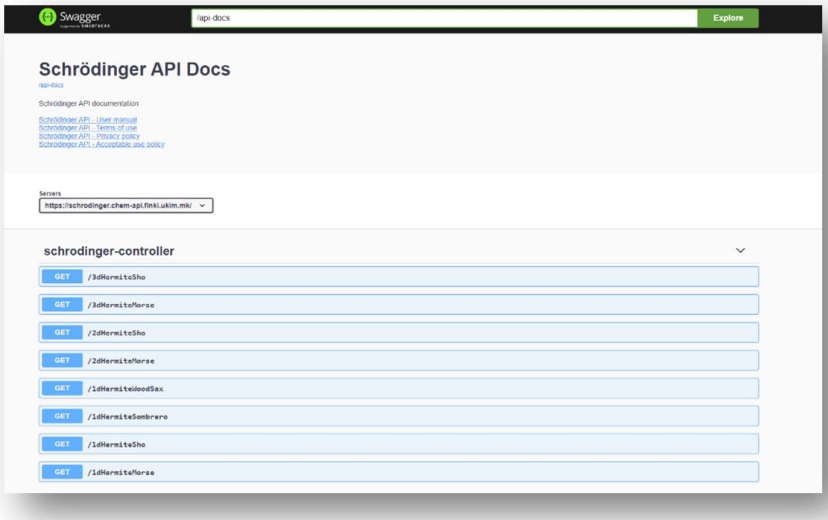

Figure 21: The Schrödinger API interface

In the following two sections we provide the results obtained by the above two approaches.

5.1. Testing the service through a survey

The survey described in Section 2, was sent to over 50 people all over the European region. All of them are researchers in the field of computational physics. However, only seven people provided full responses to the survey. Below, we will provide a brief summary of the responses, for every one of the four sections of the survey namely (1) familiarity with API concepts, (2) documentation, (3) code and structure, as well as (4) Jupyter Notebooks, Schrödinger API usefulness, and technical problems.

In the first section of the survey we asked two questions, with the first one being whether people are familiar with REST API concepts in which 3 users responded yes and 4 said no. The second question asked whether users are familiar with Jupyter notebooks to which 4 people said that they are familiar with Jupyter notebooks (which the Schrödinger service uses), while 3 people said they are not familiar.

In the next question researchers were asked questions related to the documentation of the service. The majority of the users responded that they either agree or strongly agree to questions such as whether the documentation is clear, contains sufficient examples, and that the examples are easy to understand and reproduce. In Figure 22, we see for example the responses to the question whether the documentation is easy to understand, where the answer strongly agree is shown with blue color, the answer agree with orange, and with green color we see one response for the answer neither.

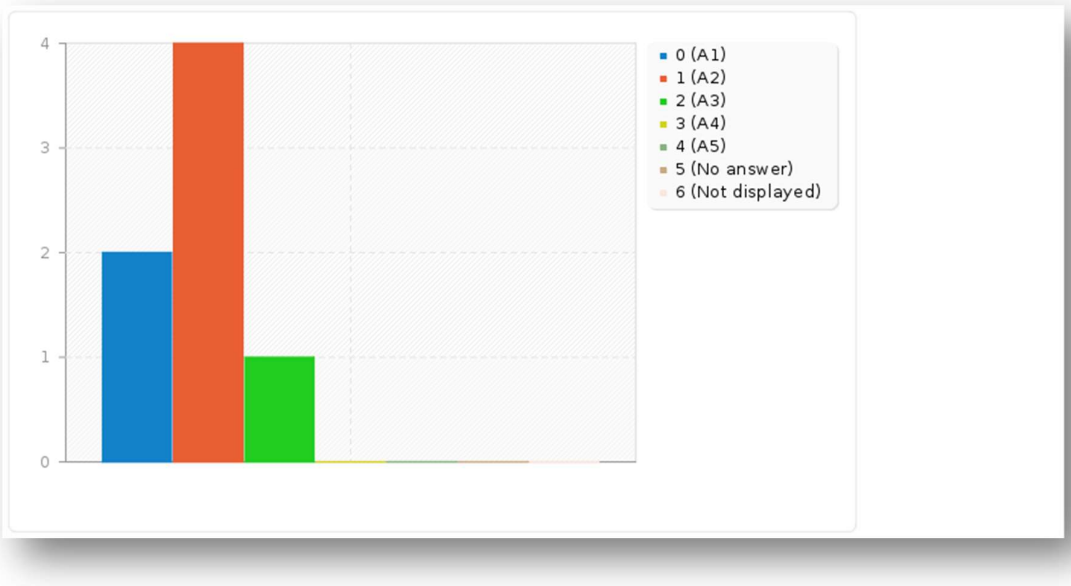

Figure 22: Replies to the question whether the documentation is easy to understand where A1 corresponds to strongly agree, A2 to agree, A3 to neither, A4 to disagree and A5 to strongly disagree

In the next section the users were asked to provide answers related to the code used inside the Jupyter notebooks. Again the majority of users responded that they found the method names intuitive, easy to understand and use, etc. In Figure 23, we see the

response to the question whether methods are easy to use to which the blue color again denotes that the users strongly agree with the statement, the orange color that they agree, and the green that they neither agree nor disagree.

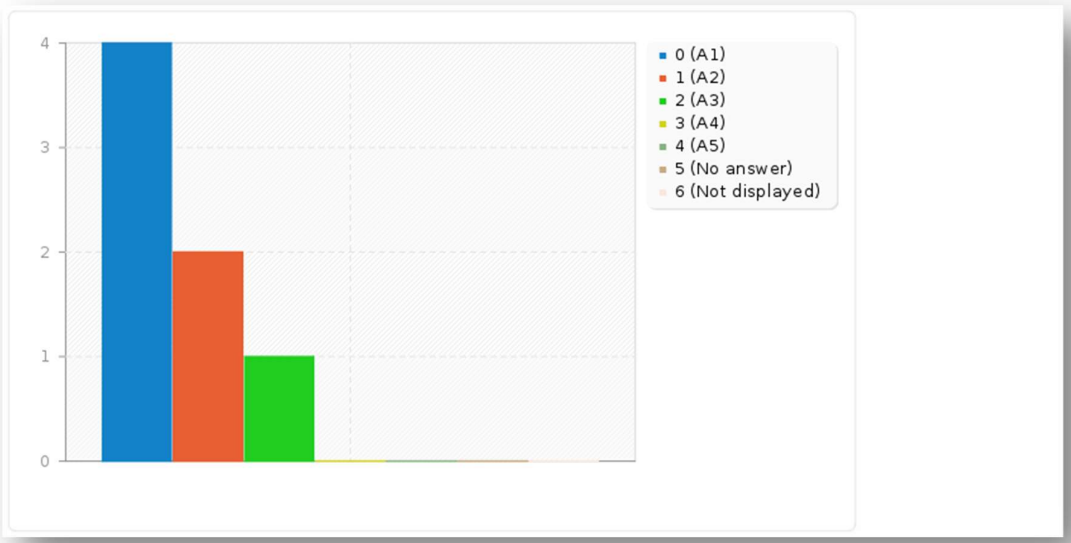

Figure 23: Replies to the question whether methods are easy to use where A1 corresponds to strongly agree, A2 to agree, A3 to neither, A4 to disagree and A5 to strongly disagree

In the following section users were asked to report their answers to questions related to Jupyter notebooks. Specifically, they answered to questions such as whether the representation format in Jupyter Notebooks is easy to read, whether the combination of software code, computational output and explanatory text into a single document is useful, etc. The majority of researchers answered that they find the Jupyter notebooks useful.

In the next section the researchers provided answers related to the usefulness of the Schrödinger API. Specifically, some of the questions asked were related to whether the researchers find the service useful to their work, whether they would recommend the service to other researchers, and whether the service would help them increase their productivity. The majority of researchers either strongly agreed or agreed with these statements.

Finally, in the last section the researchers where asked to report any technical difficulties they faced. Specifically, to the questions whether they experienced technical problems with the service connection, whether they experienced technical problems while running the service, whether they experienced technical problems with the Jupyter Notebook connection, or whether they experienced technical problems while running the Jupyter Notebook all the researchers responded that they either disagreed or strongly disagreed indicating that the service run smoothly.

In the open-ended question about reporting technical problems one researcher indicated that the user manual in this webpage is not working: https://schrodinger.chem-

api.finki.ukim.mk/swagger-ui/index.html?configUrl=/api-docs/swagger-config#/schrodinger-controller/HermiteSho3D. However, upon examination of the aforementioned url by the task leader, Dr. Chrysovalantis Constantinou the user manual appears to be working properly. Finally, in the open-ended question asking the researchers for suggestions for service improvement one researcher suggested that the developers of the service should provide a description of what the service does and who the target group is.

5.2. Testing the service through a scientific investigation, a brief introduction to the investigated problem

Understanding the gas-phase structure and dynamics of biomolecular species is the first step towards better insights into their inclusion in more complex, biologically relevant media. The advent of infrared multiple photon dissociation (IRMPD) spectroscopic techniques has enabled numerous size-selected species to be explored in detail [4]. In this technique, a molecular ion is first mass-selected, trapped in either Penning or radio-frequency trap, and subsequently irradiated with tunable infrared laser radiation. Once the laser frequency is resonant with an intramolecular vibrational mode of the ion, the corresponding IR photons are absorbed, which leads to its heating and subsequent fragmentation. A mass spectrometer is used to detect the fragment ions, serving as photon absorption indicators.

The IR action spectrum can thus be obtained if one follows the dependence of the extent of fragmentation on the laser wavelength. Alternatively, the IR action spectra can be obtained by a combination of UV and IR lasers. The UV laser is initially used to cause ion fragmentation provided that the particular ion has got an appropriate chromophore. Taking the scan of the IR laser over a given vibrational band causes absorption of the radiation and subsequent reduction of the ground state population and consequent decrease of the fragmentation signal induced by the UV irradiation.

In general, the number of absorbed photons that will eventually lead to ion fragmentation depends on numerous factors, such as: the photon energy, the dissociation barrier, the initial temperature of the ion etc. In the case of weakly bound clusters with small dissociation energies cold ions can be dissociated upon absorption of even a single IR photon. This method is usually applied by employing a weakly bound so called "molecular messenger" to "tag" the investigated ion in a molecular beam or in an ion trap. H_2, He, Ne, Ar are all typical examples of tags that attach to ions at very low temperatures and that can be detached upon absorption of one or more IR photons [4]. Such photodissociation technique, demonstrated for the first time by Okumura et al. [5, 6] has been applied to numerous ionic species. Aside from the obvious advantages of this method, there is also an implicit assumption that the tagging species should not actually cause significant perturbation of the ion of interest. For example, Ne and Ar are known to cause quite large perturbations of the vibrational spectra of investigated ions, because of their large polarizabilities. Molecular hydrogen, on the other hand, was first thought to cause only minor perturbations, as demonstrated in the studies of protonated dipeptides, dicarboxylate anions etc. A recent study of protonated glycine and betaine, however, seemed to have demonstrated quite the opposite. Tagging with even a single H_2 molecule appears to affect significantly some vibrational modes of these species.

Results from the experimental spectroscopic observations, however, lead to differences in the vibrational eigenstates, not revealing their exact origin. Most of the band assignments, therefore, are usually based on empirical arguments and on "chemical intuition". As a consequence of this, any serious attempt to provide an in-depth understanding of molecular force fields should be based on sound theoretical analysis and argumentation.

Most of the simple and straightforward approaches towards this aim, which are automated in the available quantum mechanical computational codes, are based on algorithms involving the double-harmonic normal mode approximation, and are being used often in conjunction with a density functional theory (DFT) – based method. While the harmonic approximation may suffice for many of the intramolecular modes, this is not the case with modes involving stretching vibrations of the X-H fragments. There is a firm evidence that systematic errors inherent to widely-used DFT methods may cancel out with those due to the harmonic approximation, thus hampering any possibility to get a deeper insight into the nature of the molecular vibrational force field [7, 8].

Therefore, any rigorous calculation aiming to clarify the experimental observations and, at the same time, give a correct insight into the factors determining the vibrational frequency shift of a particular mode of interest must be based on an anharmonic treatment of the oscillators in question. This is particularly true for X-H oscillators, due to the small mass of the H atom and consequently the relatively high vibrational amplitudes, spanning a significant segment of the notably anharmonic part of the vibrational potential. Though theoretical approaches have been established for this purpose, such as the vibrational self-consistent field methods and the perturbation theoretic approach by Barone [9], these are computationally rather demanding and expensive, so that having a simpler, yet rigorous theoretical approach for this purpose is highly desirable. We have therefore developed and implemented the approach described further in the text.

5.3. The systems under study

The present use-case illustration is a part of a bigger project aiming to provide an in-depth theoretical understanding of the tagging of protonated glycine with H_2 and He in gas phase. To avoid any ambiguities in the data interpretation, it is rather useful to consider the tagging of protonated glycine's analogue in which all amine group hydrogen atoms have been replaced with methyl groups, known as betaine [10]. The structure of betaine is shown in Figure 24 below.

Specifically, in the present study we have focused on the process of betaine(+) tagging with molecular hydrogen in gas phase.

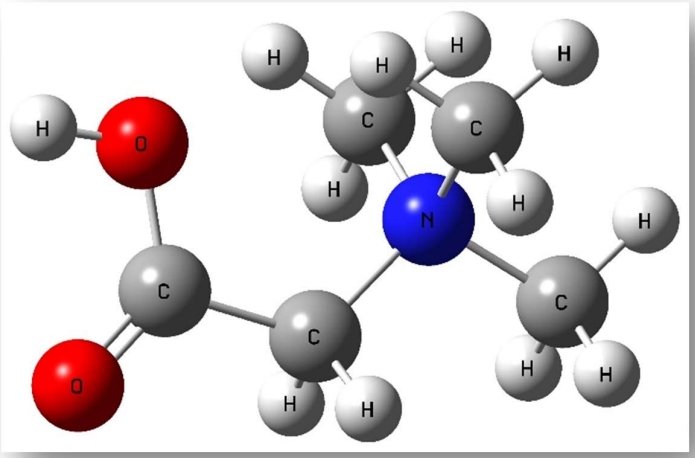

Figure 24: The structure of betaine(+)

5.4. Theoretical approach to the computation of anharmonic vibrational frequencies

To calculate the anharmonic vibrational frequencies of the O-H stretching modes of the free and interacting systems in the present study, the researchers have adopted two approaches: a static and a dynamic one.

Within the static approach, they first carried out thorough explorations of the potential energy hypersurfaces (PESs) of the investigated systems using the combination of Becke's three-parameter adiabatic connection exchange functional (B3 [11]) with the Lee-Yang-Parr (LYP [12]) correlation functional – B3LYP, as well as the long-range corrected variant of B3LYP by the Coulomb-attenuating method developed by Handy's group (CAM-B3LYP [10]). The aug-cc-PVTZ basis set was used for orbital expansion to solve the Kohn-Sham equation iteratively. The "ultrafine" pruned (99,590) grid (99 radial and 590 angular integration points) was used in numerical integration in all DFT calculations. In parallel, just for comparison purposes, they have also done the same calculations with the density functional tight binding method (DFTB). In particular, they have used the DFTB variant with analytical computation of relevant the matrix elements (DFTB-A [14]).

All geometry optimizations and searches through the explored PESs were done with the Schlegel's gradient optimization algorithm [15]. The character of the located stationary points on the PESs was further on tested by performing analytical calculations of the Hessian matrices. Absence of negative eigenvalues of the Hessians indicated that a true minimum on the PES is in question. At the same time, by diagonalization of the mass-weighted Hessians they have calculated the harmonic vibrational frequencies.

Subsequently, to compute the anharmonic vibrational potential of the O-H stretching mode, which is essentially localized and decoupled from all other intra- and intermolecular modes, they have generated a series of configurations in which the positions of all atoms

and the center of mass of the O-H oscillator were kept fixed, while only the O and H atoms were moved in opposite directions, according to the following equations of motion $\dot{\vec{r}}(O) = \frac{m_H}{m} \cdot \dot{\vec{r}}$ and $\dot{\vec{r}}(H) = \frac{m_O}{m} \cdot \dot{\vec{r}}$ where $\vec{r} = \vec{r}(H) - \vec{r}(O)$.

Further, they have carried out a series of single-point energy calculations for each of the generated configurations in the case of all considered O-H oscillator. Total of 20 single-point calculations were carried out for each of the O-H oscillators considered, varying the O – H distance from 0.70 Å up to 1.65 Å. The obtained vibrational potential energy curves $V = f(r_{OH})$ were subsequently fitted to the Morse function of the following functional form $V(r_{OH}) = D \cdot \{1 - \exp[-a \cdot (r_{OH} - r_{OH,0})]\}^2 - D$.

Finally, using the 1D Morse function, the vibrational Schrödinger equations were solved by the discrete variable representation approach, as implemented in the **Schrödinger API** web service. The usage of this service is explained in the following section.

The dynamic approach was based on performing a series of molecular dynamics simulations for free betaine(+) as well as for the betaine(+)…H_2 dimers in gas phase, at DFTB level, employing the atom-centered density matrix propagation scheme (ADMP [16, 17]). ADMP simulations were performed at 40 K (which is close to the experimental conditions reported in Ref. [10]), in the NVE ensemble. Appropriate amount of initial nuclear kinetic energy corresponding to the desired temperature of the system was injected and distributed among the atoms. No thermostats were used to keep the temperature constant throughout the simulations. Initial equilibration phase of the MD simulations of 2 ps was followed by production phase of 11 ps. Cholesky decomposition scheme has been used to obtain the orthonormal basis set. The time step that has been used to integrate the equations of motion was set to 0.1 fs. Electron fictious mass parameter was set to 0.1 atomic mass units. The particular choices of the last two parameters were aimed to keep the adiabaticity index and the conditions for idempotency of the electronic density matrix within the literature recommended values throughout the whole trajectory. From the productive phase of simulation (the equilibrated ADMP runs), total of 100 statistically independent configurations (snapshots) were extracted and subjected to subsequent quantum mechanical calculations of the O-H stretching potentials in a manner identical to that explained before.

5.5. Usage of the Schrödinger API web service developed within the framework of the NI4OS-Europe project

To solve the vibrational Schrödinger equation, the researchers have used the Schrödinger API for multidimensional time-independent Schrödinger equation using Hermite DVR approach (https://schrodinger.chem-api.finki.ukim.mk/). This web service is hosted on the openstack FINKI cloud.

For the purpose of their study, they have used the module for solving the one-dimensional (1D) Schrödinger equation for Morse potential. In particular, they have called the function for 1D Schrödinger equation in case of Morse potential – 1dHermiteMorse which depends on the parameters D (dissociation depth), a (inverse "width" of the potential) and r_0 (equilibrium bond distance) as well as npts (number of points) and prec

(precision). An example of introducing these numbers in the 1dHermiteMorse function is presented in Figure 25.

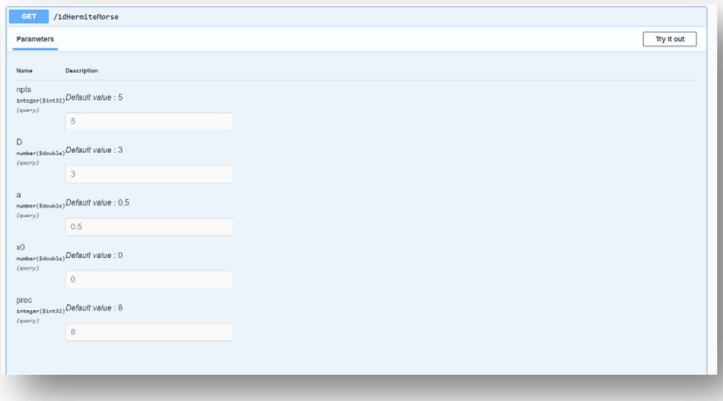

Figure 25: An example of the 1dHermiteMorse input interface

As they have computed the vibrational potential energy curves in a pointwise manner, and subsequently fitted the potential to the Morse model function, for each oscillator they have different values of the parameters D, a and r_0 (denoted as x_0 in the API). The number of points for DVR computations (npts), as well as the precision (prec) remain the same for all cases, to keep the computational consistency throughout the study and to enable comparability of the results.

The advantage of using the Schrödinger API becomes particularly obvious when it comes to automating the process of analyzing the results from the molecular dynamics trajectories. For each trajectory, as mentioned before, a total of 100 snapshots have been extracted and the OH stretching vibrational potentials have subsequently been computed and fitted to a Morse function. One has, therefore, to solve a total of 100 vibrational Schrödinger equations per trajectory. Using the Schrödinger API simplifies the computational efforts to a great extent, and also allows for an automation of the process.

A script that calls the 1dHermiteMorse method was created for automation purposes. This script is executed each time with different input values for the D, a and r_0 parameters. For example, if the provided values are D = 4.0, a = 0.5 and r_0 = 0.0, the call has the following form: https://schrodinger.chem-api.finki.ukim.mk/1dHermiteMorse?npts=10&D=4.0&a=0.5&x0=0.0&prec=6

The obtained eigenvalues were used to calculate the anharmonic vibrational frequencies (i.e. wavenumbers) of the fundamental $|0\rangle \rightarrow |1\rangle$ vibrational transitions corresponding to the O-H stretching modes, from the energy differences between the ground and the first excited vibrational states. Of course, in the same manner one could straightforwardly compute the wavenumbers of the higher-order transitions (the vibrational "overtone" transitions $|0\rangle \rightarrow |2\rangle$).

5.6. Results, discussion, and future work directions

In the left panel of Figure 26, one can see the typical vibrational energy curves for free betaine(+) as well as for the betaine(+)OH … H_2 dimer where tagging is realized on the O-H oscillator side calculated as explained before.

One can readily observe a "downshift", i.e. a shift of the potential energy curve to lower energies at higher O-H distances, implying increase of the anharmonicity and a red shift of the O-H stretching wavenumber in the case of betaine(+)OH … H_2 dimer, as compared to free betaine(+) cation. A typical fit of the O-H stretching potential energy function with Morse function of the form (5) is shown in right panel of Figure 26. All the Morse function fits were excellent, judging from all statistical figures of merit (e.g. adjusted R^2 was always higher than 0.999).

The frequencies (i.e. wavenumbers) of the fundamental $|0\rangle \rightarrow |1\rangle$ vibrational transitions corresponding to the O-H stretching modes in the case of free betaine(+) and betaine(+)…H_2, with the H_2 tagging realized on the CH_3 and OH sides, and the corresponding frequency shifts calculated according to the described. It appears that the noncovalent interaction of betaine (+) with molecular hydrogen "tag" from the O-H oscillator side leads to substantial downshift (red-shift) of the frequency of the O-H stretching vibration. The implemented rigorous theoretical approaches so far, clearly overestimate the frequency shift of this mode (computed with reference to free betaine (+)). Including the long-range corrections into the B3LYP combination of functionals (i.e. using the CAM-B3LYP functional) leads to even a more pronounced disagreement with the experiment. The semi empirical DFTB-A method, however, seems to outperform the other two.

Tagging the betaine(+) molecule with molecular hydrogen from the charged group side, on the other hand, leads to a very small blue-shift of the O-H stretching frequency of the betaine(+) intramolecular O-H oscillator. While all theoretical approaches implemented in the present study predict correctly the sign of this shift, the quantitative agreement between the two advanced DFT methods (B3LYP and CAM-B3LYP) and the experiment is much poorer than that of semi-empirical DFTB-A.

The previously outlined results clearly indicate that tagging with molecular hydrogen in photodissociation techniques can lead to notable perturbation of the intramolecular vibrational force field of the tagged molecular system. This is particularly pronounced in the case of typical vibrational chromophores, such as the O-H group.

Rigorous quantum mechanical calculations, including the vibrational anharmonicity of the relevant mode predict the correct trend of the vibrational frequency shifts upon tagging at different molecular sites. However, agreement is not quantitative and the corrections to the long-range behavior of the B3LYP functional by the Coulomb attenuation method does not improve the results.

Further efforts will therefore be directed towards achieving a quantitative agreement with the available IRMPD spectroscopic data. Their current studies proceed towards inclusion of dispersion corrections to the B3LYP functional (mainly by the D3 version of Grimme's dispersion with Becke-Johnson damping [21]), as well as testing of some other specifically developed functionals (such as e.g. different HCTH functionals by the Handy's group [19, 20]).

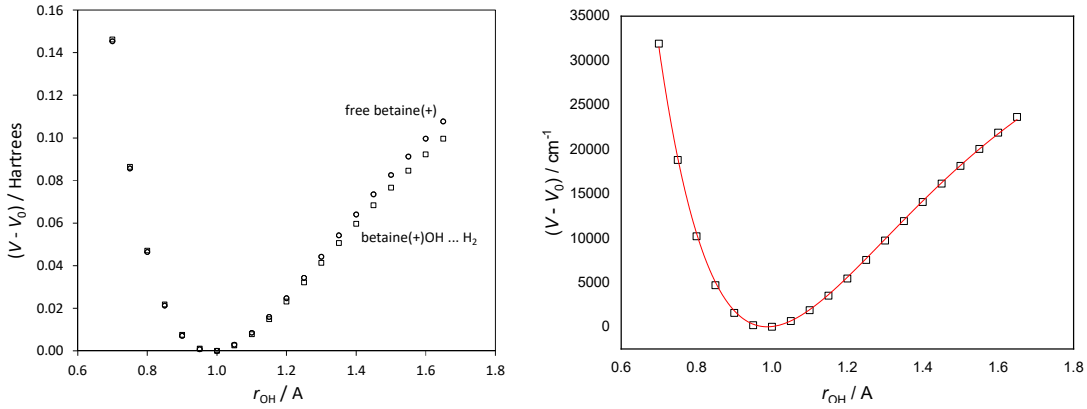

Figure 26: *Left Panel*: Vibrational energy curves for free betaine(+) as well as for the betaine(+) … H_2 dimer where tagging is realized on the O-H oscillator side. *Right Panel:* A typical fit of the O-H stretching potential energy function with Morse function of the form (5)

5.7. Overall Impression for Schrödinger

Since FINKI-cloud was extensively evaluated throughout the life sciences use-case, and for this use case it was only used indirectly through Schrödinger API, the efforts of this investigation focused mostly on the thematic service. Closing this use case, we provide the overall impression of the scientific team which conducted in investigation. The RESTful web service Schrödinger API is, already at present stage of development, a user-friendly platform that excellently serves its purpose – automated solution of multidimensional Schrödinger equation. The web service could be easily integrated into more complex codes in which the Hamiltonian spectra need to be computed for a wide variety of systems. Direct usage of the REST API methods is enabled directly through the web access bar, which is an excellent advantage. Alternative way of its incorporation into more complex sequential algorithms is by consuming a particular method/approach in the user's source code. As yet another alternative, it can be also used interactively through a web browser.

Further improvements of the service should perhaps involve extension of the DVR approaches (at present, only the Hermite DVR methodology has been implemented), as well as the available analytic potentials in 1, 2 and 3 spatial dimensions that have been used for illustration of the approach. Of course, the service could be also adapted to enable a direct solution of the multidimensional Schrödinger equation on the basis of numerical specification of the potential energy part of the Hamiltonian (without its analytical specification).

6. RePol Use Case

6.1. Using RePol for creating the repository policies

In addition to testing the thematic services, generic services as well as the repositories on-boarded and to be on-boarded in NI4OS-Europe, an opportunity of testing an ORDM tool which has been developed within NI4OS-Europe showed up. Namely, The Cyprus Institute, took the decision of setting up an institutional repository. Members of the responsible committee attended a number of NI4OS-Europe training events on RePol (Repository Policy Generator) and realised that policies of their repository can be created using RePol. Hence, they used RePol and provided the following report. The members who conducted the use-case of RePol are:

- Dr. Andreas Gavrielides (CYI)
- Dr. Dimitrios Katsouras (CYI)
- Dr. Michalis Papadopoulos (CYI)

6.2. What is the RePol

A trustworthy repository should have a transparent policy, informing users about the roles, responsibilities, rights and procedures aimed at ensuring that their deposited data are preserved and disseminated in line with the FAIR principles. RePol online tool uses a lightweight online form to guide you through the process of defining a repository policy. By choosing options in the form, you choose sets of predefined policy clauses formulated in line with the current best practice. The resulting policy document may be downloaded, additionally customized, and integrated into your repository. With RePol one can proceed with the following actions:

- Access, data entry and document creation – You can access RePol through a web browser, without any authentication. Data entry starts with the selection of a form for a document that is to be generated. Once all fields in the form are properly filled, the document can be generated. The same fields are shared across several forms.
- Editing of generated documents – The generated HTML document is a draft that must be read and edited before it is considered final. The sections that must be revised are clearly marked. If you want to re-upload and re-use the document in RePol, please refrain from deleting or altering the portion between tags because that is where the machine-readable data is stored.
- Saving all forms – At any time, you can download an XML (standalone) file containing all entered data – the latest values of the input elements, from all forms at once. Each generated document also contains its data in a machine-readable format.
- Later changes of data and documents – Both standalone XML exports and generated documents can be re-uploaded to the RePol and their data parsed to fill the appropriate input elements in all forms that have them, allowing to update the data or the same data to be re-used in other templates or newer versions of the

existing ones. All user-made customisations of the earlier created text will have to be repeated in the new document.

The baseline implementation of RePol is fully operational and is available at https://repol.ni4os.eu. The layout of the repol tool is presented in Figure 27.

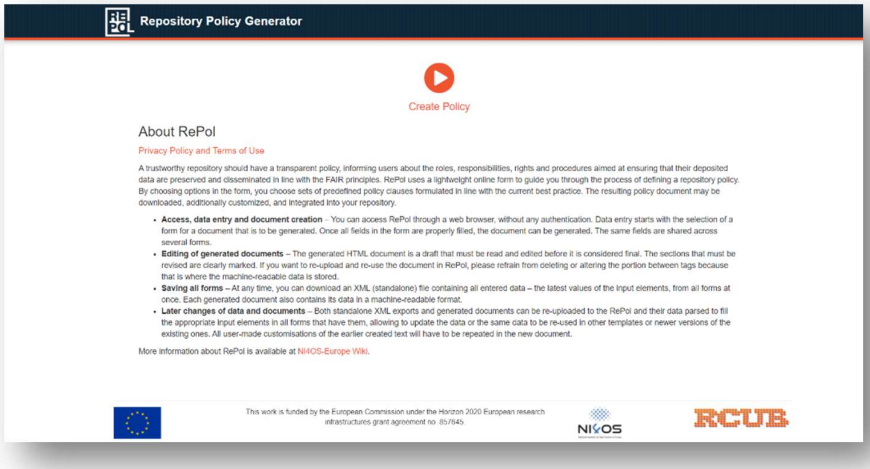

Figure 27: The layout of the Repository Policy Generator

6.3. RePol Report

The Cyprus Institute, as a partner of NI4OS-Europe and as a proud supporter of Open Science, has proceeded in designing a local repository for its researchers and research support personnel to deposit their documents and general data for public access. We adhere to the fact that a trustworthy repository should have a transparent policy, informing users about the roles, responsibilities, rights and procedures aimed at ensuring that their deposited data are preserved and disseminated in line with the FAIR principles.

Being new in the repository management field, we needed help in forming a repository policy. This is where the RePol tool developed in the context of NI4OS-Europe came to save the day. Even though the repository policy is an amalgamation of a variety of policies (including Sherpa), RePol was used as the basis and to introduce us to how a correct repository policy should be constituted.

RePol is an intuitive tool which is one of the reasons that it was chosen by us after attending a training seminar that explained in detail how to use it optimally. The interface is easy to use and only uses minimal input to produce a very comprehensive policy.

The versatility of RePol to adapt to any parameter necessary for a rigorous policy is one of its very strong points. Being amateurs in the repository field we found that we should take provision of certain parameters that we would not be looking for had we started drafting the policy ourselves and thus it would have taken us a lot of unnecessary time to find our footing for drafting a solid policy. This versatility helped us overcome a major

conundrum that we had stemming from the fact that The Cyprus Institute is a heavily multidisciplinary research institute. Each of our four research centres employs scientists that have extremely diverse research interests and backgrounds with their own uniqueness on how they treat data and publications as well as how open science is translated. Some of our fields of research include theoretical physics, archaeology, climate science and modelling, etc.

The Cyprus Institute is also developing its own innovation office centrally as well as research support centres within the Research Centres themselves and thus a repository policy needs to encompass the as open as possible but as closed as necessary philosophy, in which RePol has successfully tackled. A screen shot from the procedure followed in order to create the policies using the RePol is provided in Figure 28.

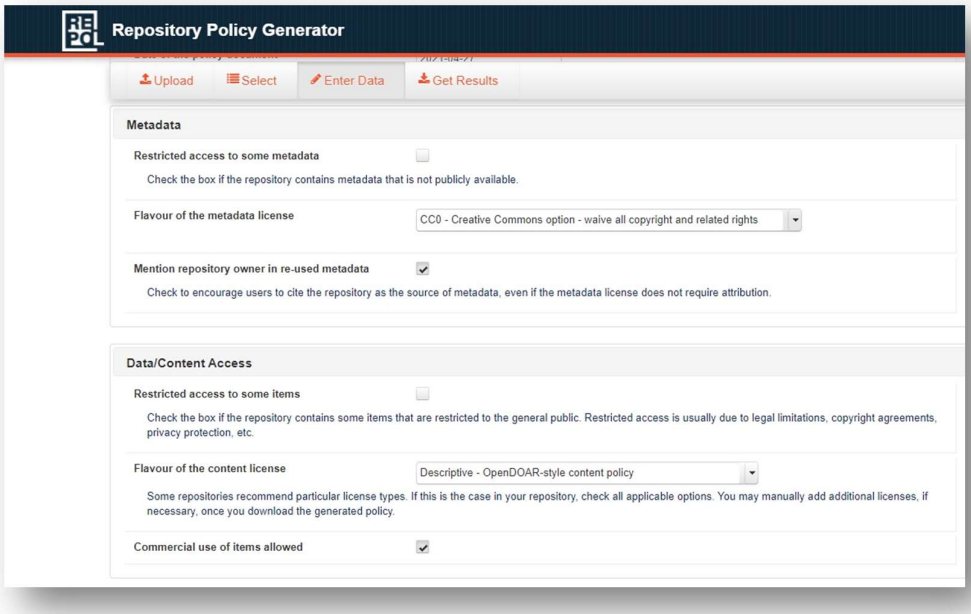

Figure 28: A screenshot from the procedure of creating the repository policies for The Cyprus Institute institutional repository using RePol

Using the RePol tool has not only simplified our task and accelerated the release of our repository but it has also helped us recognise the important points that such a policy needs to be addressed that has in turned led us to take a basic policy and enrich it by deepening to the focal points that are suited to The Cyprus Institute's needs. NI4OS-Europe has thus once again provided us with a tool that is highly recommended to any organisation that wants to either develop or enhance its repository policy and upgrade its open science initiatives integration that is becoming more and more necessary in a continuously evolving European and international landscape. The Cyprus Institute's institutional repository can be accessed with the link: https://repository.cyi.ac.cy/; its layout as this appears on the web is presented in Figure 29. Even though it is accessible it is not fully functional.

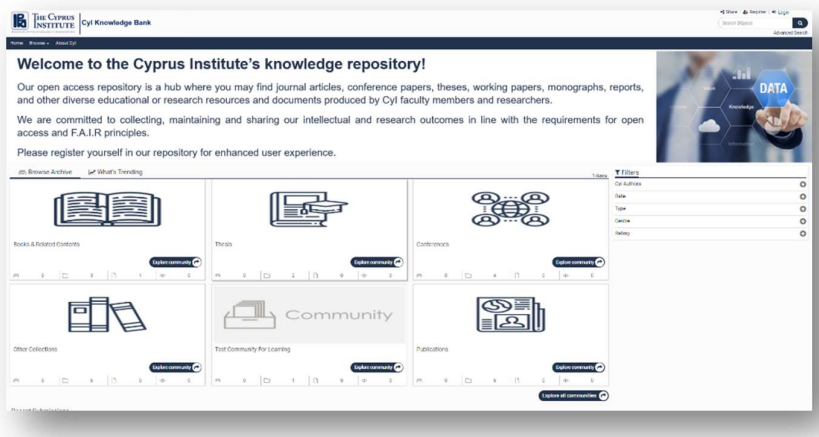

Figure 29: The portal to access the institutional repository of The Cyprus Institute

As a next step, the repository policy formed using RePol as its core component described above, will be sent by the Research, Innovation and Management Support (RIMS) office which serves as a horizontally structured management for the scientific coordinators and project officers, to the upper management of The Cyprus Institute for further review and for any comments that need to be addressed.

After the first comments are made and for any changes are requested, we will use RePol again to give us further guidance to submit a reviewed version before getting the final approval and then releasing the repository policy to The Cyprus Institute community (research and non-research related).

The repository will then be launched on an institute-wide basis after assigning administrative roles to each Centre's/Department's designated research support personnel in order for them to have overview and authorization rights for the repository to be managed more efficiently.

7. Status of the on-boarded services

In this section we will provide a brief overview of the services which have been on-boarded, however, for a complete and detailed description of the status of both on-boarded and non-on-boarded services can be found in Deliverable D5.2, "First report on provider and repository integration". By the term on-boarded services, we mean services described in terms of EOSC resource profiles specification.

As far as repositories go, the NI4OS-Europe repository service provided by the National Infrastructures for Research and Technology, Greece, and the SZTE repository of publications provided and hosted by the University of Szeged, Hungary are fully on-boarded. The NI4OS-Europe repository has been evaluated in the life science use case. In addition, we attempted to also evaluate the later via a survey with no success since researchers of the University of Szeged did not respond to our request. In addition to the above two repositories, twenty-six other repositories are registered within the catalogue, but some aspects of resource description are still under development. The candidate repositories are listed in Table 7, together with the links to the resource on-boarding dashboard page where the up-to- dated details are provided.

Repository	On-boarding dashboard page
Repository of Faculty of Science, University of Zagreb	https://catalogue.ni4os.eu/?_=/resource/id/0eed0440-917e-44a9-a324-aee1d0873744
Meteorological and Hydrological Service of Croatia Repository	https://catalogue.ni4os.eu/?_=/resource/id/33c74095-a1e1-4326-9511-77a794a39489
Repository of the Institute of Public Finance, Zagreb	https://catalogue.ni4os.eu/?_=/resource/id/3d3a3b00-0936-44f6-b01a-bfcda9b0145d
HELIX Data	https://catalogue.ni4os.eu/?_=/resource/id/4c7896d0-0945-42d0-bdf7-c31a090f00ae
UKIM Repository	https://catalogue.ni4os.eu/?_=/resource/id/ee7096b3-6aa7-468e-accd-e6dcad722678
NaRDuS - National Repository of Dissertations in Serbia	https://catalogue.ni4os.eu/?_=/resource/id/01426fe3-8783-47f2-97e6-757bcd70e1be
RIFDT - Repository of the Institute for Philosophy and Social Theory	https://catalogue.ni4os.eu/?_=/resource/id/1ff579c7-2cdd-484a-8eef-afed0d60db87
CHERRY - CHEmistry RepositoRY	https://catalogue.ni4os.eu/?_=/resource/id/22b44397-6edc-4c65-81f5-b165ba5085ab
Institutional Repository of Economic Knowledge	https://catalogue.ni4os.eu/?_=/resource/id/252b6aca-616a-4a3f-9ee8-c481b1d62b79
RIStocar - Repository of the Institute of Animal Husbandry	https://catalogue.ni4os.eu/?_=/resource/id/38659100-c15b-4416-905a-758d4ac0c763
RADaR - Digital Repository of Archived Publications of the Institute for Biological Research	https://catalogue.ni4os.eu/?_=/resource/id/4c9ae669-8b1f-48e0-831c-27f77481004a
RIK - Repository of the Maize Research Institute Zemun Polje	https://catalogue.ni4os.eu/?_=/resource/id/5250b151-2cc5-4714-9151-e1f1d10c14fd
RAF - Repository of the Faculty of Architecture	https://catalogue.ni4os.eu/?_=/resource/id/5585e9a5-34ad-4ac8-b037-b99c1009975e
Jakov - Repository of the University of Criminal Investigation and Police Studies	https://catalogue.ni4os.eu/?_=/resource/id/62acbf7b-1b98-4347-8fdb-eda0e9804492

Repository	On-boarding dashboard page
Armenian NOAD	https://catalogue.ni4os.eu/?_=/resource/id/6e109cc5-7d0f-42b6-924f-e55fbb6262fa
RAUmPlan - Repository of Architecture, Urbanism and Planning	https://catalogue.ni4os.eu/?_=/resource/id/71473eb3-8ae3-444e-af9b-ea083236f6bf
RhinoSec - Repository of the Faculty of Security Studies	https://catalogue.ni4os.eu/?_=/resource/id/7fd00650-d0d3-4784-8d1b-4da74c981322
FarFar - Repository of the Faculty of Pharmacy	https://catalogue.ni4os.eu/?_=/resource/id/806cdfd8-ce66-48a6-bcec-0e316260d2a0
Georgian Integrated Library Information System Consortium 2017	https://catalogue.ni4os.eu/?_=/resource/id/833d08c3-3ab8-4502-9f29-b8d3e0edfa48
DAIS - Digital Archive of the Serbian Academy of Sciences and Arts	https://catalogue.ni4os.eu/?_=/resource/id/9b7a3e3c-7160-4166-b13c-79c9d323cb6f
PlantaRum – Repository of the Institute for Plant Protection and Environment	https://catalogue.ni4os.eu/?_=/resource/id/a1ccf5c4-edcf-471d-99cc-38d939277dfb
CeR - Central Repository of the Institute of Chemistry, Technology and Metallurgy	https://catalogue.ni4os.eu/?_=/resource/id/c5742170-b800-48f2-9c00-57d573a48ceb
SMILE – School of dental Medicine dIgitaL archivE	https://catalogue.ni4os.eu/?_=/resource/id/c9646c31-f6e1-4a10-9fad-346e6f5db4e6
GraFar - Repository of the Faculty of Civil Engineering	https://catalogue.ni4os.eu/?_=/resource/id/ccf1af79-a7af-4d99-923d-f330207ec1f2
Veterinar – Repository of the Faculty of Veterinary Medicine	https://catalogue.ni4os.eu/?_=/resource/id/cea37a6e-4d61-43f8-b1e0-5d87d9179d2b
VinaR - Repository of the Vinča Institute of Nuclear Sciences	https://catalogue.ni4os.eu/?_=/resource/id/e805372a-9b44-4aae-ba06-eca32294a141

Table 7: On-boarding repository candidates registered within the Agora catalogue

Thematic Service	On-boarding dashboard page
Live Access Server	https://catalogue.ni4os.eu/?_=/resource/id/3d243e51-9224-4ae3-810b-9d3c5a311193
Clowder4DCH	https://catalogue.ni4os.eu/?_=/resource/id/d0c2baae-45dc-496f-922c-dcb051ce5fc0
DREAM	https://catalogue.ni4os.eu/?_=/resource/id/4799d746-e9cd-4a61-be59-2e0a2aa88600
Airpolution prediction	https://catalogue.ni4os.eu/?_=/resource/id/ce2f1db4-c0d2-436c-bb45-ba9ca04fb625
OMApp	https://catalogue.ni4os.eu/?_=/resource/id/7919d83d-a02d-4b39-87cb-a1db40419f06
IoT Cloud Platform	https://catalogue.ni4os.eu/?_=/resource/id/923fb0a9-253f-4ddc-a71c-3fb1dbf017c7
High Energy Solar Particle Events foRecastIng and Analysis	https://catalogue.ni4os.eu/?_=/resource/id/c41e1ab7-bf1e-488b-bb00-c0edc0b844e2
MelGene	https://catalogue.ni4os.eu/?_=/resource/id/04b06b6f-e3a1-490b-94ea-8a1ab03092

Thematic Service	On-boarding dashboard page
DICOM Network	https://catalogue.ni4os.eu/?_=/resource/id/0b702315-d082-4ab8-a1ad-3ce7b0d8c2e6
Reduce and Visualize Gene Ontology	https://catalogue.ni4os.eu/?_=/resource/id/6d231512-984c-4fe9-933a-842f9d4c9fbe
IoT Cloud Platform	https://catalogue.ni4os.eu/?_=/resource/id/923fb0a9-253f-4ddc-a71c-3fb1dbf017c7
LOCKDOWN SCENARIOS TOOL based on ATVBG-SEIR Model	https://catalogue.ni4os.eu/?_=/resource/id/97065b62-e5ed-4b96-8a3c-d412e2647c75
Atlas of prokaryotic traits	https://catalogue.ni4os.eu/?_=/resource/id/a5eac7b9-217a-4f6c-a8a4-2e98ad94dcde

Table 8: On-boarding thematic service candidates registered within the Agora catalogue

As far as thematic services go, within the Agora catalogue, 19 thematic services have been registered, 6 of them are marked as on-boarded, while the other 13 as candidates. The candidates together with the links to the on-boarding dashboard page are listed in Table 8. The on-boarded services are ChemBioServer, FEPrepare, Schrödinger API which have been evaluated as well as NanoCrystal, DREAMM and Gaussian API.

As far as generic services go, 4 generic services have been fully described in terms of EOSC profiles specification, while an additional 12 are partially described. So, in total 16 generic services are registered within the project's catalogue. The 4 generic services which are fully on-boarded are Data analysis service or PARADOX Hadoop, AVITOHOL HPC, PARADOX IV cluster as well as FINKI-cloud; all these four services have been tested by the user communities. Table 9 lists partially on-boarded services which we refer to as candidate services.

Generic Service	On-boarding dashboard page
Simple storage service	https://catalogue.ni4os.eu/?_=/resource/id/986f7994-e45a-4f07-b05f-df40f78a8508
Data discovery service	https://catalogue.ni4os.eu/?_=/resource/id/117c2e24-c2af-47b4-8dcc-7a2892d88e84
AVITOHOL cloud	https://catalogue.ni4os.eu/?_=/resource/id/45ba1dd6-6bc5-4974-b20c-c27dbbb6aa6f
ARIS	https://catalogue.ni4os.eu/?_=/resource/id/fe5d5668-e435-4bae-ac93-b44c7e04f155
Archival service	https://catalogue.ni4os.eu/?_=/resource/id/a40c074d-f507-48b8-9970-8276bc5b0c87
CyI Cloud Facility	https://catalogue.ni4os.eu/?_=/resource/id/43b7bd0a-39e1-47d8-807a-5cf5554ece2e
Cyclone	https://catalogue.ni4os.eu/?_=/resource/id/7ef76b56-0fbc-4bb6-883a-127539b535d6
ONYX	https://catalogue.ni4os.eu/?_=/resource/id/9695ac3f-8169-4646-be7e-372778dec0e5
ICIPRO Cloud	https://catalogue.ni4os.eu/?_=/resource/id/34b5bb82-769b-4fd9-8c67-a5395f6dc184

Generic Service	On-boarding dashboard page
ASNET-AM Cloud	https://catalogue.ni4os.eu/?_=/resource/id/ab063ffd-54bc-4d18-a1c1-b62466e85a9d
RENAM Scientific Cloud	https://catalogue.ni4os.eu/?_=/resource/id/bd4a722c-1d1c-440b-b6ff-10513451d240
RENAM Storage Service	https://catalogue.ni4os.eu/?_=/resource/id/f5ceb838-3c0c-4b79-aa63-97b1caf4664c

Table 9: On-boarding generic service candidates registered within the Agora catalogue

Finally, it is worth mentioning that NI4OS-Europe will continue scouting for services which will enhance the added value of the service catalogue. Such services have been identified and will soon be included to the service catalogue of NI4OS-Europe, Agora.

8. Conclusions

This deliverable has been dedicated to the test of the services which have been on-boarded by NI4OS-Europe as well as services which will be soon on-boarded (candidate services) or will be on-boarded by the end of the project. This deliverable provides the report resulting by all the activities conducted within task T6.2 "Execution of Demonstrators" which served as a continuation of the task T6.1 "Definition of demonstrators". Throughout all these activities ten NI4OS-Europe partners have been mobilized and contributed towards the success of this task. In addition, researchers and stakeholders from other European Countries such as Germany, Norway and Finland as well as from countries all over the world such as USA and Egypt have participated in the use-cases strengthen thus the cross-board character of the services and NI4OS-Europe in general.

The use cases focused in three highly cross-disciplinary scientific areas, those of Life Sciences, Climate Sciences, Digital Cultural Heritage which serve as NI4OS-Europe flagship scientific communities. In addition, a use case in the highly cross-disciplinary as well as inter-disciplinary field of Computational Physics has been included. Finally, we report a use case dedicated to Open Research Data Management and more specifically in creation of repository policies for the establishment of an institutional repository. Each scientific community was directed by a scientific community leader who planned, organised and coordinated the execution of each use case.

Each scientific community leader adopted a different strategy for testing the thematic services related to her/his scientific domain, the generic services as well as the repositories which can be used to simulate a real-life research project. For instance, Professor Zoe Cournia who is leading the Scientific Community of life sciences, decided to set up a team which consisted of junior researchers who went through all the details of the services and tested their on-board properties as well as their technical functionalities. This provided an extremely detailed analysis which is well documented in Section… An alternative approach was that followed by the Climate scientific community leader Professor Theodoros Christoudias who decided to approach the climate community within the region of NI4OS-Europe, through the associate consortium and ask from senior researchers to carry out tests of services by executing tasks related to their research. Although this led to a less detailed investigation of the functionalities of the services, still, it provided a good idea whether the services used have faults and what kind of fine-tuning they require. In addition, this approach led to the mobilisation of the NI4OS-Europe consortium towards the successful execution of the demonstrations. Professor George Artopoulos, the scientific community leader of Digital Cultural Heritage, adopted another approach, that of contacting an international consortium of Digital Cultural Heritage stakeholders and ask anonymously the testing of the services participated in his use case. Moreover, Bojana Koteska, the scientific community leader of Computational Physics adopted a hybrid approach; that of assigning a research project to a particular team to carry it out and at the same time to ask a consortium of researchers working on this topic to anonymously test the services. Finally, a use case resulted mainly by the need of a NI4OS-Europe partner, The Cyprus Institute (CYI), to create an institutional repository. More specifically, CYI members responsible for the creation of the aforementioned repository attended training events where RePol was presented and motivated by the user-friendly nature of the tool, they decided to use it in order to create policies for the repository. This

demonstrates the important role training events play in spreading out ORDM and facilitating it.

It appears that a successful approach for testing a service, especially a thematic one, is if the service provider organizes a team of researchers and assign them the detailed investigation of each single feature of the service. This is a very efficient way of finding out what are the faults, what are the weaknesses and what are the needs to be fine-tuned in a service. As a matter of fact, this is a good strategy of testing a service in order to ensure its technical as well as on-boarding maturity before this is listed as a fully on-boarded service in NI4OS-Europe and EOSC. NI4OS-Europe will adopt this approach.

Testing the services enabled NI4OS-Europe consortium to achieve the following:

- Ensure that the decision of on-boarding (in a full manner) of a service was right or wrong.
- To ensure that fully on-boarded services are not suffering major faults or weaknesses.
- To indicate whether fine-tuning is required for fully on-boarded services.
- To identify what development needs to be done for candidate services before these are fully on-boarded in NI4OS-Europe.
- To identify whether services which are expected to be on-boarded suffer major faults and they should be discarded from the on-boarding workflow.

The overall conclusion regarding the fully on-boarded services is that users are satisfied using them and the services do not suffer any major weaknesses. It also appears that some of the services already on-boarded are providing added value to the NI4OS-Europe service catalogue due to their unique functionalities. Problems of minor importance have been reported and thus WP3 in cooperation with WP5 and WP6 will work closely so that they address all the subtleties reported.

Regarding the candidate services which have been tested, once more, no major weaknesses have been spotted. However, since these are not yet fully on-boarded in NI4OS-Europe service catalogue effort will be invested in contacting the relevant service providers and requiring the fixing/development of the flawed operations that the service evaluators encountered.

Regarding services which will be on-boarded in the future, nothing indicates that any of these services scheduled to be on-boarded should be excluded from our service catalogue. Naturally, all the problems encountered within this effort should be resolved before NI4OS-Europe consortium considers their on-boarding.

This deliverable can serve as a reference for future service providers who can study this manuscript and reach conclusions on how to improve their services and bring them closer to the on-boarding phase.

