H2020-EINFRA-2015-1

VI-SEEM

VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean

Deliverable D6.4

2nd Report of open calls and integration support

Author(s): E. Athanasaki (editor)

Status –Version: Final – d

Date: September 30th, 2018

Distribution - Type: Public

Abstract: Deliverable D6.4 – "2nd Report of open calls and integration support" is a report on the implementation of the 3rd call, the outcome of the review process, and the required efforts for the integration of the selected applications in the VRE. Results of the 2nd call are also included. SME and pay-per-use call are also described.

The VI-SEEM Consortium consists of:

GRNET	Coordinating Contractor	Greece
CYI	Contractor	Cyprus
IICT-BAS	Contractor	Bulgaria
IPB	Contractor	Serbia
NIIF	Contractor	Hungary
UVT	Contractor	Romania
UPT	Contractor	Albania
UNI BL	Contractor	Bosnia-Herzegovina
UKIM	Contractor	FYR of Macedonia
UOM	Contractor	Montenegro
RENAM	Contractor	Moldova (Republic of)
IIAP-NAS-RA	Contractor	Armenia

GRENA	Contractor	Georgia
BA	Contractor	Egypt
IUCC	Contractor	Israel
SESAME	Contractor	Jordan

The VI-SEEM project is funded by the European Commission under the Horizon 2020 e-Infrastructures grant agreement no. 675121.

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
September 10th, 2018	a	Evangelia Athanasaki	TOC and initial contents provided
September 21st, 2018	b	Dusan Vudragovic, Andreas Athenodorou	Expanded the required information
September 28th, 2018	c	Ognjen Prnjat	Review and check
September 30th, 2018	d	Evangelia Athanasaki	Final version

Table of contents

References

[1] Project VI-SEEM-675121 - Annex I - Description of the Action

List of Figures

List of Tables

Glossary

API	Application Programming Interface
CBIR	Content-Based Image Retrieval
CORDEX	Coordinated Regional Downscaling Experiment
CRYSTAL	Computational tool for solid state chemistry and physics
DICOM	Digital Imaging and Communications in Medicine
ECHAM	Global Climate Model developed by the Max Planck Institute for Meteorology
EMAC	ECHAM/MESSy Atmospheric Chemistry
Ferret	Interactive Computer Visualization and Analysis environment
GrADS	Grid Analysis and Display System
GROMACS	Molecular Dynamics Software Toolkit
HPC	High Performance Computing
OCR	Optical Character Recognition
SEEM	South East Europe and Eastern Mediterranean region
SLES	Sodium Lauryl Ether Sulphate
VRE	Virtual Research Environment
VI-SEEM	VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean
WRF	The Weather Research & Forecasting Model

Executive summary

What is the focus of this Deliverable?

The focus of this deliverable is to provide a report on the implementation of the 3rd call for proposals for projects accessing VI-SEEM resources and services, the outcome of the review process and the required efforts for the support of the selected applications. In addition we provide the intermediate results of the 2nd open call – applications are still running.

We provide a detailed description of the applications coming from the three scientific communities of Climate, Digital Cultural Heritage and Life Sciences. Moreover, a detailed flow chart which demonstrates how the calls have been implemented is provided.

What is next in the process to deliver the VI-SEEM results?

This is the final deliverable of WP6, due to the end of the project. The operational management of running applications beyond the project lifetime require self-funded effort by the participating organizations which will be offered according to the call plans.

What are the deliverable contents?

The deliverable "2nd Report of open calls and integration support" provides a description of the implementation of the 3rd call for proposals for projects accessing VI-SEEM resources and services. It consists of all the information relevant to their implementation. The description starts with a short introduction on the basics details of the calls and more specifically of the 3rd call. Subsequently, Section 2 provides a short summary on the adopted procedure and the objectives of the calls. Then, Section 3 provides all the information on the available computational and storage resources as well as services. Section 4 gives the list of all the applications that have been received in the call as well as description of the review process. Section 5 provides the outcome of the review procedures as well as the plans for supporting the deployment and the operation of the applications. Section 6 gives a description of the scope of the selected applications. Section 7 provides a description of the intermediate results obtained in the 2nd Open Call as well as information on the publications, presentations as well as PhD and MCs thesis resulting from this call. Finally, Section 8 provides the conclusion of the deliverable.

Conclusions and recommendations

This deliverable provides details regarding the implementation of the 3rd call for proposals for projects accessing VI-SEEM resources and services. The goal of all three open calls implemented in VI-SEEM project is the capacity building by producing fresh results which will be published as well as by enriching further the VRE portal with datasets and codes.

The calls have been organized according to the outcome of the deliverable D6.1 "Framework for VRE resource and service provision".

The third call enabled 22 scientific applications to get access to the VI-SEEM resources and services. From the applications 6 belong to Climate Research, 8 to Life Sciences, 5 to Digital Cultural Heritage and 3 to Cross-Disciplinary Scientific fields. The SME call enabled 6 scientific applications to get access to the VI-SEEM resources and services. From the applications 2 belong to Climate Research and 4 to Life Sciences. VI-SEEM consortium members assigned service enablers for each participant - the role of which guide researchers to access the allocated resources and services and advise the progression of the projects.

The projects which are awarded resources in this call will have access to VI-SEEM services, infrastructure and the overall VRE well beyond the end of the project: the computational resources will be awarded for a period of 12 months, and access to the VI-SEEM data repository service for up to 2 years, beginning July 2018 (for the 3rd call) and immediately after technical evaluation (for the SME call). Provision of guaranteed service to the selected communities beyond the timeline of the project is ensured by a Memorandum of Understanding (and thus also partially funded by the local providers/governments) which was signed in the spring of 2018. This also supports the maintenance of the core data services (PID, VI-SEEM repository) to serve the 3rd call applications beyond project lifetime and also ensures long term preservation of data, with the guarantees and quality of service offered by the local resources providers. This demonstrates a baseline level of VI-SEEM sustainability.

VI-SEEM has also launched a continuous SME call. This is a "try-before-you-buy" program based on the idea to provide possibility for SMEs / industrial organizations to test the VI-SEEM VRE services free of charge for their purposes before making any decision about possible purchase. Finally, a pay-per-use call has also been launched, offering the possibility for industrial players to buy a full access and professional support for the offered service.

1 Introduction

This deliverable describes the setup and implementation of the 3rd open call and the continuous SME call for proposals for projects accessing VI-SEEM services. These two calls belong to the series of calls for projects that require the use of VI-SEEM services and resources in order to create scientific output relevant to the three communities in the South Eastern and Eastern Mediterranean region, namely Climate Research (CR) community, Life Sciences (LS) community and Digital Cultural Heritage (DCH) community.

The design and implementation of the 3rd as well as the SME calls have been based on the outcome of deliverable D6.1 "Framework for VRE resource and service provision" where the main principles and objectives of the VI-SEEM service and resources access framework have been developed.

The defined framework is being used in the context of the VI-SEEM project for providing access to users of the VI-SEEM services and the associated resources. It provides the foundation for the access policies that are used throughout the duration of the project to ensure the best possible utilization of the VI-SEEM services and resources, serving also as a long-term guide for the provision of Virtual Research Environment (VRE) services.

The VI-SEEM VRE is composed of several types of services ranging from access to multiple types of computational services such as HPC, Cloud and Grid, storage services and underlying resources, as well as software, data, workflow repositories and application specific services. All those services have different characteristics in terms of availability, costs and therefore access modalities.

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

- To make services and resources provided by the VRE available for free to as many as possible researchers from the region of South Eastern Europe and the Eastern Mediterranean (SEEM) and the wider user community (for open and free resources).
- To promote the best and most efficient usage of the underlying infrastructure provided by the VI-SEEM resource provider partners.
- To promote scientific excellence and improve the competitiveness of the researchers in the SEEM region.
- To promote scientific collaboration and exchange of knowhow between the experienced research groups and the less experienced but potentially excellent new research teams of the region.
- To open up the knowledge and data produced in the region to all researchers in Europe and beyond where possible.
- To offer user and application enabling support to user communities which require it.

- To serve the scientific fields of Life Sciences, Climate Research and Digital Cultural heritage which are identified as most relevant for the region.
- To provide the opportunity to researchers of all countries to have access to the services offered by VI-SEEM.

All the above principles have been used to implement the 3rd VI-SEEM and the SME calls.

Section 2 provides a summary of the procedure and the objectives of the calls. Section 3 provides all the information on the available computational and storage resources as well as services. Section 4 gives the list of all the applications that have been received in the two calls as well as description of the review process. Section 5 provides the outcome of the review procedures as well as the plans for supporting the deployment and the operation of the applications. Section 6 gives a description of the scope of the selected applications. Section 7 provides a description of the results obtained in the 2nd Open Call as well as information on the publications, presentations as well as PhD and MCs thesis resulted within this call so far. Finally, Section 8 provides the conclusion of the deliverable.

2 Summary of the process and objectives of the calls

VI-SEEM offers a broad set of generic as well as application-specific services in the region of South-eastern Europe and the Eastern Mediterranean, with special focus on the scientific communities of Life Sciences, Climatology and Digital Cultural Heritage. Such services are in the areas of Compute resource provisioning (HPC, Grid and Cloud), Storage and Data services provisioning, Dataset provisioning, Software and Scientific Workflow provisioning as well as Application Specific service provisioning. These services create a unique Virtual Research Environment, thus improving research productivity and competitiveness on the pan-European level.

The projects which are awarded resources in these calls will have access to VI-SEEM services, infrastructure and the overall VRE well beyond the end of the project. Provision of guaranteed service to the selected communities beyond the timeline of the project is ensured by a Memorandum of Understanding which was signed in the spring of 2018. This also supports the maintenance of the core data services (PID, VI-SEEM repository) to serve the 3rd call applications beyond project lifetime and also ensures long term preservation of data, with the guarantees and quality of service offered by the local resources providers.

2.1 The 3rd open call

The 3rd open call was addressed to scientists and researchers that work in academic and research institutions in the region of South Eastern Europe and the Eastern Mediterranean. More specifically these are (in alphabetical order): Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Cyprus, Egypt, FYR of Macedonia, Georgia, Greece, Hungary, Israel, Jordan, Lebanon, Moldova, Montenegro, Romania, Serbia and Turkey.

The project proposals should address open research topics in specific fields of Life Sciences, Climate research, and Digital Cultural Heritage.

Via this call VI-SEEM opened possibilities for regional scientists from the selected scientific fields to have access to the advanced resources and services that it offers.

The list of services and resources offered by the VI-SEEM VRE can be found at: https://services.vi-seem.eu and in the VI-SEEM VRE at: https://vre.vi-seem.eu.

Access to underlying computational resources could be awarded for a maximum period of 12 months, while access to underlying storage resources was provided for up to 2 years. The third call enabled researchers from selected countries and research fields to obtain access to the advanced services of the VI-SEEM Virtual Research Environment.

2.1.1 Applicable scientific fields

Eligible projects were only the ones that addressed one of the following scientific and/or societal challenges:

In the field of Life Sciences

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

In the field of Climate Research

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

In the field of Digital Cultural Heritage

- DCH Area A: Online services and access to repositories in order to enable studies of the immense cultural heritage assets in the region (e.g., searchable digital libraries; with support of meta-data and OCR for Latin characters).
- DCH Area B: Online visualization tools and data management systems to drive breakthrough contributions to art historical problems (e.g., interactive visualization viewer of RTi files and 3D models with digital libraries integration).
- DCH Area C: Unsupervised feature learning in photogrammetric techniques, data processing for image classification; semantic referencing; and geo-referencing.
- DCH Area D: Architecture, Urban Modelling and Planning.
- DCH Area E: Bioarcheology and Natural Sciences.
- DCH Area F: Material Science.

In the Cross-disciplinary scientific field:

- CD Area A: Data Visualization
- CD Area B: Data Analytics and Processing
- CD Area C: Remote Sensing and Photometric Techniques

or any other combination of above.

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

- Scientific Excellence.
- Innovation potential.
- Scientific and/or social impact of the proposed research.

- The need for usage of the selected services and resources.
- The ability to provide the project results (mainly data sets but also services and software) as services for other future VRE users.
- Maturity and experience of the principal investigator and his/her team in the research field as well as in the using the selected resources and services.
- Feasibility of the project based on a technical evaluation and the availability of resources.
- Potential for the collaboration among scientists in the eligible countries for this call.

The criteria above were used for the evaluation of proposals at the review process.

VI-SEEM aimed at a balanced provision of resources to the whole spectrum of scientific fields between the three target communities that this call addressed, as well as to as many as possible countries in the South-eastern Europe and Easter Mediterranean region.

2.1.2 Eligibility

Eligible applicants (as Principal Investigators) were scientists affiliated with academic or research institutions in the following countries (in alphabetical order): Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Cyprus, Egypt, Former Yugoslav Republic of Macedonia, Georgia, Greece, Hungary, Israel, Jordan, Lebanon, Moldova, Montenegro, Romania, Serbia and Turkey.

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

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

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

2.1.3 Application process

All proposals have been submitted electronically via the VI-SEEM survey tool.

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

Support to applicants was provided during the call: namely, VI-SEEM Access Team was available to answer questions while a call was open.

2.1.4 Important dates

In the following list we provide all the important dates concerning the 3rd call of applications for access to VI-SEEM e-Infrastructure resources and services.

- Opening date: 12th February 2018
- Closing Date: 26th March 2018
- Allocation decision: latest April 2018
- Allocation Start Date of awarded proposals: latest April 2018
- Allocation end date of award: July 2019 for computational projects, July 2020 for some data projects
- Final report from successful projects: May 2020

2.2 The SME Call

VI-SEEM opened a specific continuous Call for Proposals for preparatory/development projects accessing the VI-SEEM services and associated infrastructure specifically for partnerships involving SMEs and academic/research institutions. The call was addressed to consortia of scientists and researchers that work in academic/research institutions and SMEs. Academic/research institutions should be located in the region of South Eastern Europe and Eastern Mediterranean. More specifically countries are (in alphabetical order): Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Cyprus, Egypt, Former Yugoslav Republic of Macedonia, Georgia, Greece, Hungary, Israel, Jordan, Lebanon, Moldova, Montenegro, Romania, Serbia.

The project proposals should address non-proprietary/open research topics in the specific fields of Life Sciences, Climate research, and Digital Cultural Heritage.

Via this call VI-SEEM opened possibilities for SMEs and regional scientists from the selected scientific fields to have access, via joint projects, to the advanced resources and services that it offered. The list of services and resources offered by the VI-SEEM VRE can be found at: https://services.vi-seem.eu and in the VI-SEEM VRE at: https://vre.vi-seem.eu. VI-SEEM Cloud and Grid services are not provided through this call.

Access to underlying computational and storage resources was awarded for a maximum period of 3 months. Access to the VI-SEEM data repository service was granted for up to 12 months.

Access was provided for free for research purposes and as specified in this call. If applicants required extra support for services other than those that were specified in this call they could communicate with individual service providers to negotiate their access method. In case they were interested for paid services, they were encouraged to contact individual service providers or send an email to: vi-seem-pmo@vi-seem.eu.

The continuous call enabled SMEs and researchers from selected countries and applications fields to obtain access to the advanced services of the VI-SEEM Virtual Research Environment.

2.2.1 Applicable scientific fields

Eligible projects were only the ones that addressed one of the following scientific and/or societal challenges:

In the field of Life Sciences

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

In the field of Climate Research

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

In the field of Digital Cultural Heritage

- DCH Area A: Online services and access to repositories in order to enable studies of the immense cultural heritage assets in the region (e.g., searchable digital libraries; with support of meta-data and OCR for Latin characters).
- DCH Area B: Online visualization tools and data management systems to drive breakthrough contributions to art historical problems (e.g., interactive visualization viewer of RTi files and 3D models with digital libraries integration).
- DCH Area C: Unsupervised feature learning in photogrammetric techniques, data processing for image classification; semantic referencing; and geo-referencing.

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

- The need for usage of the selected services and resources.
- The results of the project should be of no direct commercial value and should be publishable.

- Feasibility of the project based on a technical evaluation and the availability of resources.

VI-SEEM aimed at a balanced provision of resources to the whole spectrum of scientific fields between the three target communities that this call addressed, as well as to as many as possible countries in the South-Eastern Europe and Easter Mediterranean region.

2.2.2 Eligibility

Eligible applicants (as Principal Investigators) were scientists affiliated with academic or research institutions in the following countries: (in alphabetical order): Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Cyprus, Egypt, Former Yugoslav Republic of Macedonia, Georgia, Greece, Hungary, Israel, Jordan, Lebanon, Moldova, Montenegro, Romania, Serbia.

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

Industrial partners (SMEs) – preferably from the eligible countries – were required to be participants in the proposals. They could be participating as collaborators, in proposals with academic or research institutions in the eligible countries, and only if the aims and objectives of the project were compatible with the criteria set in this call.

Applicants should commit to using the resources that were allocated to them, as well as to providing reports of their work based on the proposed time schedule. Further to that, users should acknowledge the use of the VI-SEEM VRE services in all publications presenting results obtained from using the allocated resources and should make these results publicly available.

2.2.3 Application process

Proposals could be submitted at any time. The evaluation team provided a final evaluation of the proposals within a month from application submission. The evaluation team had the right to ask the applicant for clarification during the evaluation process.

Applications stopped being accepted at the end of the VI-SEEM project (end of September 2018). The VI-SEEM consortium could terminate this call and stop accepting further applications at any point by publishing this decision in the relevant web address of the call. This could mainly happen if demand for successful proposals exceeded the current resource availability.

2.3 The Pay-per-use Call

VI-SEEM project has also built a bundle catalogue to attract customers coming from the private market/industrial sector. VI-SEEM opened a pay-per-use call where it offers a broad set of generic as well as application-specific services, with special focus on the scientific communities of Life Sciences, Climatology and Digital Cultural Heritage. These services are in the areas of Compute resource provisioning (HPC, Grid and Cloud), Storage and Data services provisioning, Data Set provisioning, Software and Scientific Workflow provisioning as well as Application Specific service provisioning. All services can be purchased electronically via the following link: https://vi-seem.eu/bundle/

The bundle catalogue offers the ability to procure these services. It contains specific bundles for HPC services, IaaS services and application-oriented services. The format of these services makes them attractive and affordable within the existing market. VI-SEEM marketplace gives the opportunity to interested customers to navigate the site, be aware of what these services can offer and at what price, and make their decision. Potential customers can choose the option "purchase order", after which a document informing who the Service Provider is displayed, containing the service description, the monthly charges and the initial terms.

The general terms and conditions under which these services as well as the relevant Service Level Agreement are provided in relevant documents online. All above documents (terms and conditions, SLA, Purchase Order) must be physically signed by both parties (customer and provider) prior to service delivery.

Applications are accepted beyond the lifetime of the VI-SEEM project, in order to potentially contribute to the long-term sustainability of the VI-SEEM offering.

3 Available resources and services

The infrastructure of the VI-SEEM VRE consists of resources of various types – HPC resources – clusters and supercomputers with different hardware architectures, Grid sites, Clouds with possibility to launch virtual machines (VMs) for services and distributed computing, and storage resources with possibility for short and long term storage.

3.1 Computational and storage services available

HPC resources: The HPC resources of the project consist of clusters with low-latency interconnection or supercomputers. Most of the systems are based on CPUs with x86_64 instruction set, some of them equipped with accelerators (GPUs and Intel Xeon Phi coprocessors), but also there are BlueGene/P systems, as well as one based on the Cell processor (PS3 cluster IMAN1-Booster/King). HPC resources are offered by the following countries: Albania, Armenia, Bulgaria, Cyprus, Egypt, FYR of Macedonia, Greece, Hungary, Jordan, Romania and Serbia. In total 15 million CPU core hours, 300 million GPU core hours and 15 million Phi core hours will were provided in this call. 7 out of 15 million CPU core hours were allocated at the Blue Gene supercomputer. Information regarding the HPC resources can be found in the following link:
https://wiki.vi-seem.eu/index.php/Main_Page#HPC_Resources

Cloud resources: The Cloud resources available in the call can be used in two ways. Those clouds that provide the ability to launch VMs with public IPs give the possibility to deploy VRE services for their main or backup/fail-over instance. VMs that possess only private IPs can be used for distributed data processing where necessary. Cloud resources are provided by the following countries: Albania, Armenia, Bosnia and Herzegovina, Bulgaria, Cyprus, FYR of Macedonia, Greece, Israel, Moldova and Romania. In total around 300 VM cores are to be provided in this call. Information on the Cloud resources can be found in the following link:
https://wiki.vi-seem.eu/index.php/Main_Page#Cloud_Resources

Grid resources: The Grid resources, available in this call, are provided mostly from smaller clusters. Grid resources for the VI-SEEM VRE are provided from the following countries: Armenia, Bulgaria, FYR of Macedonia, Georgia, Greece, Moldova, Montenegro and Serbia. Information on the Grid resources can be found in the link:
https://wiki.vi-seem.eu/index.php/Main_Page#Grid_Resources

VI-SEEM Simple Storage service: The VI-SEEM Simple Storage service (VSS) is a secure data storage service provided to VI-SEEM users for storing and sharing research data as well as keeping it synchronized across different computers. Data sharing will be possible with other registered VI-SEEM users or with anyone else by using public links which can be

protected with passwords if needed. Each user will be provided with 50 GB of storage for up to two years from the beginning of its project.

VI-SEEM Repository Service: The main storage service that will allow the users of the VI-SEEM VRE to deposit and share data is the VI-SEEM Repository Service (VRS). This is VI-SEEM main repository for hosting the Regional Community Datasets. It can also be used to host publications and their associated data, as well as software or references to software and workflows used to generate such data and publications. The VRS is also the service designated for storing simplified data formats such as images, videos or others formats suitable also for the general public. Each project is eligible for up to 10 TB of storage for up to two years from the start of the project.

VI-SEEM archival service: Data archiving is the practice of moving data that is no longer being used or are being used in a less frequent fashion into a separate storage service. It is a single set or a collection of historical records specifically selected for longer term retention and future reference. Additionally, data archives contain data that are important for future reference or it is important to preserve them for regulatory and audit purposes. In science, archived data are important for future reference and reproducibility of scientific simulations. Each project will be eligible for storing up to 10 TB in the archival service of the project for up to two years from the start.

VI-SEEM work storage space / local storage and data staging: This service refers to storage space available by the computational resource providers to store temporary data for the purposes of processing them, or for storing results of computations. The service will be available for 12 months from the start of the project. The maximum capacity depends on the service provider.

VI-SEEM data analysis service: This service provides the capability to the VRE users to carefully and efficiently investigate and analyze even very large, unstructured datasets. The VI-SEEM data analysis service is based on Hadoop technology, and provides access to 60 CPU cores, 180 GB of RAM and 5.3 TB of storage in HDFS.

3.2 Application-specific services

3.2.1 Application-specific services for Climate

Live Access Server (LAS)

The Live Access Server is a highly configurable server designed to provide flexible access to geo-referenced scientific data. It can present distributed data sets as a unified virtual data base through the use of DODS networking. Ferret is the default visualization application used by LAS, though other applications (Matlab, IDL, GrADS etc) can also be used.

LAS enables the web user to:

- Visualize data with on-the-fly graphics.

- Request custom subsets of variables in a choice of file formats.
- Access background reference material about the data (metadata).
- Compare (difference) variables from distributed locations.

3.2.2 Application-specific services for Digital Cultural Heritage

VI-SEEM Clowder

Clowder is a research data management system designed to support any data format and multiple research domains. It contains three major extension points: pre-processing, processing and previewing. When new data is added to the system, pre-processing is off-loaded to extraction services for extracting appropriate data and metadata. The extraction services attempt to extract information and run pre-processing steps based on the type of the data, for example to create previews. This raw metadata is presented to the user in the Clowder web interface. Users can upload, download, search, visualize and get various information about these data.

Data in the case of VI-SEEM and more specifically in the field of Digital Cultural Heritage can be of very diverse types and formats.

More specifically users can upload massively (zipped) or individual files of:

- 3D Models: where extractors clean up and prepare for visualization on the platform itself.
- Scanned books and their metadata: OCR algorithms will be used to extract the text in the documents so that users can find books using both metadata information and the book's contents.
- Image, text and sound files and their metadata, organised in collections.
- Advanced documentation data, such as Reflectance Transformation Imaging, and analysis of material properties of structures, works of art and artefacts.

3.2.3 Application-specific services for Life Science

ChemBioServer

ChemBioServer is a web-application for effectively mining and filtering chemical compounds used in drug discovery. ChemBioServer allows for pre-processing of compounds prior to an in silico screen, as well as for post-processing of top-ranked molecules resulting from a docking exercise with the aim to increase the efficiency and the quality of compound selection that will pass to the experimental test phase.

It provides researchers with the ability to:

- Browse and visualize compounds along with their properties.
- Filter chemical compounds for a variety of properties such as steric clashes and toxicity.
- Apply perfect match substructure search.
- Cluster compounds according to their physicochemical properties providing representative compounds for each cluster.
- Build custom compound mining pipelines.

- Quantify through property graphs the top-ranking compounds in drug discovery procedures.

AFMM

AFMM provides an automated platform with which the users can generate parameters for modelling small molecules with Molecular Dynamics simulations. The method used fits the molecular mechanics potential function to both vibrational frequencies and eigenvector projections derived from quantum chemical calculations. The program optimizes an initial parameter set (either pre-existing or using chemically-reasonable estimation) by iteratively changing them until the optimal fit with the reference set is obtained. By implementing a Monte Carlo-like algorithm to vary the parameters, the tedious task of manual parameterization is replaced by an efficient automated procedure. The program is best suited for optimization of small rigid molecules in a well-defined energy minimum, for which the harmonic approximation to the energy surface is appropriate for describing the intra-molecular degrees of freedom.

Due to the abundance of organic molecules, no parameters have been created for the full chemical space. Thus, there is a great need for molecule parameterization before proceeding to Molecular Dynamics calculations. AFMM allows users to access parameters for their Molecular Dynamics simulation of small organic molecules that can be used as drugs or materials.

NANO-Crystal

NANO-Crystal is a web-based tool, is implemented for the construction of spherical nanoparticles of a given radius.

More specifically, the goal is to find the number and the Cartesian coordinates of smaller spheres that fit on the surface of the nanoparticle and visualize the output morphology. The home page (http://nanocrystal.vi-seem.eu/) menu allows two selections for the user:

i. the radius of the nanosphere (nm), and
ii. the radius of smaller spheres (nm), that will cover the surface of the nanoparticle

The program computes the number of smaller spheres that fit on the bigger surface and the user can download their Cartesian coordinates (output format .xyz). The program code is implemented using PHP server-side scripting language, which is embedded into the HTML and CSS code. JQuery, a cross-platform JavaScript library, is also used. For local host of the webpage tool, the Wamp server is used. Moreover, we have developed a crystal computational morphology toolbox for constructing and modelling different crystal nanoparticle shapes. We use computational approaches for computing the macroscopic morphology of any periodic crystal by forming different shapes based on Miller indices and the distance measure from the centre of the crystal and visualizing the resulting crystal. That crystal is a polyhedron that is created as the intersection of multiple polyhedra and

individual planes via the steps that follows. This tool is planned to be imported in the NANO-Crystal webserver within 2017.

This tool enables users to construct spherical nanoparticles as well as different crystal nanoparticle shapes based on Miller indices and the distance measure from the centre of the crystal.

Subtract
Subtract is an online tool that can calculate the volume of a binding site found in a protein. Subtract accepts an atom selection in the form of a PDB file and computes the three-dimensional convex hull of the atoms points with the help of SciPy library. The next step of the algorithm is to compute the volume of the convex hull and the volume of the atoms that are included in the solid based on their van der Waals radii. The subtraction of those two volumes yields the volume of the investigated cavity. The algorithm computes cavity volumes of trajectory frames in parallel for maximum efficiency and speed. It requires minimal usage of memory due to the fact that it follows a buffering strategy of reading file chunks and therefore there is no need to load the entire file into memory. There is a wide support of trajectory formats like Gromacs trajectory files and multi-model PDB files due to its dependency to the MDTraj library.

The measurements are evaluated for statistical significance using Wilcoxon Signed-Rank test and had their null hypothesis rejected (p-value < 0.005). Subtract is a tool that has been created to solve the problem of accurate measurement of the protein binding sites, and works both for crystal structures downloaded from the Protein Data Bank and for protein structures arising from Molecular Dynamics simulations trajectories.

3.2.4 Assignment of resources

Finally, successful applicants get access to all public services, data sets, workflows and codes available in the VI-SEEM VRE – https://vre.vi-seem.eu.

The number of accepted projects depends on the technical and scientific merit of the proposals and the availability of resources. HPC projects were expected to be assigned a maximum of 15 Million core hours. A larger number of requested core hours requires elaborate and well documented justification.

4 Applications and review

4.1 Applications from the 3rd Open Call

In total 22 applications have been received in this call from 8 different countries of the region. 18 of the applications required HPC services, 7 Grid and Cloud service, 16 of storage services, 7 application specific services. The applications underwent technical review and where all the requirements were clarified they were sent for the scientific review. For confidentiality reasons the following tables show only the accepted applications.

Application Acronym	Scientific Discipline	Country	HPC	Grid	Cloud	Simple Storage	Repository	Archive	Data Analysis
ClimData	CR Area A	Bulgaria					Y		
RNA_LUPUS	LS Area C	Greece	Y			Y	Y	Y	
EuGenia	LS Area C	Greece	Y			Y	Y	Y	Y
SARISI	CD Area C	Albania			Y	Y			
BSI	LS Area C	Greece	Y			Y	Y	Y	
CAT-ICE	Other	Serbia	Y			Y	Y	Y	Y
EXMED	CR Area C	Greece	Y						
GQL	LS Area A	Serbia	Y			Y		Y	
CHERE	DCH Area A	BIH			Y	Y			
OPENFOAM	CR Area A	Albania	Y						
AT1R	LS Area A	Greece	Y						
CPI	DCH Area A	Cyprus				Y	Y	Y	Y
CORDEX-FPS Ph2	CR Area A	Greece	Y						
RCM-MENA-CORDEX-II	CR Area A	Cyprus	Y						
RXRa	LS Area A	Greece				Y	Y	Y	
DESPHOTMAT	DCH Area F	Greece	Y	Y	Y	Y			
CDPrSc	Other	Greece	Y			Y	Y		Y
CMMC	LS Area C	Greece			Y	Y	Y	Y	
CDFOLK	DCH Area B	FYRoM	Y		Y	Y	Y		
ArmWRF	CR Area C	Armenia	Y	Y	Y	Y		Y	
GAMMOS	DCH Area F	Armenia	Y						
SEMaCD	LS Area C	Serbia	Y		Y	Y			Y

Table 1 – 3rd Open Call applications and requested computational and storage services

Application Acronym	Scientific Discipline	Country	Live Access Severs	Clowder	ChemBio Server	AFMM	Nano Crystal	Subtract
ClimData	CR Area A	Bulgaria	Y					
RNA_LUPUS	LS Area C	Greece						
EuGenia	LS Area C	Greece						
SARISI	CD Area C	Albania						
BSI	LS Area C	Greece						
CAT-ICE	Other	Serbia						
EXMED	CR Area C	Greece	Y					
GQL	LS Area A	Serbia						
CHERE	DCH Area A	BIH						
OPENFOAM	CR Area A	Albania						
AT1R	LS Area A	Greece			Y	Y		Y
CPI	DCH Area A	Cyprus		Y				
CORDEX-FPS Ph2	CR Area A	Greece						
RCM-MENA-CORDEX-II	CR Area A	Cyprus	Y					
RXRa	LS Area A	Greece			Y			Y
DESPHOTMAT	DCH Area F	Greece						
CDPrSc	Other	Greece						
CMMC	LS Area C	Greece						
CDFOLK	DCH Area B	FYRoM		Y				
ArmWRF	CR Area C	Armenia						
GAMMOS	DCH Area F	Armenia						
SEMaCD	LS Area C	Serbia						

Table 2 – 3rd Open Call applications and requested computational and storage services

4.2 Applications from the SME Open Call

In total 6 applications have been received in this call from 4 different countries of the region. All 6 applications required HPC services, 4 of storage services, 1 application specific services. The applications underwent technical review and where all the requirements were clarified they were sent for the scientific review.

Application Acronym	Scientific Discipline	Country	HPC	Simple Storage	Repository	Archive	Data Analysis	Nano Crystal
AOGD	CR Area B	Bulgaria	Y					
IntelliRent	CR Area B	Bulgaria	Y	Y			Y	
BRFAA-NGS	LS Area C	Greece	Y					
COMP-SS-COMP	LS Area B	FYROM	Y	Y				
MAGIA	LS Area C	Greece	Y	Y				
NPssDNA	LS Area A	Hungary	Y	Y				Y

Table 3 – SME Call applications and requested computational and storage services

5 Selected applications and integration efforts required

5.1 Applications from the 3rd Open Call

In total 22 applications got access to VI-SEEM resources and services. The distribution of applications per country is as follows: Albania: 2, Armenia: 2, Bosnia and Herzegovina: 1, Bulgaria: 1, Cyprus: 2, FYR of Macedonia: 1, Greece: 10, Serbia: 3. In total, 9.3 Million CPU core hours, 0.2 Million GPU card hours, and 1 Million Xeon Phi core hours have been awarded to projects requiring among others access to the HPC services. WP6 manages the provision of resources and services to the successful projects and assigned service enablers for each one of them.

Application Acronym	Scientific Discipline	Country	HPC Resource	Grid Resource	Cloud Resource
ClimData	CR Area A	Bulgaria		GE-01-GRENA	
RNA_LUPUS	LS Area C	Greece	ARIS		
EuGenia	LS Area C	Greece	ARIS		
SARISI	CD Area C	Albania			Avitohol
BSI	LS Area C	Greece	ARIS		
CAT-ICE	Other	Serbia	BA-HPC		
EXMED	CR Area C	Greece	Cy-Tera		
GQL	LS Area A	Serbia	PARADOX		
CHERE	DCH Area A	BIH			MK-04-FINKI_CLOUD
OPENFOAM	CR Area A	Albania	Avitohol		
AT1R	LS Area A	Greece	ARIS		
CPI	DCH Area A	Cyprus			
CORDEX-FPS Ph2	CR Area A	Greece	ARIS		
RCM-MENA-CORDEX-II	CR Area A	Cyprus	Cy-Tera		
RXRa	LS Area A	Greece			
DESPHOTMAT	DCH Area F	Greece	Avitohol	HellasGrid	Okeanos
CDPrSc	Other	Greece	ARIS		
CMMC	LS Area C	Greece			Okeanos
CDFOLK	DCH Area B	FYRoM	InfraGRID		MK-04-FINKI_CLOUD
ArmWRF	CR Area C	Armenia	Armcluster		IIAP Cloud
GAMMOS	DCH Area F	Armenia	NIIFI SC		
SEMaCD	LS Area C	Serbia	PARADOX		

Table 4 – Selected Applications and awarded computational services in the 3rd Call

5.2 Applications from the SME Call

In total 6 applications got access to VI-SEEM resources and services. The distribution of applications per country is as follows: Bulgaria: 2, FYR of Macedonia: 1, Greece: 2, Hungary: 1. In total, 430,000 CPU core hours, 6,500 GPU core hours, 100,000 Xeon Phi-core hours have been awarded to projects requiring among others access to the HPC services. WP6 manages the provision of resources and services to the successful projects and assigned service enablers for each one of them.

Application Acronym	Scientific Discipline	Country	HPC Resource
AOGD	CR Area B	Bulgaria	Avitohol
IntelliRent	CR Area B	Bulgaria	Avitohol
BRFAA-NGS	LS Area C	Greece	ARIS
COMP-SS-COMP	LS Area B	FYROM	InfraGRID
MAGIA	LS Area C	Greece	ARIS
NPssDNA	LS Area A	Hungary	NIIFI SC

Table 5 – Selected Applications and awarded computational services in the SME Call

5.3 Review Process

The proposals underwent a technical review and a scientific review in order to determine the eligibility and suitability of applications for the requested services and systems.

Applications requesting large amounts of HPC resources were also peer reviewed by independent scientific experts from the region taking into account conflict of interest and fairness issues. Reviewers were allowed to ask applicants for clarifications. Applications not requiring HPC resources underwent a more lightweight review from scientific community leaders or other scientists. All reviews where based the criteria set in "Scope and criteria of access" of the call.

The VI-SEEM access committee comprising of the VI-SEEM project technical board prioritized the applications based on the criteria set in "Scope and criteria of access" of the call.

The applicants were notified of the final results of the evaluation. Successful applicants received further details regarding the services and resources and the process to obtain user accounts.

6 Summary of the applications

6.1 Applications from the 3rd Open Call

6.1.1 Applications in Climate research

6.1.1.1 ClimData

Suite of STARDEX and ETCCDI Climate Indices Datasets based on E-OBS and CARPATCLIM Gridded Data
Climate research – Area A

The oncoming climate changes are the biggest challenge the mankind is faced with. The impacts of climate change are manifold and vary regionally, even locally, in their severity. For decades, most analyses of long-term global climate change using observational temperature and precipitation data have focused on changes in mean values. However, immediate damages to humans and their properties as well as the ecosystems, are not obviously caused by gradual changes in these variables but mainly by so-called extreme climate events. The relative rare occurrence of extremes makes it necessary to investigate long data records to determine significant changes in the frequency and intensity of extreme events. There are various methods to investigate extreme events, but the computation and analysis of climate indices (Cis) derived from daily data is probably the most widely used non-parametric approach. In order to detect changes in climate extremes, it is important to develop a set of indices that are statistically robust, cover a wide range of climates, and have a high signal-to-noise ratio. The Cis are numerical indicators, which are carefully designed to encompass magnitude (e.g., hot-day threshold), frequency (e.g., heavy rainfall days) and persistence (e.g., longest dry period) of climate extremes. They include absolute-thresholds indices, percentile-based indices, and indices based on the duration of an event. They are used in several projects on climate change with focus on at different spatial scales, from planetary to continental, regional, national or local scale, as prevailing indicators of changes of the extreme events. As far as many of these studies uses partially pre-existing datasets of CIs, the availability of such databases could facilitate any future work, which relies more or less on Cis-based analysis of the present climate. The objective of the present project is to construct and present to the expert community for barrier-free use a present project is to construct and present to the expert community for barrier-free use a comprehensive suite of climate indices datasets (called ClimData), computed from reliable and up-to-date input data from one side and well elaborated and internationally accepted methodology from other. Hence the importance of assessing trends in climate extremes is often emphasized (see literature), estimations of the magnitude of the trend as well as its statistical significance, are accepted as 'natural' supplement to the Cis-time series. Thus, such information for all indices on seasonal and annual basis, is included in ClimData also.

6.1.1.2 EXMED

Extreme weather events over Mediterranean
Climate research – Area C

Within an environment characterized by constantly changing weather conditions, the study and analysis of extreme weather events is of great importance. Both the frequency and intensity of these kind of events, play a crucial role with many social and financial aspects and implications. Moreover, the question is still open regarding the impacts of climate change on the characteristics of extreme weather phenomena.

The scope of this work lies in the study of extreme weather events based on the numerical analysis of a number of characteristic cases. Furthermore, new methodologies focused on the definition of the evolution, intensity and repeatability of severe weather events will be pursued. The study is centralized in the meteorological parameters of wind speed and wave height, employing state of the art atmospheric and wave prediction models.

The first part of this work comprises of the simulation of characteristic extreme weather events which have taken place in the last decades over Mediterranean Sea. To this end, several sensitivity cases and methodologies will be tested in order to investigate the effects of the different inputs and physical parameterizations. This can include the use of different forcing data, model set up and resolution. In addition, the atmospheric model will be interfaced with the wave model allowing the online communication and exchange of meteorological parameters and utilization of advanced air-sea interaction physical schemes. The goal is to identify the optimum configuration necessary for the proper description and representation in these kind of events. The derived information and findings will be employed for the second stage of the study. Precisely, the maximum number of the documented extreme weather events that have taken place in the Mediterranean Sea area the last thirty years will be simulated. The final target is the application of the Extreme Value Theory in the derived results of wind speed and wave height. This allows the determination of the probability of occurrence in future horizons. Relevant modeling studies are limited and mainly based on simulations of lower resolution.

The results derived from the performed experiments will be further utilized in order to quantify the frequency of occurrence of extreme weather events in the study area, as well as the categorization based on wind speed and wave height magnitudes.

6.1.1.3 OPENFOAM

Wind simulation over rugged terrain with OpenFOAM
Climate research – Area A

The project goals are the analysis of wind flow phenomena over rugged terrain in Albania, the impact of mountains and valleys in wind flows and eddies, important for environmental and green energy activities.

6.1.1.4 CORDEX-FPS Ph2

Convective phenomena at high resolution over Europe and the Mediterranean – Phase 2
Climate research – Area A

Damaging weather events are often associated with extreme convective precipitation. Rapidly ascending motions in areas of moisture convergence where the atmosphere is conditionally unstable lead to the development of strong convective cells. Such cells can develop anywhere in Europe and the Mediterranean, over homogenous and flat terrain, or can develop due to orographic barriers, as well as due to land/sea and/or urban/rural contrasts. A variety of potentially severe consequences arise from convection, including heavy rainfall, flash-floods, short-lived windstorms, hail and/or lightning. Convection will be potentially affected by climate change, intensifying extreme precipitation and possibly influencing large-scale conditions, including atmospheric circulation and stratification. Consequently, the occurrence of convection may be favoured or hindered. This, in turn, could result in important changes in the return periods of precipitation extremes. Within the above context, the proposed research is anticipated to improve significantly our understanding of the mechanisms and the factors that influence the location, intensity and frequency of occurrence of convective precipitation events under changing climatic conditions. Further, the expected results will assist in better constraining estimates of future changes convective extremes and associated processes across a range of climatic regions. The conduction of high-resolution climate simulations, at convection-permitting scales, will also contribute to bridging the gap between regional climate models and impact models (e.g. hydrological, ecosystem models, etc.), while added-value will be provided for the decision-making process through analysis of risk and opportunities associated with changes in extreme convective events.

6.1.1.5 RCM-MENA-CORDEX-II

Regional Climate Modelling for the MENA-CORDEX Phase II
Climate research – Area A

Most global climate change projections suggest that the region that encompasses the Middle East and North Africa (MENA) will be greatly affected by climate change. Substantial increases in temperature and changes in the hydrological cycle will likely induce serious impacts in various societal and economic aspects including human health, agriculture, tourism, energy production and more. However, current global models do not provide the necessary local-scale information needed for the design and implementation of vulnerability assessments and the design of impact strategies. Climate simulations of high spatial resolution are necessary to fully understand the regional and local climate effects and perform impact assessments or design adaptation strategies on a regional or national level. Moreover, despite the general agreement in temperature projections, global models are less certain regarding precipitation changes. This is likely related to the more local nature of

these phenomena that cannot be sufficiently represented from the coarse global models. This type of regional information can be extracted by applying the dynamical downscaling technique on the output of GCMs using limited area models. This study is the continuation of a previous VI-SEEM project. In the previous effort, we have performed regional climate simulations over the Middle East, North Africa and West Asia regions using the CLWRF set of modifications to the version 3.6.1 of WRF. The model was set up in a 50-km horizontal resolution and it was driven by a bias-corrected version of CESM1 earth system model forced by two RCP scenarios (RCP4.5 and RCP8.5).

Using as point of reference our previous work we are planning to use VI-SEEM resources in order to continue our projections of future climate over the region and extent our contribution to the MENA-CORDEX initiative. Since the next phase of CORDEX suggests an increase in the horizontal resolution of regional climate models we will investigate the model's skill in horizontal resolutions between 15 and 25 km. We will use the latest version of the WRF model (3.9.1) and we will apply several optimizations relevant for climatic runs. These include the use adaptive time step options, the exclusion of irrelevant variables in order to reduce the time-consuming I/O processes, the addition of climate diagnostics that are part of the standard CORDEX output protocol and the consideration of aerosols in radiation and microphysics parameterizations. The latter are expected to play a great role in this crossroad of dust and polluted air, while their role in the MENA region has not been extensively studied for climate simulations with WRF, as far as we know. At first stage, we will drive the model with the ERA-Interim reanalysis data for validation and further optimization regarding improved features of the latest model version. This task will help us understand better the model behavior in finer resolutions and prepare us for the downscaling of the next generation of global earth system models that will also contribute to the upcoming IPCC report.

6.1.1.6 ArmWRF

High-resolution convection-permitting modelling in Armenia by Weather Research and Forecasting (WRF) model
Climate research – Area C

The aim of the study is to simulate local strong severe weather events (convective rainfalls, hailstorms, strong wind gusts) over a mountain topography of Armenia through sensitivity experiments with the WRF various parametrization schemes and initial and lateral boundary conditions. Data assimilation is needed to be applied to further improve the simulation results (conventional, satellite and radar data assimilation). The research project is being implemented under the collaboration between the Hydromet Service of Armenia and the Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia.

6.1.2 Applications in Life Sciences

6.1.2.1 RNA LUPUS

Molecular profiling in SLE using RNA-Seq
Life Sciences - Area C

Systemic lupus erythematosus (SLE) is a heterogeneous disease whereby an interplay of environmental, genetic and epigenetic factors leads to perturbation of complex biological networks culminating into diverse clinical phenotypes of varying severity. High throughput methods have allowed an "initial glimpse" into pathogenesis and have laid the foundations for a molecular-based taxonomy for personalized therapy. Based on our experience with the molecular characterization of SLE, a recently completed RNA sequencing analysis of 150 patients, the subsequent sequencing of 800 more patients, and our track- record of "paradigm shift" trials in subsequent sequencing of 800 more patients, and our track-record of "paradigm shift" trials in SLE, we will design gene expression panels and "expression profile"/"clinical trait" correlation matrices for diagnostics, personalized immunotherapy and improved clinical trial design. The utility of SLE research extends beyond its boundaries, by providing unique insights as to how the immune system recognizes self-constituents and maintains its homeostasis, and how gender impacts on disease biology.

6.1.2.2 EuGenia

Multi Level whole genome analyses in population genetics
Life Sciences – Area C

The advent of Next Generation Sequencing (NGS) has revolutionized modern population genetics and evolutionary biology. We are now able to perform analyses on thousands of individuals from a multitude of species, model and non-model organisms. The availability of data is crucial for increasing the accuracy of the results, thus obtaining a deeper insight into the correct evolutionary history of the organisms. Moreover, we now extend our knowledge beyond the level of the species. We are able to see how multiple species interact and affect each other's evolution. Such data availability comes with a computational cost. It is extremely time consuming and often not possible (due to space and RAM requirements) to perform analyses on off-the-shelf personal computers or on small servers. Thus, accessing larger computer infrastructures is absolutely necessary.

The goal of this project is to analyze and provide insights into the evolution of species at multiple levels. We will start with the simpler form of information, i.e., sequence information. The goal is to detect the genomic regions that have evolved under selection and define the selection forces. Also, using sequence information we will infer the demographic model of the population, which is absolutely necessary for the accurate inference of the selection forces. Then, we will move to the gene expression level, asking how sequence evolution affects the evolution of gene expression, and vice versa how gene

expression guides the evolution of sequences via the action of natural selection. From the sequence, and gene expression levels, we will move to gene regulatory networks and pathways in order to understand how evolution affect populations and species in different levels of biological organization.

In conclusion, the project will facilitate the understanding of biological organization and its evolution by providing insights in multiple levels of biological organization, from sequence to gene regulatory networks and pathways. It also extends the population genetics theory to more complex entities, beyond the gene level.

6.1.2.3 BSI

Bio Signature identification: empirical evaluation of bio-signature discovery methods
Life Sciences – Area C

Biological mechanisms are often regulated by a handful of molecular quantities which jointly act in tight coordination. Slight disruptions of the equilibrium between these quantities can lead to several diseases, including cancer. Identifying the exact subset of quantities, a.k.a. biological signatures, associated with the mechanism under study is thus of paramount importance. Machine learning offers promising approaches for identifying biological signatures through the analysis of high-dimensional omics datasets. Thousands of datasets measuring gene expression, micro-RNA, epigenomics, proteomics and metabolomics profiles in several conditions and diseases are currently available via public on-line repositories.

In this project we investigate the effectiveness of statistical and machine learning methods in retrieving biological signature by performing a large scale experimentation on thousands of datasets. The results will provide a strong benchmark for selecting the best method in future studies.

6.1.2.4 GQL

Computer assisted identification of potential anti-tumor drugs: modeling guanine-quadruplex – ligand interaction
Life Sciences – Area A

Guanine-quadruplex represents a non-canonical nucleic acid structure, whose binding with ligands is shown to be potentially significant in antitumor treatment. This project focuses on theoretical analysis of guanine-quadruplex interaction with a series of ligands. Classical molecular dynamics simulations will be used for systems' equilibration and selection of structures that will be subsequently treated with Density Functional Theory. Topological analysis of electron density, including Reduced Density Gradient (for analysis of non-covalent interactions between the binding structures) and Interacting Quantum Atoms/Fragments methodology (which rests on Quantum Theory of Atoms in Molecules and enables determination of nature and strength of atom pair interactions), will be used to

provide insight into guanine-quadruplex – ligand interaction. Identification of dominant energy terms will be performed with the Relative Energy Gradient procedure. We will consider the following ligands: squaraine-based compounds, benzimidazole-carbazole molecules and complexes with transition metals (such as platinum). Most of the proposed ligands were synthesized last year, and to the best of our knowledge have not been theoretically thoroughly investigated. We hope that our results could lead to the identification of efficient G-quadruplex binders.

6.1.2.5 AT1R

Molecular simulations and drug binding examination on AT1R
Life Sciences – Area A

Nowadays, especially in the developed world, cardiovascular disease and hypertension in particular affect the vast majority of the population and can lead to serious and life threatening condition and even death if they remain untreated. This project is aiming to investigate the action of angiotensin II (angII) receptor AT1 (AT1R) antagonists, also known as angiotensin receptor blockers, ARBs with the aid of computational chemistry. The AT1R antagonists belong to the class of sartans, with the prototype losartan and there are plenty of commercially available molecules such as irbesartan, olmesartan, candesartan, valsartan, azilsartan et al. This type of drugs targets the renin-angiotensin system (RAS) and their main application is the treatment of hypertension, diabetes related kidney failure and hyperaemic episodes. It was not until recently, that x-ray crystallography was performed on AT1R and its structure bound to olmesartan is available, with PBD ID: 4ZUD (Zhang et al. J. Biol. Chem. 2015, 290, 29127-29139). We are planning to perform long MD simulations (up to 1jns) in order to study the binding mechanism of the antagonist candesartan, as well as it interactions and behaviour when bound to the drug vessel 2-hp-beta-cyclodextrin. The results will shed light upon the way that the drug interacts with the receptor and cellular membrane which will provide crucial information on the design of optimal drugs and formulations in the future which will improve thousands of lives that suffer from cardiovascular diseases.

6.1.2.6 RXRa

Investigating the effect of the RXRα S427F mutant on the structure and dynamics of the wild-type RXRα
Life Sciences – Area A

Urothelial carcinoma of the bladder is a common malignancy that has been estimated to cause approximately 165,100 deaths per year worldwide (results refer to data collected until 2012). A frequently occurring mutation in bladder cancer is found on the retinoic X receptor alpha (RXRα) protein, a nuclear receptor that mediates transcription when activated. This mutation results from the substitution of a serine to phenylalanine (S427F) in the ligand binding domain (LBD) of RXRα. The tumours where this mutation is found showed increased expression of genes involved in adipogenesis and lipid metabolism, suggesting that the

mutations cause constitutive activation. The aim of the current project is to gain insights into the mechanism of oncogenesis of the S427F RXRα mutant by studying the overall effects of the oncogenic mutation on the structure and dynamics of both the RXRα homotetramers and the RXRα-RARα heterodimers. For this purpose, we used extensive, all-atom Molecular Dynamics (MD) simulations spanning to the microsecond time scale of both the wild type and the mutant proteins. With this approach, provided an atomic-level detailed description of the structural and dynamic behaviour of the mutant protein that may lead to its oncogenic effects and which would be impossible to assess with standard experimental techniques. The simulation of such large biomoleular systems is computationally demanding and produced 4TB of data that we would like to store in the VI-SEEM resources.

6.1.2.7 CMMC

Circular Messenger RNA in Memory formation in Chicks
Life Sciences – Area C

One of the objectives of neuroscience is unravelling the molecular mechanisms underlying memory formation. In this project we will investigate changes in the transcriptional activity during memory formation in domestic chicks. Particularly, we will study the changes in expression for circular mRNA in the neural tissue of chicks before and after visual imprinting. The analysis will be performed on Next Generation Sequencing (NGS) data, which ensure unparalleled precision and breadth of investigation, yet demand huge computational power for processing and analysis.

6.1.2.8 SEMaCD

See Monogenic and Complex Disease Database
Life Sciences – Area C

Characterization of special genotypes/phenotypes of human sub-populations of the region can enable the description of the medical needs in the target area. Such characterization requires a development of a set of interfaces and tools that will structure and enhance access to a reliable information on the patients. Towards the establishment of the e-Infrastructure capable to facilitate research for understanding disease mechanisms in the populations, we have developed a database of biological sample data that contains data for over 6,000 patients with a wide range of diseases - from rare to common diseases, e.g. cystic fibrosis, alpha-1 antitrypsin deficiency, cardiovascular diseases, to several types of cancer. In this project, we will deploy previously (during the integration phase) developed the structure of the database, as well as the framework and interfaces able to facilitate identification of the disease mechanism pathways.

6.1.3 Applications in Digital Cultural Heritage

6.1.3.1 CHERE

Cultural Heritage Repository
Digital Cultural Heritage – Area A

CHERE (Cultural HEritage REpository) is a collection of individual applications and services aimed at simplifying the handling of digital cultural heritage resources. It currently consists of module for repository and management of entries (based on Omeka-S), module for structure-from-motion reconstruction (based on VisualSFM, vlfeat,PMVS/CMVS, SPSR and texrecon) provided as both hosted web service as well as semi-automated Docker container, measurement module and web based module for basic editing of 3D meshes (Meshlab.js). It also provides basic tools to generate web based panoramas from 360 photos (pannellum).

This service significantly lowers the required technical knowledge required to start using digital repositories and conduct structure from motion reconstructions without installation of additional software. It is aimed at museum staff with limited technical know-how but can also be used in standalone version via Docker containers for more technically inclined users. It can generate tiled models of panoramas suitable for web consumption without the restriction of maximum resolution supported by the browser.

This project is the continuation of CHERE from Integration phase 2.

6.1.3.2 CPI

Corpus of Ptolemaic Inscriptions
Digital Cultural Heritage – Area A

The Corpus of Ptolemaic Inscriptions project began in October 2013, with the aim of creating a Corpus of up-to-date editions of the Greek, bilingual and trilingual inscriptions on stone from Ptolemaic Egypt (323-30 BCE), numbering around 450 items, based on material collected and annotated by the late Peter Fraser FBA (1918-2007), who was the leading authority of the 20th century on the history and epigraphy of Ptolemaic Egypt. The project is funded by the Arts and Humanities Research Council.

The project will make available a full corpus of scholarly editions, replacing older publications and other partial collections organised by specific region or theme, and will offer for the first time a full picture of the Greek epigraphy of the Ptolemaic period. The new corpus will give proper weight to the importance of public and private documentation on stone, which, for Egypt, has tended to be overshadowed by the volume of papyrus documents. The corpus will illustrate the ways in which epigraphical modes of public pronouncement and display became important in what was originally a language culture alien to the Greeks, not merely in Greek cities such as Alexandria, Ptolemais and Naukratis, but also in indigenous Egyptian towns In taking its inspiration from the work of Peter Fraser,

who recognised the importance of this material and provided the basis for understanding and exploiting it, the CPI project aims not only to bring to completion the work of a great scholar but also to offer a deeper understanding of the history, culture and society of Ptolemaic Egypt

6.1.3.3 DESPHOTMAT

Design of some novel photonic materials
Digital Cultural Heritage – Area F

The proposed project involves the rational design of graphene derivatives for photonic applications. The motivation of the present proposal is to apply an integrated approach for the systematic study of the linear and nonlinear optical properties (L&NLO) of graphene and functionalized graphene derivatives.

The design will take place by employing an array of computational techniques. The proposed research involves two interconnected steps. The first one involves the development of a set of methods (which define a protocol) for the reasonably accurate computation of the linear and nonlinear optical (L&NLO) properties of graphene derivatives and the second step involves the design/selection of derivatives for photonic applications (such as Optical Limiters, Saturable Absorbers), employing the already defined protocol.

The key parameters for the proposed design are the nonlinear optical (NLO) properties. The increasing demand for faster data processing, storage and distribution can only be fulfilled by ongoing miniaturisation of the basic electronic devices. The scientific innovation potential of the proposed project is connected with the rational design and development of the proposed protocol.

6.1.3.4 CDFOLK

Cultural Diversity Through the Traditional Folklore Costumes
Digital Cultural Heritage – Area B

The preservation of tangible cultural heritage, especially when it is in the form of degradable artifacts such as folk costumes, is of utmost importance for many museums worldwide. Traditional clothing used for weddings and in daily life represent the culture, history and artistry of the groups of people who populate a nation. In the Balkans, layering of intricately decorated and crafted textiles is the customary method of dressing. Innovative forms of presentation of such artifacts, where multiple layers of clothing are applied one over another, are required to achieve better understanding and appreciation of the cultural heritage and traditional motifs and symbols. Museum visitors need this type of presentation but just as important are other interested parties such as students, teachers, historians, and designers. Digitizing this heritage and visualizing it in some highly interactive and engaging form will enable long time preservation, active appreciation and deeper understanding of the cultural significance of these textiles.

6.1.3.5 GAMMOS

Gas adsorption mechanism on metal oxide sensors
Digital Cultural Heritage – Area F

In the modern world, sensors technology is in great demand due to the existence of many varieties of toxic, explosive chemicals in industry and environment. There are various not harmful and hazardous matter vapours (for example H2O2, acetone, toluene, propylene glycol, dimethylformamide and etc.), which have a major role in diverse spheres such as environmental protection, industrial manufacture, medicine. As an illustration, propylene glycol (PG) is an excellent solvent for many organic compounds and is used as an active ingredient in engine coolants and antifreeze, brakes, paints, enamels and varnishes, and in many products as a solvent or surfactant. It can also be found in cosmetics, perfumes, as well as in pharmaceuticals. Acetone and PG have a huge impact on human organs. Taking into account above information, toxic and harmful gases' sensors have a huge application for detecting and continuous monitoring of these gases, in the spheres where they are used.

The detection of such gases in the environment is not only technological problem but also fundamental. One of the fundamental or theoretical problems is connected with the prediction of the materials of the sensor for detecting particular gases. Another theoretical problem is an explanation of adsorption mechanisms of gases on metal oxide. The new approach of simulation will help to solve the problem and theoretically predict the results of experiments, saving time and money for creation new sensors.

6.1.4 Cross-disciplinary applications

6.1.4.1 SARISI

SAR Interferometry for Subsidence Investigation
Cross-disciplinary research – Area C

The project goals are the study of ground displacements and subsidence in western lowlands of Albania (Preadriatic Depression), for a better understanding of factors behind significant sea transgression and loss of lands in parts of Adriatic Sea beaches. The methodology of the study consists in production of differential interferograms and phase differences from two or more ground scattering of satellite radar wave data taken in different days, and identification of terrain objects that serve as persistent scatterers. Phase difference data represent displacements of terrain points both in horizontal and vertical directions, suitable for evaluation of ground subsidence and prediction of landslides. Volume of radar satellite images and related processed data reach hundreds of gigabytes, and exploitation of cloud resources offered by VI-SEEM are suitable for this purpose.

6.1.4.2 CAT-ICE

Cluster Analysis Tools for Interstellar Civilizations Expansion
Cross-disciplinary research – Area A and B

We develop simulation models of the astrobiological evolution of the Milky Way in general and the interstellar civilizations' expansion in particular. This nascent field of numerical astrobiology has already been crowded with loosely constrained models that cover large chunks of the total parameter space. This highlights the need for efficient implementations and parallel computing in reaching the new level of precision and predictability. Our various spatial and temporal numerical models are based on both the cellular automata platform and the N-body-like particle formalism, pieces of which have already successfully been used in modeling Galactic habitability in restricted contexts. Hence, in order to reach the next stage, we have developed further numerical algorithms which encompass a wider range of astrophysical and astrobiological phenomena.

The main scope of the proposed project is twofold:

1. The efficient parallel computing of all classes of models and resulting simulation runs on the best possible resolution of input parameters;
2. Development of the efficient visualization and data analysis algorithms, as well as the development of their parallel computing implementations.

Apart from the obvious scientific benefit of the project, all algorithms and codes will be freely available with detailed documentation in order to provide interested researchers in astrobiological and SETI-related studies the same advantages as enjoyed by users in the cosmological structure formation studies and similar numerical-intensive fields.

6.1.4.3 CDPrSc

Computer Design of Protein-based α-helical Scaffolds
Cross-disciplinary research – DCH Area F and LS Area A

One of the main challenges in the field of Bio-Inspired Materials is to develop a fundamental understanding of protein-based biological structures and their self-assembly mechanisms and consequently to engineer innovative bio-inspired materials for a wide range of applications. In this context, the rational design of protein scaffolds that self-assemble into defined nanostructures is of utmost importance in materials science.

In this project, we model the homodimeric RNA-binding ColE1 Repressor of Primer (Rop) protein, which is a paradigm of a canonical 4-α-helical bundle. Rop consists of four a-helices packed in an antiparallel fashion. Moreover, we study loop-mutants of Rop, which according to experimental findings, present remarkable structural and physicochemical changes.

The computational approach that we propose is based on detailed atomistic Molecular Dynamics simulations. Our main goal is to predict, from first principles, the various structures formed from different biomolecules (proteins) in aqueous solutions. In order to

model such realistic systems, in all-atom detail, it is clear that long time simulations of large systems are required.

We further aim to apply our analysis scheme and computational tools on specific parts of amino acid sequences of proteins which are expected to alter conformational and/or dynamical properties of proteins. The application of this methodology on biological systems will hierarchically link together different levels of description and will provide the opportunity to study realistic systems and at the same time to predict their properties directly from the molecular structure.

A synergy between simulation methodologies and experiments is met in our project where virtual experiments will be followed by the "real production" (in the laboratory) some of the computer designed systems, which have been found appropriate for desired applications.

6.2 Applications from the SME Call

6.2.1 Applications in Climate research

6.2.1.1 AOGD

Assessment of Impact of Fleet Management on Air Quality
Climate research –Area B

The company has been developing software and hardware, related to fleet management. The goal of the project is to assess options for achieving "greener" fleet management, through the use of data about meteorology and air pollution in the big cities. The project will focus on methods for data acquisition, aggregation and processing, which enable data produced from researchers or publicly available from the municipalities to be combined with proprietary data, in order to enable clients to optimise their traffic from the point of view of potentially harmful emissions and impact on the environment within the city limits.

6.2.1.2 IntelliRent

Assessment of Impact of Fleet Management on Air Quality
Climate research –Area B

The project aims to investigate the current and potential impact of air quality on the prices of properties, especially of rental prices. It will combine results from air quality modeling and air quality index assessments, performed within the VI-SEEM project with proprietary data, in order to evaluate the most significant factors and relationships. The focus will be on the big cities Sofia and Plovdiv, which have, on one hand, a developed market for rental properties, and on the other hand, high levels of pollution, especially in certain regions. The project focus will be on using Mahine Learning Techniques in order to obtain these results.

The company has established relationships with investors in rental properties and the results of this investigation should be used as one of the inputs when taking investment decisions.

6.2.2 Applications in Life Sciences

6.2.2.1 BRFAA-NGS

BRFAA_NextGenSequencing
Life Sciences research –Area C

The aim of the project is to establish automatically functioning bioinformatics pipelines that will be used for the detection of germline and somatic variants. We will use a combination of publicly available tools in sequential order and our final goal will be to produce not only an annotated list of variants but also a complete QC report with emphasis in the metrics that are most valuable in both research and clinical setting. The accessibility to a large computational infrastructure will provide the opportunity to test several programs in parallel and through the use of available "gold standard" datasets determine each algorithm's sensitivity and sensibility. The project will also estimate each algorithm's performance at different depths of sequencing. It will also allow to have accurate estimates of the time and computer processing power needed for the different bioinformatics steps which will eventually lead to a better allocation of available resources.

6.2.2.2 COMP-SS-COMP

Computational exploration of the solid-state compatibility between active pharmaceutical ingredients and excipients by quantum chemistry, multivariate statistics and machine learning techniques
Life Sciences research –Area B

The issue concerning compatibility between the key physiologically active component of a pharmaceutical product – the active pharmaceutical ingredient (API) and the excipients involved to produce a solid-state formulation thereof is of fundamental importance in pharmaceutical sciences and engineering. It affects the very effectiveness of the API, and it also tackles the issues of the stability of the formulation itself.

In this project, solid-state compatibility between several active pharmaceutical ingredients (APIs) and excipients that are most frequently used to design either immediate- or sustained-release solid-state dosage formulations is studied by a combined usage of quantum chemistry, multivariate statistics and machine learning techniques. Quantum chemical techniques are implemented to investigate the possibility of reactive intermolecular interactions between APIs and excipient species under realistic conditions (finite temperature, electronic excitation caused e.g. by sample illumination, presence of water etc.). Both static quantum mechanical (QM) methods, based primarily on density functional

theory (DFT), as well as various on-the-fly ab initio molecular dynamics (AIMD) methodologies are developed and implemented to address these issues. AIMD methodologies are based on both Born-Oppenheimer molecular dynamics (BOMD) as well as on atom-centered density matrix propagation scheme (ADMP). Static QM approaches are mainly based on careful detailed analyses of the potential energy hypersurfaces (PESs) of clusters containing API molecules, excipient molecular or ionic species, as well as certain number of water molecules. Special emphasis in this context is put on the location and identification of specific points on the PESs, corresponding to either real minima or to various order saddle points and the possibility for reactive transformation thereon. The influence of various number of water molecules included in the interacting cluster on the potential energy landscapes is addressed in details as well, with an emphasis on the possibility of water-induced reactive events on the explored PESs. To mimic the realistic in-crystal environment within which these transformations are expected to take place, various embedding techniques are implemented by a careful choice of the relevant part of the in-crystal surroundings. On the other hand, analyses of the complete on-the-fly AIMD trajectories of the studied supermolecular clusters, as well as relevant snapshots thereof, enable us to pinpoint at certain alternative reaction pathways (e.g. roaming mechanism), as well as to account for the finite-temperature effects on the possibility of reactive transformations. At the same time, along with the QM methods, also advanced multivariate statistical methods and machine learning techniques are developed and implemented to provide an in-depth understanding of the results from various screening analytical techniques that have been used to assess the solid-state compatibility between APIs and excipients. In this context, most of our efforts are directed towards the analysis of the experimental results obtained by Fourier transform infrared spectroscopy (FTIR) and differential scanning calorimetry (DSC).

6.2.2.3 MAGIA

Mutation cAlling algorithm IntegrAtion
Life Sciences research –Area C

Variations in the DNA sequences of humans (also called mutations) can affect how humans develop diseases and respond to various health-related factors like several chemicals, drugs and the environment. They can also affect the function of human genes and their participation in networks involved in human development and genetic disease formation and progression. Therefore, proper identification of the underlying disease gene mutations can have major implications for diagnostic and therapeutic approaches.

During the last few years, continuous technological advances allow life scientists to obtain high quality information regarding the most fundamental source of our genetic diversity which is our DNA sequence, through a technique called Next Generation Sequencing (NGS). Also, due to the rapid adoption of NGS, the operating costs and turnaround times are constantly dropping, rendering the acquisition of genomes affordable and accessible. At the

same time, the expanding genomic data generation has boosted applied clinical research by discovering new gene-disease associations.

As genetic variations are key mediators towards the realization of personalized medicine approaches, their accurate identification is imperative. To this end, certain NGS techniques such as Targeted Gene, Whole Exome and Whole Genome Sequencing are applied. However, these techniques yield huge amounts of data which require trained bioinformaticians, a compilation of many computationally intensive algorithms and considerable amounts of time spent in the filtering of spurious results and the clinical interpretation of the remaining data.

One of the most fundamental steps in DNA mutation detection is the identification of the genomic areas which are altered with respect to a common reference genome, as this list drives the rest of the clinical interpretation. Although several algorithms exist for the implementation of this step, each one presents its own advantages and disadvantages, leading, researchers, clinicians and bioinformaticians to spend great amounts of valuable time in the evaluation and calibration of several of these algorithms. This is further complicated by the continuous appearance of new algorithms which add to the complexity level, as little effort is devoted into combining multiple computational algorithms towards harnessing the advantages and mitigating the disadvantages of each.

The purpose of the MAGIA project is to systematically combine several DNA mutation detection algorithms by weighting their outcomes based on simulated data, where the result is known a priori and each algorithm can be ranked accordingly. As these algorithms operate on data of large volumes, they are computationally and time expensive. The VI-SEEM infrastructure will mitigate this bottleneck by providing the necessary computational resources to perform the computational greedy simulations as well as the required storage for the intermediate results which will be needed to evaluate the performance of each algorithm and rank appropriately. The expected outcomes of MAGIA are predefined weights which will be used to rank the result of each mutation detection method and applied to a novel computational pipeline which will detect mutations based not only in one algorithm but to an ensemble of algorithms, guaranteeing more accurate results and sparing the time of researchers and clinicians to evaluate and calibrate individual methods.

6.2.2.4 NPssDNA

Atomistic simulations of immune stimulatory single stranded bacterial DNA
Life Sciences research –Area A

Among immune stimulators of microbial origin, oligodeoxynucleotides (ODNs) represent the most advanced potential adjuvants. ODNs are unmethylated single stranded DNA sequences with CpG-motifs, which are able to activate the innate immunity and also responsible for mounting acquired responses by binding to their cognate TLR9 receptors. Adjuvant effects are optimized by maintaining ODNs and vaccine antigens in close proximity, which can be

achieved by loading the immune stimulator and the antigen cargo to an appropriate carrier. We propose inorganic nanoparticles as carriers, such as hydroxy-apatite, which is a bioactive and biocompatible material used in current medical applications. The designed drug delivery system should provide a stable attachment of the ODNs to the carrier HAP surface, while not obstructing the multivalent presentation of the ODN molecules to their cognate receptors. Incorporation of ODN immune stimulators into drug delivery systems requires knowledge of their conformational properties, as it has been observed that dsDNA and ssDNA with strong secondary structure tendencies has stronger binding affinity to HAP surfaces, while for TLR9 binding, it has been suggested that only ssDNA can act as an agonist, while dsDNA containing the CpG motif have greatly reduced affinity. The solution structure of ODN immune stimulators, however, has not been experimentally assessed so far. The project will carry out atomistic molecular dynamics simulations on ODN immune stimulators using various force fields.

6.3 Application support

6.3.1 Service enabler

6.3.1.1 Responsibilities

Service enablers are responsible to implement the smooth integration of services into the VI-SEEM Virtual Research Platform (VRE). Each partner assigned a service enabler who is responsible to coordinate and assist researchers from their own county during the process of the service integration. In particular each service enabler is responsible for the following:

1. Explain the integration procedure to researchers and ensure that the integration follows the agreed timelines as set in the integration plan.
2. Assist researcher's access to the VI-SEEM infrastructure and ensure the smooth initiation of service integration.
3. Provide technical support and address any problems encountered during the integration possibly with the support of other experts from the project. The service enabler together with the SC leader and WP5 leader will assign more experts (service integration team) to the project in case the service enabler does not have the full capacity to assist in all aspects of the project.
4. Ensure that researchers receive all required support from the partners that provide the computing resources.

All service enablers are responsible to inform the WP5 and Task 5.4 leader about the progress of the integration and initiation of the services, and promptly report any problems. All the current service enablers are listed in Table 7**Error! Reference source not found.**.

Country	Name	Contact Information
Greece	Kyriakos Gkinis	kyrginis@admin.grnet.gr
Cyprus	Andreas Athenodorou	a.athenodorou@cyi.ac.cy
Bulgaria	Mariya Durchova	mabs@parallel.bas.bg
Serbia	Petar Jovanovic	petarj@ipb.ac.rs
Hungary	Tamas Maray	vi-seem-support@niif.hu
Romania	Silviu Panica	silviu.panica@e-uvt.ro
Albania	Neki Frasheri	nfrasheri@fti.edu.al
Bosnia and Herzegovina	Mihajlo Savic	badaboom@etfbl.net
FYR of Macedonia	Anastas Mishev	anastas.mishev@finki.ukim.mk
Montenegro	Luka Filipovic	lukaf@ac.me
Moldova	Alexandr Golubev	galex@renam.md
Armenia	Wahi Narsisian	wahi@sci.am
Georgia	Temur Maisuradze	temur@grena.ge
Egypt	Youssef Eldakar	Youssef.Eldakar@bibalx.org
Israel	Zivan Yoash	zivan@iucc.ac.il
Jordan	Salman Matalgah	salman.matalgah@sesame.org.jo

Table 6 – The list of Service Enablers

It is clear that the service enablers play an important role during the integration phases. Nevertheless, service enablers also contribute importantly towards the open calls. In the context of the 3rd open call and the SME call each service enabler is responsible for the following:

1. The service enabler contacts the primary investigator and guides him/her through the process of getting access to the VI-SEEM resources which have been allocated through the review process.
2. For any technical questions regarding the initiation of the project the primary investigator refer to your service enabler.
3. After the primary investigators gain access on the allocated VI-SEEM resources and services, whenever they experience a trouble or issue they first contact the service enabler to address the trouble.

The 3rd open call and the SME call for proposals for projects accessing VI-SEEM resources and services and the service enabler responsibilities at all stages of the call are outlined in Figure 1.

6.3.2 Software

Applications that requested computational time either in HPCs or Cloud or Grid, require the existence of particular software tools which must be installed in the machines. Hence, the

allocation of resources took into account the software requirements for each application and assigned the project according to the software needs to the right machines. In case that a machine requires the installation of a software, this is done with the support of the local contact site of the machine. Lists with the software required for each application is provided in Table 7 and Table 8 for the 3rd open call and the SME call respectively.

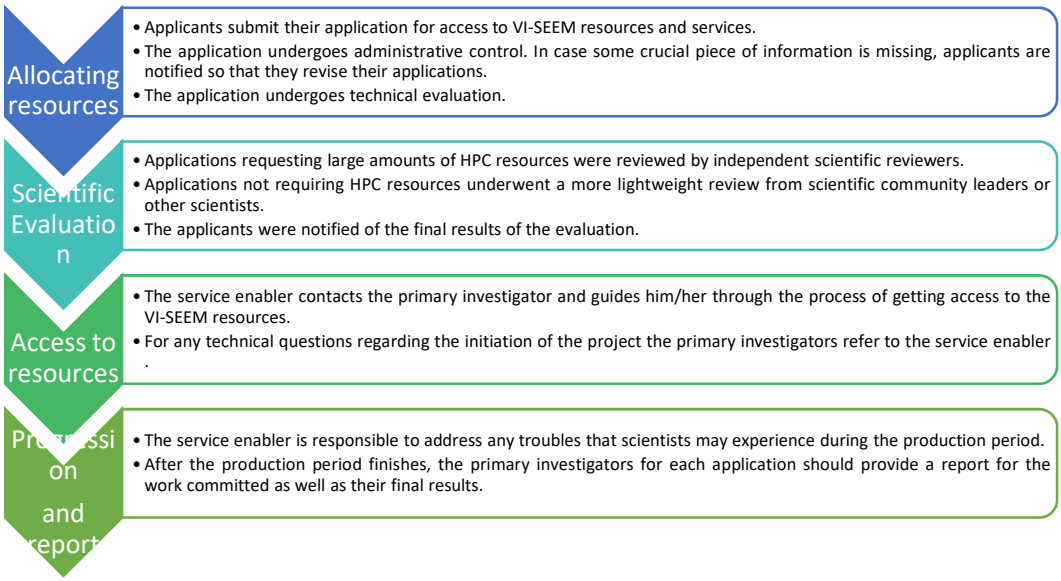


Figure 1 - The flowchart for the integration of a service in the VI-SEEM VRE

Application Acronym	Required Software
ClimData	Software provided by the following organizations and institutes: CARPATCLIM, ECA&D, STARDEX, ETCCDI, Unidata.
RNA_LUPUS	Cutadapt, star, htseq
EuGenia	SweeD, OmegaPlus, EVONET
SARISI	ESA Cloud Tool Box
BSI	Just Add Data, autosklearn, MetaboAnalystR
CAT-ICE	mcav
EXMED	RAMS 6.0 atmospheric model, WAM, OASIS.3-MCT
GQL	CP2K, GAUSSIAN, AIMAll
CHERE	-
OPENFOAM	OpenFOAM, DEM2FOAM
AT1R	Schrondiger Glide/Desmond, gromacs
CPI	-
CORDEX-FPS Ph2	WRF model
RCM-MENA-CORDEX-II	WRF
RXRa	GROMACS

DESPHOTMAT	NWCHEM, DALTON, GAMESS
CDPrSc	GROMACS
CMMC	CIRI, CIRCExplorer, KNIFE
CDFOLK	COLMAP
GAMMOS	Quantum ESPRESSO
SEMaCD	REDCap

Table 7 – The required software for each selected application of the 3rd Open Call

Application Acronym	Required Software
AOGD	Hadoop environment
IntelliRent	LIBSVM
BRFAA-NGS	Bwa, samtools, GATK
COMP-SS-COMP	CPMD, Crystal 17
MAGIA	VarSim, FastQC, NEAT
NPssDNA	AMBER

Table 8 - The required software for each selected application of the SME Open Call

7 Results of the applications

7.1 Applications from the 2nd Open Call

7.1.1 Applications in Climate research

7.1.1.1 KPP_CLIMATE_GPU

KPP/KP4 Climate GPU Acceleration
Climate research – Area A

The outcome of the study was the development of a publicly released source code that automatically ports the kinetics calculations on GPU architectures. Each GPU thread calculates the chemical concentrations of an individual atmospheric grid box, massively parallelising the workload. The achieved performance showed up to 22.4× improvement of the kernel execution time over the old implementation.

One of today's great scientific challenges is to predict how climate will change locally as gas concentrations change over time. The study of chemistry-climate interactions represents an important and, at the same time, demanding task of global Earth system research.

The global climate model EMAC is a modular model that simulates climate change and air quality scenarios. EMAC uses a general-purpose chemical kinetic module to calculate the concentrations and the interactions of chemical species in the atmosphere. Solving atmospheric chemical kinetics is one of the most computationally intensive tasks in atmospheric chemical transport simulations.

To address the computational challenge, an activity was initiated under the VI-SEEM project to accelerate the performance of the chemical kinetics calculations on modern high-performance supercomputers using Graphics Processing Units (GPU).

The VI-SEEM project provided an opportunity to develop and test the software using the Cy-TERA Supercomputer System in Cyprus. Cy-TERA provides computational resources for researchers in the wider South East Europe and Eastern Mediterranean Region with the installation of a hybrid cluster machine of peak performance 30 Teraflops.

The VI-SEEM project provided the necessary e-Infrastructure resources, training and support to develop the building blocks for a new generation of advanced and well-evaluated high-resolution global climate models, capable of simulating and predicting regional climate with unprecedented fidelity.

Vastly improved spatial and temporal simulation accuracy will allow assessment of the impacts of climate change on the environment, human health, agriculture and energy

sectors for the years to come and will enable informed policy planning for adaptation and mitigation.

We developed software that generates CUDA kernels for numerical integration in the global climate model EMAC, used to study climate change and air quality. We focus on atmospheric chemical kinetics, the most computationally intensive task in climate–chemistry simulations. This approach can serve as the basis for hardware acceleration of numerous geoscientific models that rely on KPP for chemical kinetics applications.

The project also granted HPC access to the Juelich Supercomputing Center JURECA supercomputer after successful application.

For the granting period from 1 November 2017 to 31 October 2018 the following quota has been allocated:

- 0.20 million core-h on compute nodes without GPUs
- 0.60 million core-h on compute nodes with GPUs
- 0.20 million core-h on the JURECA booster

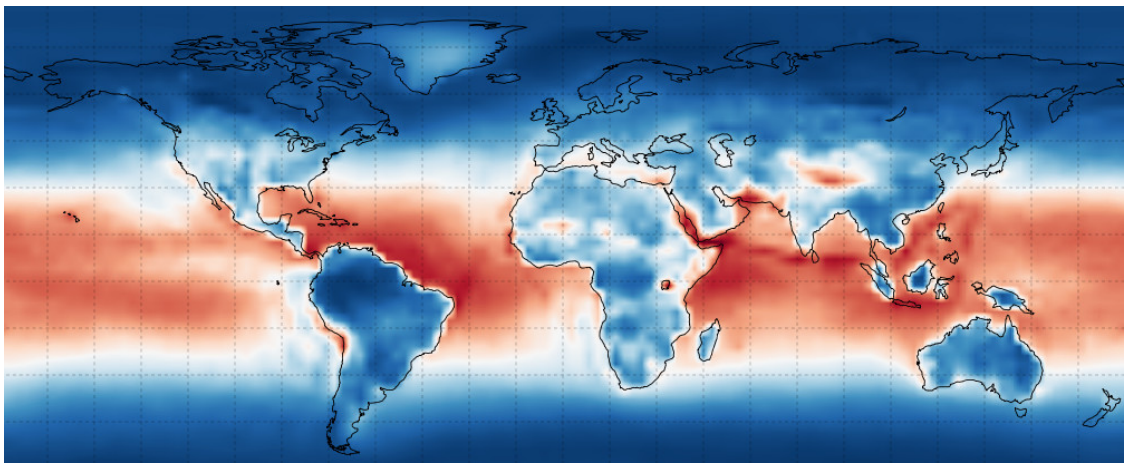


Figure 2 - Hydroxyl radical global surface concentrations simulated on GPU accelerators

7.1.1.2 WheAirCYEM

Weather and air pollution forecasting for Cyprus and Eastern Mediterranean
Climate research – Area C

The aim of this project was to provide meteorological, air quality/chemical weather forecasts, for the region of Eastern Mediterranean, focusing on Cyprus. The forecasts included extreme event (e.g., dust storms) warnings in order to help mitigate the impacts of such weather phenomena to daily activities. It further explored the impact/feedback of air pollution patterns on general weather parameters.

Air pollution in the region of Eastern Mediterranean and Middle East (EMME) has become, over the years, an important issue that affects the everyday lives of millions of people in the area. EMME is a well-known hot-spot of air pollution that is influenced by multiple sources, either of natural or anthropogenic origin. As a consequence of the continual poor air quality conditions the frequency and severity of respiratory related diseases has increased, as well as the number of the associated premature deaths.

The project is concerned with air quality and weather pollution forecasts based on numerical predictions. Air pollution forecasts for the EMME region and Cyprus include early warnings for extreme air pollution episodes, such as dust storms, since the region is located at the center of the so called "dust-belt", which includes two of the main dust sources-Saharan and Arabian deserts. The main objective is to positively contribute to the mitigation of the impacts from such extreme phenomena that will eventually help improve the quality of life over the region. In addition, the impact/feedback of air pollution patterns on general weather parameters will be further investigated once enough simulation data are produced.

The project conducts numerical simulations by employing a regional climate model coupled online with chemistry. Namely, it uses the WRF-Chem model (v.3.6.1 and v.3.9.1.1) that has the capacity to simulate simultaneously physical and chemical processes using nested domains. In particular, the latter feature allows us to focus on specific areas of interest where the simulation of some physical processes requires a finer resolution (e.g., resolving topography and coast lines). For instance, in the current set up the higher resolution domain air quality forecasts are provided for the island of Cyprus at a spatial resolution of about 3km. The model simulations at such high resolutions with online atmospheric chemistry often require large amounts of CPU resources and load. To this end the project is also testing and implementing specially reduced (skeletal) chemical mechanisms, which will further expedite the forecasts, without compromising the final result.

The project is currently in the testing and model tuning phase, however the preliminary results presented below have been greatly benefited from the allocated computational resources on the CyTERA HPC facility.

In the future, the project team will continue the development and refinement of air quality forecasts based on the current work and acquired knowledge.

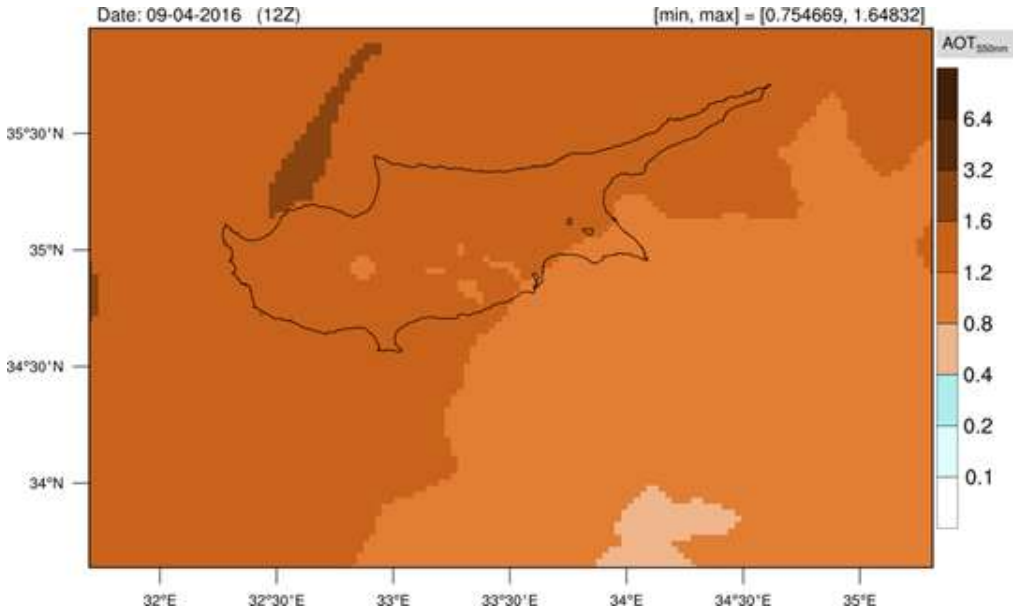


Figure 3 - A snapshot of the Aerosol Optical Thickness (AOT) at 550nm from a simulation (hindcast) of a major dust event that took place between 08-10 of April, 2016 and originated from the Sahara desert. The map depicts the high spatial resolution result over the Cyprus nest.

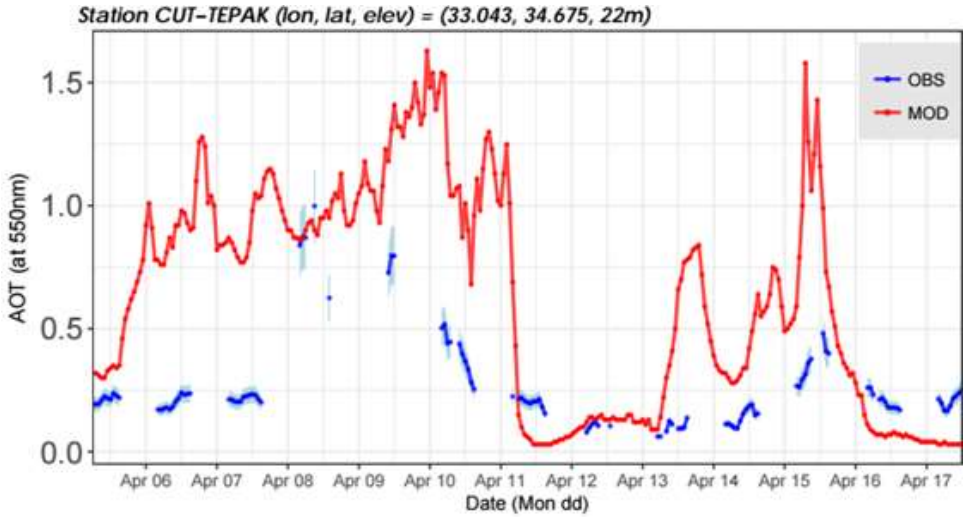


Figure 4 - Time series of the simulated Aerosol Optical Thickness (AOT) at 550nm from a simulation (hindcast) of two dust events that took place between 08-10 and 13-17 of April, 2016 and originated from the Sahara desert, over the nest of Cyprus at high spatial resolution. The simulated results (in red) are compared to some partly available measurements (in blue) at CUT-TEPAK station that are available through the AERONET network. The current model set up reproduces fairly well the two dust episodes, although it has a tendency for overestimation . Further tuning is underway.

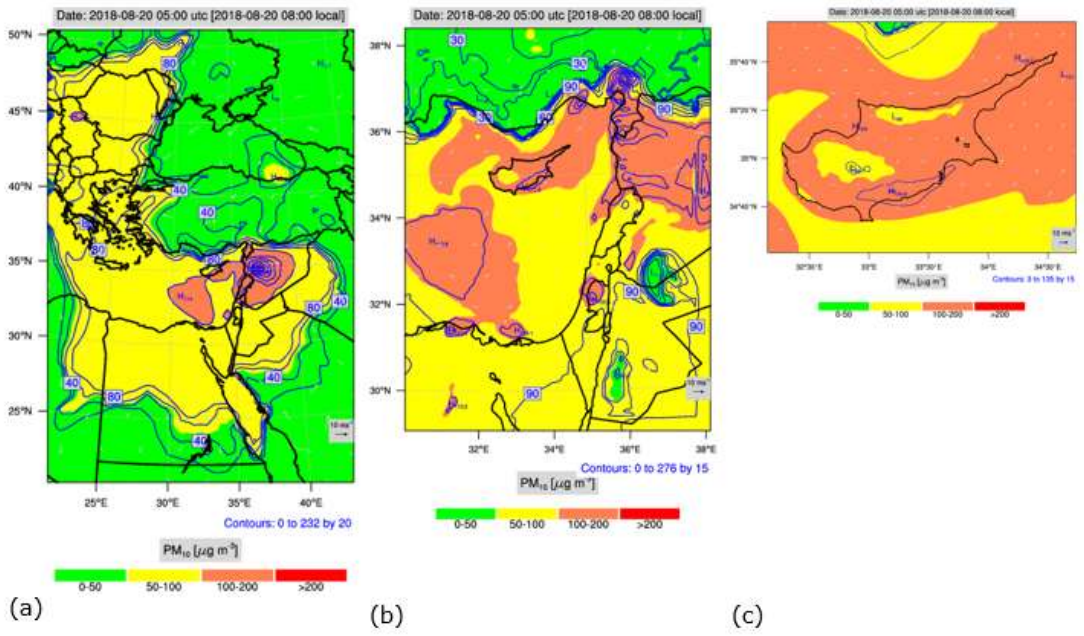


Figure 5 - Near surface PM10 concentration air quality forecast over the EMME region (a)-(b) and over the high resolution nest over the island of Cyprus (c). The contour values are grouped according to the Cyprus Department of Labor Inspection (DLI) classification scheme (See: http://www.airquality.dli.mlsi.gov.cy/). It appears that the model simulations rather overestimate the PM10 concentration over the region. More tests, mainly related to use of boundary conditions and for longer periods are still needed.

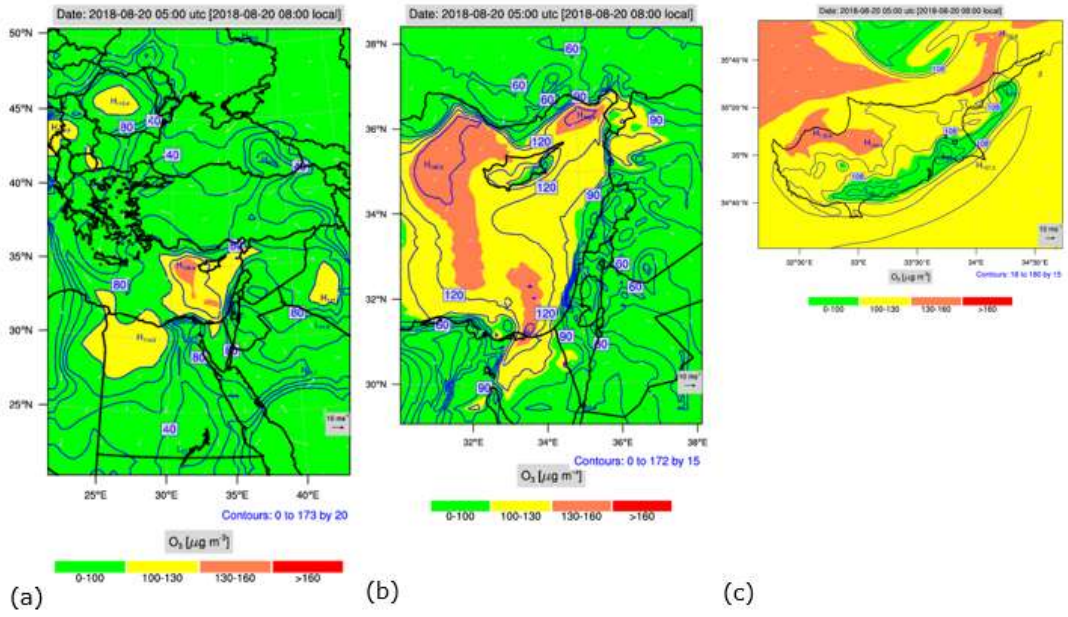
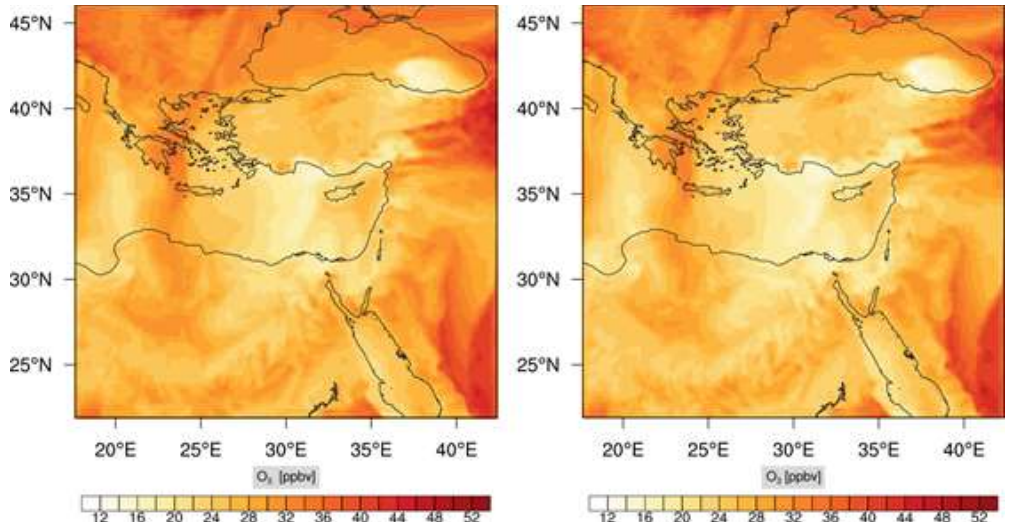


Figure 6 - Near surface Ozone concentration air quality forecast over the EMME region (a)-(b) and over the high resolution nest over the island of Cyprus (c). The contour values are grouped according to the Cyprus Department of Labor Inspection (DLI) classification scheme (See: http://www.airquality.dli.mlsi.gov.cy/).



(a) (b)

Figure 7 - Instantaneous comparison of ozone spatial mixing ratio over the EMME region, averaged over the first nine vertical layers, using: the full mechanism (a) and (b) the reduced/skeletal mechanism. Despite the fact that in (b) we have selectively reduced the number of species in the chemical mechanism, the end result does not deviate significantly from the original result (a). The overall computational speedup using the reduced mechanism approach was over 24%.

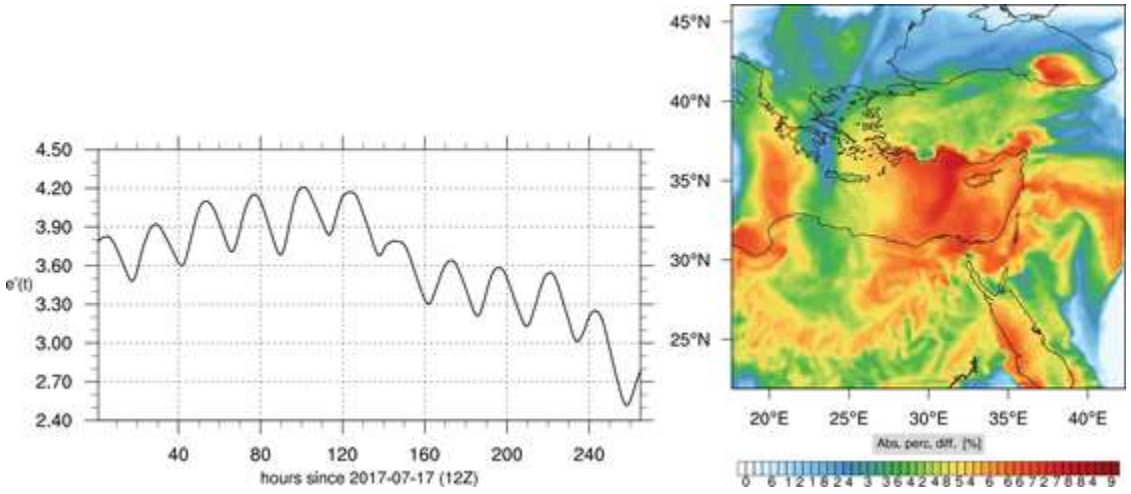


Figure 8 - The line plot on the left (a) shows the volume-weighted average of the absolute percentage difference between the full and reduced/skeletal mechanisms, e'(t), for the ozone mixing ratio. The plot on the right (b) depicts the spatial distribution of the absolute percentage difference between the reduced and the full mechanisms, w.r.t. full mechanism, for the ozone mixing ratio when e'(t) is maximum.

7.1.1.3 3DVAR WRF

Using WRF-3DVAR assimiliation system to improve the weather prediction over the Ararat Valley Region of Armenia
Climate research – Area A

This research project has been implemented by the "Hydrometeorology and Monitoring Service" of the Ministry of Emergency Situations of the Republic of Armenia with the support of the Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia.

The initial conditions have crucial impacts on the weather prediction accuracy of numerical weather prediction models, such as the WRF (Weather Research and Forecasting) model. It is especially essential for complex geographies and large-scale atmospheric circulations, like the territory of Armenia, which contains several large water bodies and significant topography, including Caucasus Mountains in the north and north-east, mountain ranges and plateaus of the Armenian Highland stretching from west to east and the Kura-Araks plain located in the east.

The number of stations located in Armenia, which are involved in the global data exchange, is limited. Only one station provides historical data and monthly updates to the Global Climate Observing System Surface Network and three meteorological stations among 47 stations provide synoptic data through World Meteorological Organization Global Telecommunication System. The study aims to analyze the effect of data assimilation with additional data observations obtained from the 47 weather stations. Assimilation experiments have also been performed to assess the impact of individual surface parameters on the model simulations. Impact of weather station observations on model simulation has been examined regarding to the control simulation and quantified in terms of root-mean-square error and mean error for air temperature (T2) and relative humidity (RH) at three selected stations Armavir, Yerevan, and Areni in Ararat valley. The three-dimensional variational (3DVAR) data assimilation system of the WRF model is planned to be used. Very high spatial resolution is required with application of triple nesting (9 km, 3 km and 1 km) or double nesting (3 km and 1 km) depending on the spatial resolution of a forcing global data (GFS ~28 km, ECMWF ~9km). The parent and nested simulation domains consist of 220-250 grid-points both by latitude and longitude. Five days period from the heat season has been used for the analysis that occurred on the territory of Ararat valley.

The assimilation of additional measurement data from ground stations improved the forecast of the weather parameters (air temperature and relative humidity), although the preliminary results were considered since the analysis was carried out for a short time interval and the period was chosen arbitrarily. The study was conducted on limited parameters. At the same time, the results of the analysis are considered satisfactory, and the project also considers continuing the data assimilation experiment. The results obtained

are preliminary, since the comparison period is very short and it is not possible to draw definite conclusions about the improvement of the weather forecast by including observational data.

Results indicate improvements in the surface air temperature and humidity simulations in the timescale of 24 h at all stations due to additional weather data assimilation into the model. RMSE (root-mean-square error) value is used to identify influence data assimilation. The RMSE value for RH has been changed from 9-13% to 6-8%. Almost all conditions with data assimilation have RMSE values that are smaller than the initial condition without data assimilation. Experiments with using additions data from stations generate the smallest RMSE values for the parameter (air temperature and relative humidity) weather. The data assimilation experiment is considered to be continued with using data from various sources of observations.

It is planned to simulate local severe weather phenomena (air temperature, convective rain, hail, strong gusts) over mountain topography in Armenia using sensitivity experiments with various WRF parameterization schemes and initial and lateral boundary conditions. The main attention will be paid to on the study of drought to improve drought assessment and forecast to the territory of Armenia.
http://meteo.grid.am

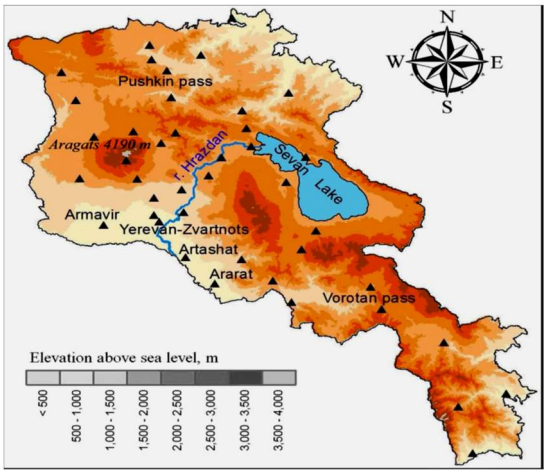


Figure 9 - The distribution of the meteorological stations in Armenia.

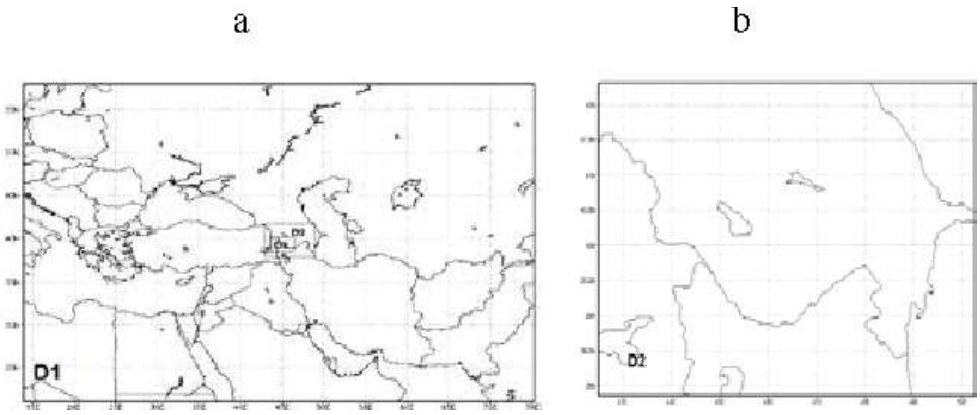


Figure 10 - WRF domains: coarse domain D1 (a) and nested domain D2 (b)

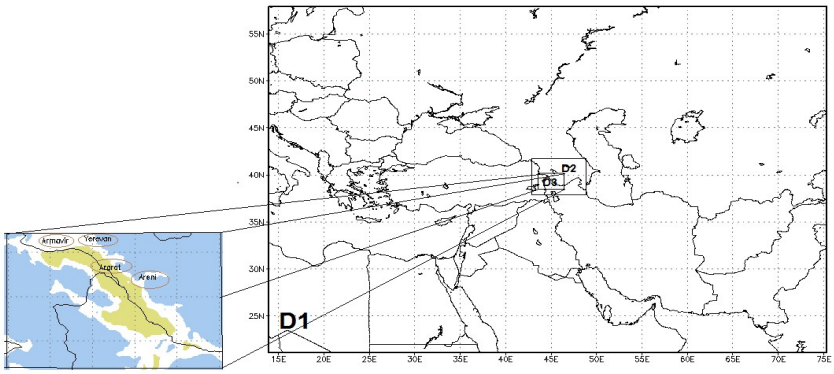


Figure 11 - WPS domain configuration: Domains for the ARW-WRF forecasts.The outer box is the coarse grid with a resolution of 18 km (D1); the inner boxes are the nested grid (D2) with a resolution of 6 km and grid (D3) with a resolution of 2 km.

7.1.1.4 NBBM4RHMS

NBBM Modeling for RHMS
Climate research – Area C

Weather significantly affects the health, safety and economic well-being of all. Climate change and global warming contribute to the more frequent occurrence of weather and hydrological disasters that have intensified intensity and adversely affect people and material goods, and hence to sustainable development. That is why continuing progress with respect to the methods and tools for forecasting the time, monitoring new technologies in development and applying the numerical forecast of time as a starting point for the preparation of timely precise and detailed weather information is of utmost importance. In order to address these climate change challenges, FYRoM, as a member of the newly established SEECOP Operational Numerical Forecast Consortium, with the commitment of the National Hydrometeorological Service and cooperation with FINKI UKIM, is accelerating

its efforts towards improving resources in the area of climate change. The preparation of the establishment and implementation of an integral contemporary prognostic system for time, seasonal and interim forecasts as well as the preparation of climate scenarios future climate in FYRoM. This will be achieved through an integrated approach, using regional climate models or available tools, with the sole aim of strengthening the use of climate information and forecasts for decision-makers in more socio-economic sectors.

Given the resolution (2km horizontally, 80 layers vertically) of the model runs, and the visualized output, there were observed vertically flatter contours of the air temperature which seemed very strange in the beginning. After analyzing the input data and the static data that was used, concluding that these topographic data is very inaccurate, especially mountain tops being ~60% of their actual height. Improving topographic data will drastically improve the accuracy of our simulations.

In the future, the project team is planning to continue the investigations in this direction and currently the different surfactant/protein systems are considered aimed to elucidate how the surfactant operates at molecular level and reveal the mechanism of protein folding in terms of atomic interactions. The aim is to improve the spatial interpolation accuracy and gather more accurate weather data.

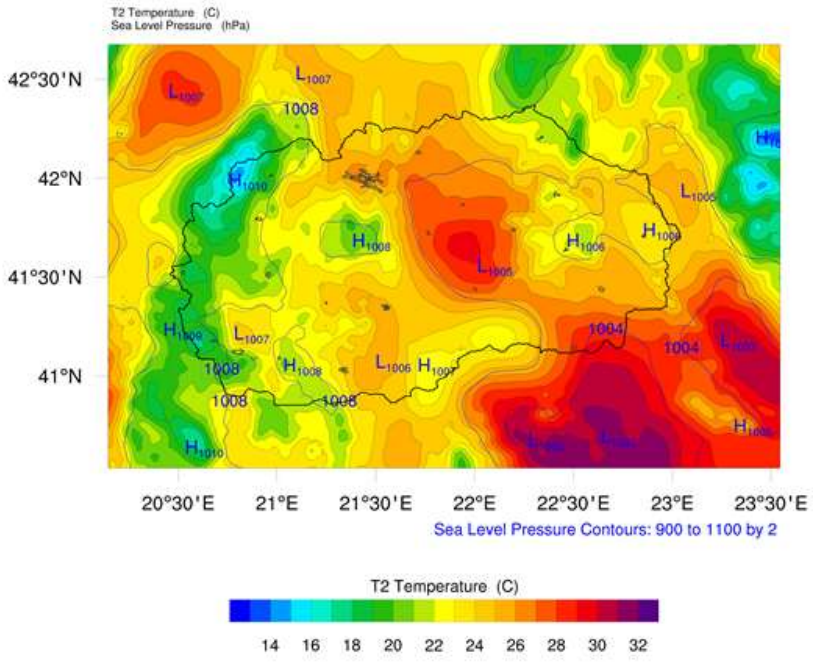


Figure 12 – Sea Level Pressure

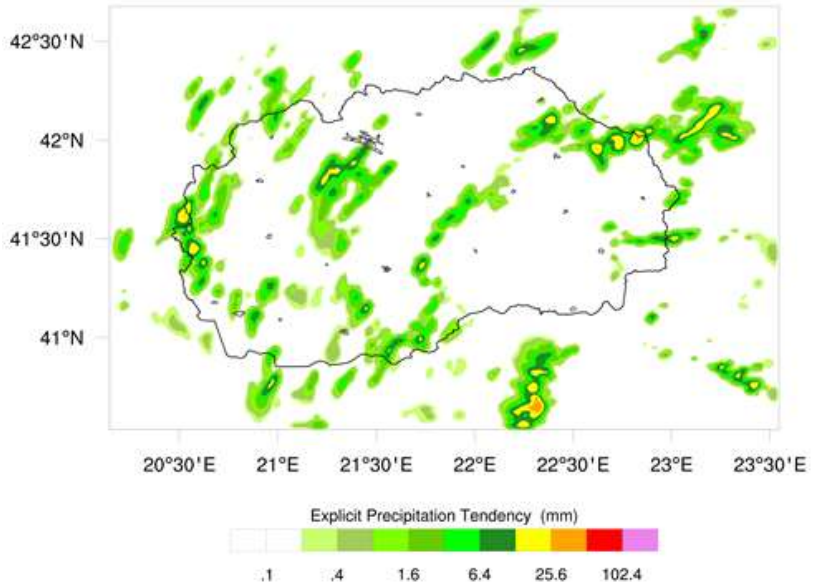


Figure 13 – Explicit Precipitation Tendency

7.1.2 Applications in Life Sciences

7.1.2.1 BRING-MD

Large scale molecular dynamics simulations of BRCA1-RING domain missense mutations
Life Sciences – Area A

Germline mutations in BRCA1 gene are associated with an increased lifetime risk for developing breast and/or ovarian cancer. Currently, 1458 BRCA1 missense mutations have been reported to the ClinVar archive. Of these, 79 have been confirmed as being pathogenic, 150 as being bening and over 1200 are characterized as Variants of Uncertain clinical Significance (VUS) i.e. variants whose pathogenic or benign effect has not been demonstrated. More than 50% of the pathogenic missense substitutions occur within the N-terminal RING domain, which is responsible for the E3-ubiquin ligase activity of BRCA1. Up to date, 62 missense mutations have been identified in the RING domain. Of these, 19 have been classified as pathogenic, 3 as benign and 40 are characterized as VUS. It is known that the BRCA1 RING domain forms a heterodimer complex with the RING domain of BARD1. The formation of this heterodimer complex has been shown to increase dramatically the E3 ubiquitin ligase activity of BRCA1. Hence, the stability and thus functionality of BRCA1 protein is related to the structural stability of the aforementioned heterodimer. Indeed, functional studies of BRCA1 variants demonstrated that pathogenic missense substitutions on the BRCA1 RING domain were also affecting the BRCA1/BARD1 heterodimerization. Furthermore, ubiquitination, which is the ligation of a ubiquitin moiety or a polyubiquitin chain to protein substrates, is mediated by interaction of an E3-ligase enzyme (such as the

BRCA1 RING domain) with an E2-ubiquitin conjugating enzyme. The BRCA1 RING domain binds to E2 enzymes through a surface that is opposite the binding interface with BARD1. The vital role of BRCA1 E3 ligase activity is demonstrated by its targets, which include estrogen receptor-alpha, progesterone receptor, CtIP and histone protein H2A. Therefore, missense BRCA1 RING domain missense mutations that disturb the BRCA1–BARD1 heterodimer complex or BRCA1-E2 interaction, may alter BRCA1 function and lead to an increased risk of cancer. For that reason, investigation of the effects of BRCA1 RING domain missense mutation upon the stability of the aforementioned complexes will contribute significantly in gaining insights into molecular mechanisms underlying the involvement of BRCA1 in tumorigenesis. However, traditional functional assays require considerable time, effort and skills and applying such demanding approaches to study each BRCA1 variant is infeasible. Here, it is suggested the implementation of molecular dynamics (MD) simulations in order to investigate the influences of BRCA1 RING domain missense substitutions upon the structure of BRCA1 RING domain as well as upon the formation of the BRCA1–BARD1 RING domain heterodimer and BRCA1–E2 complexes. In addition, it has been suggested that inhibition of BRCA1-BARD1 RING heterodimer may sensitize tumours to PARP inhibition. Using data generated by MD simulations and virtual screening, we aim to identify small molecules that could selectively inhibit BRCA1-BARD1 heterodimer and or BRCA1-E2 interaction, where BRCA1-RING domain variants are involved.

MD simulations were employed to investigate the influences of Arg71Lys BRCA1 mutation, a known pathogenic BRCA1 mutation related to carcinogenesis, upon its structural stability. Backbone RMSD data showed some flexibility and departure from the native structure within the first nanoseconds of the MD simulation. In addition, the segments of residues 59-68 of BRCA1, remained as beta-turn beta conformation throughout the simulation, whereas it became coil-turn-coil in the case of Arg71Lys BRCA1 mutant. Moreover, the segment of residues 73-77 of BRCA1 was observed to adapt a stable beta sheet conformation throughout the simulation whereas in the case of Arg71Lys mutant, the residues Ser76 and Thr77 were shown to not be part of the beta sheet and became part of the turn conformation adapted by the residue segment 78-80. Therefore, using 20 ns long MD simulations, it was shown to be adequate to show that the Arg71Lys BRCA1 mutation results in increased flexibility and departure from the native structure. Such a conformational disruption could lead to functional instability of BRCA1 that in turn would disrupt molecular mechanism monitored by BRCA1 leading carcinogenesis. As the Arg71Lys is a known pathogenic BRCA1 mutation, these result supports the hypothesis that using MD simulation we could possibly predict conformational disruption of BRCA1 caused by mutations, which could be related to carcinogenesis. This becomes even more important considering that a significant number of BRCA1 mutations remains of unknown clinical specificity. The project aims to investigate more BRCA1 mutations.

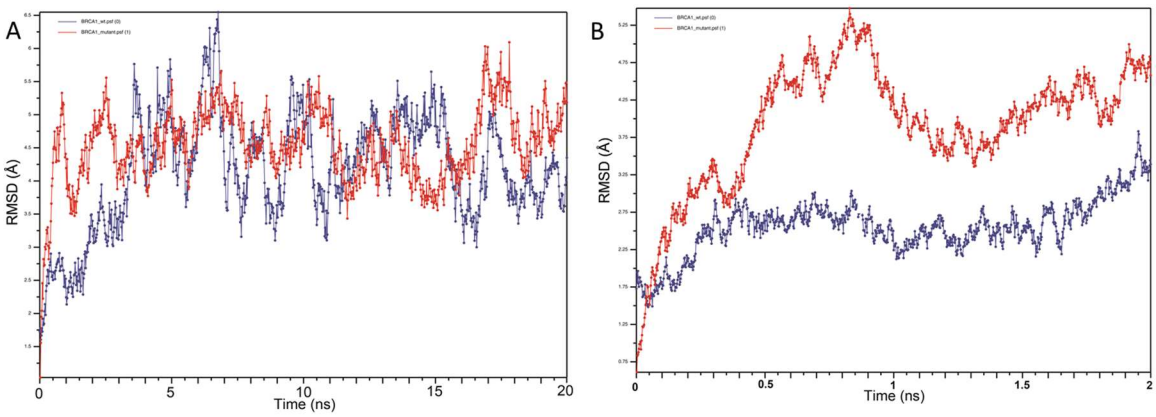


Figure 14 - RMSD values of the position difference of the backbone atoms between the mutant (red) and the wt (blue) structures of BRCA1 RING domain during the MD simulations (A) 20 ns (B) first 2 ns.

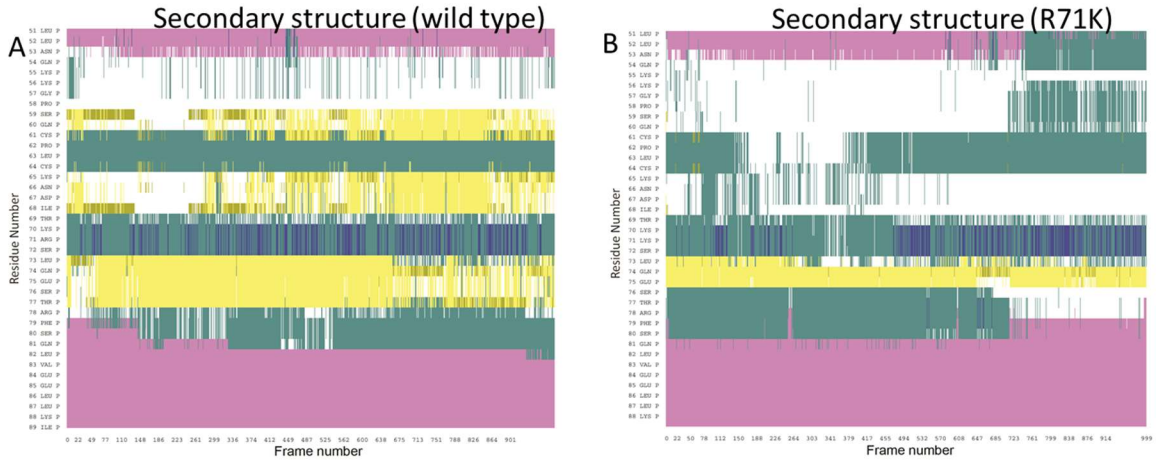


Figure 15 - Structural analysis of the MD simulation results. (A) Time evolution of the secondary structural elements of the wt BRCA1 RING domain and BRCA1 Arg71Lys mutant.

7.1.2.2 OP4D

OrthoPhoto4D
Life Sciences – Area E

In orthodontic diagnostic process, the orthodontist needs to perform a series of measurements, traditionally by manual process using callipers. Recently there is a movement to digitize study models (plaster casts of teeth) and then perform measurements. Although the scanners have come down in price recently, they are still quite expensive and out of reach for most, especially those that are precise enough for this use. This is why the project is developing a method to use photogrammetry in measurements. Operator takes 4 photographs using marked device for holding measured object and then

processes the photographs in the program that is also used for measurements in 3D space based on four 2D photographs. Device is designed to accomodate relatively small objects (approx 80x65x40mm). The software also preforms two analysis used in orthodontics (Bolton and Lundstromm). Also a web based software for measuring 3D objects (generic) is being co-developed with existing DCH efforts related to digitalization of Mueseum of RS collection.

Rapid developments in information technologies lead to wider use of digital representations of dental study models in orthodontics. Most popular way of digitizing the models is to use a 3D scanner and then perform measurements on 3D models, which requires additional and expensive hardware and software resources. In this study an alternative approach was used based on the use of photogrammetry in newly developed OrthoPhoto4D software that calculates and corrects perspective distortion errors.

The individual tooth width for 24 teeth was measured on 50 models, 12 two-teeth segments as well as inter-molar and inter-canine distances. Measurements are performed in OrthoPhoto4D software that uses four photographs of each model for measurements, uses QR codes for automation, calculates the camera position and corrects for perspective distortion caused errors in measurements. Obtained measurements are compared to ones obtained on models generated by structured light 3D scanner.

Statistical analysis strongly indicates that there is no statistically significant difference between two methods. The recorded differences also have no clinical impact as they have mean values of 0.2mm for individual tooth widths, approximately 0.2mm for two teeth segments, and under 0.3mm for both inter-canine and inter-molar distances. All recorded differences fall within the expected measurement error.

Obtained results strongly suggests that OrthoPhoto4D software is of comparable quality to measuring of 3D scanned models in diagnostic and clinical uses in orthodontics. No statistically significant differences in measured values have been found, values produced by two methods strongly correlate and measured differences are on the order of 0.2mm and as such irrelevant in practice. These results illustrate the possibility for a more affordable access with lower requirements for orthodontists in order to conduct digitization of study models and perform computer aided analysis on them.

During this study, through the VI-SEEM project, we were granted access to Steinbichler Comet L3D 5M structured light 3D scanner as well as access to compute cluster at Faculty of Electrical Engineering at University of Banja Luka. Access to this infrastructure enabled the performance of 3D scanning, processing, analysis, visualization and measurements. Two software packages have also been developed for web based measurement and for photogrammetry based measurements which would not be possible without help from the project staff.

In the future, it is planned to explore the automation process of data acquisition as well as to expand the process to use structure from motion.

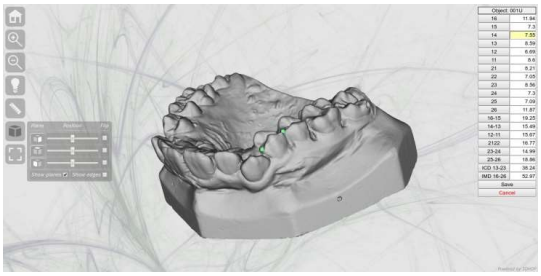


Figure 16 - OP4D Web Application - Measurement UI

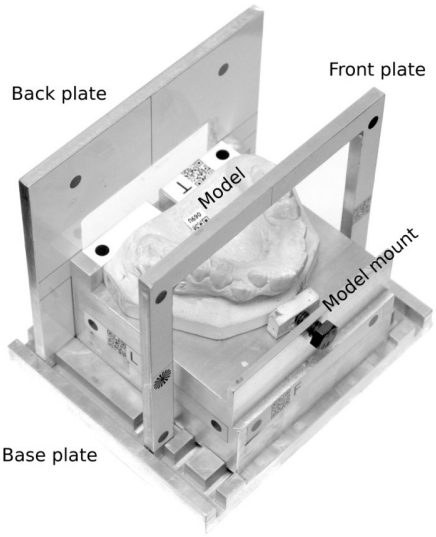


Figure 17 - Measurement apparatus

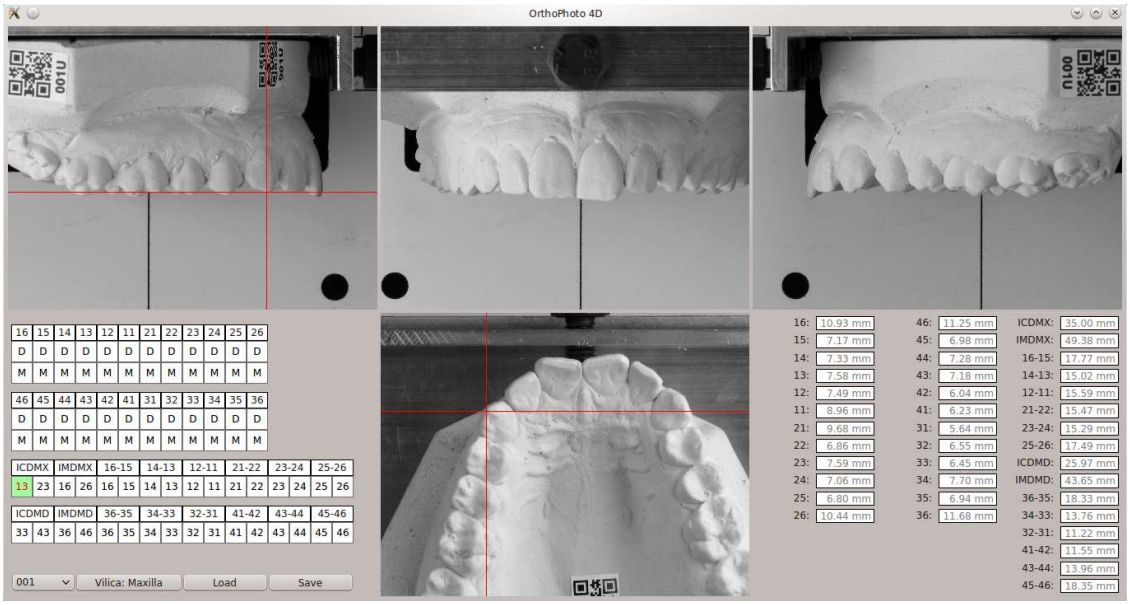


Figure 18 - OrthoPhoto4D User Interface

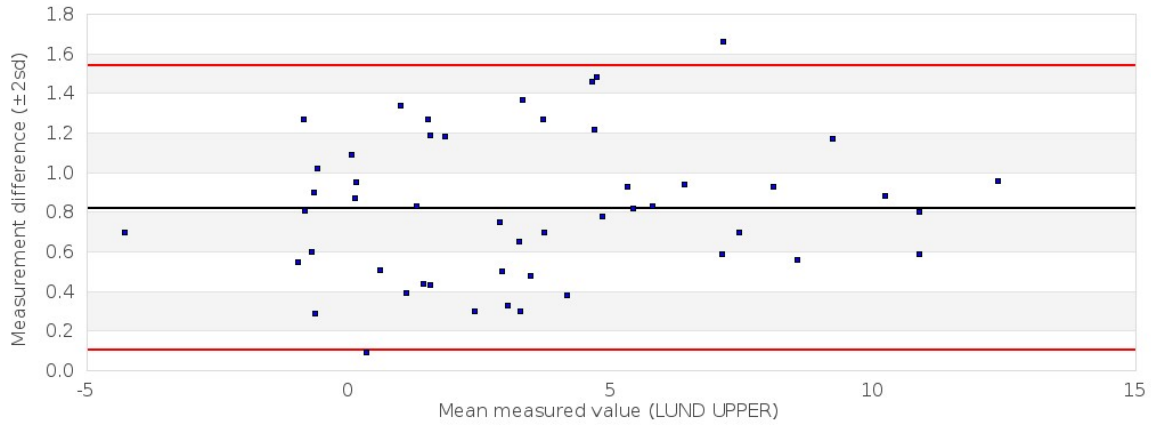


Figure 19 - Bland-Altman plot for Lundstrom analysis for maxilla

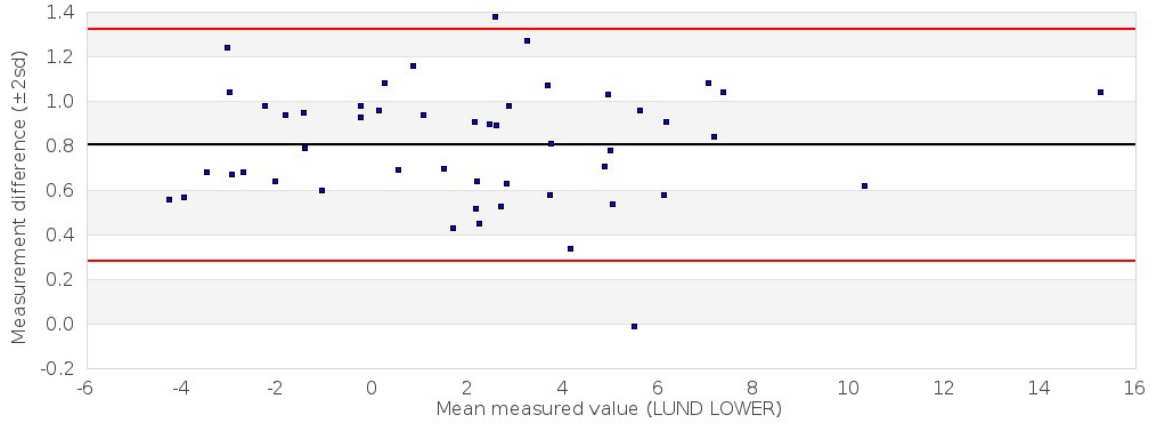


Figure 20 - Bland-Altman plot for Lundstrom analysis for mandible

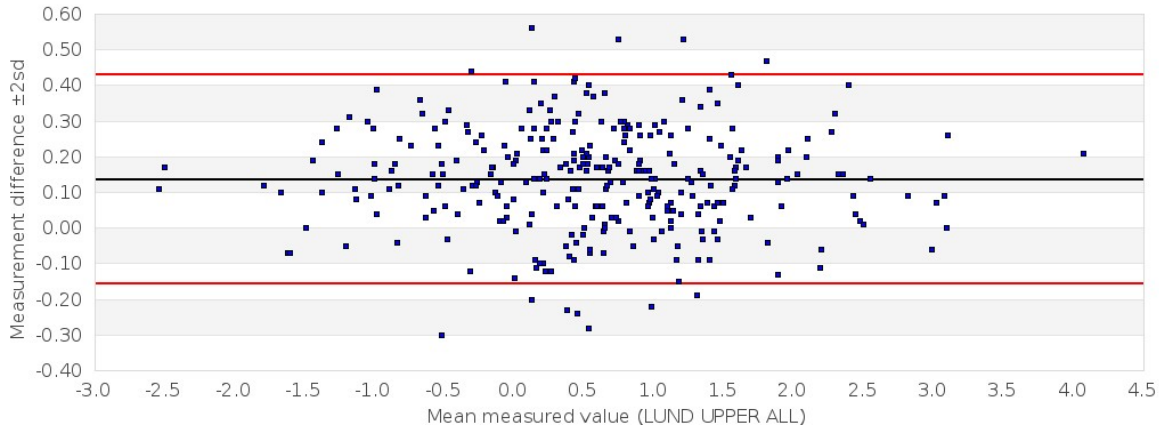


Figure 21 - Bland-Altman plot for Lundstrom analysis for all segments for maxilla

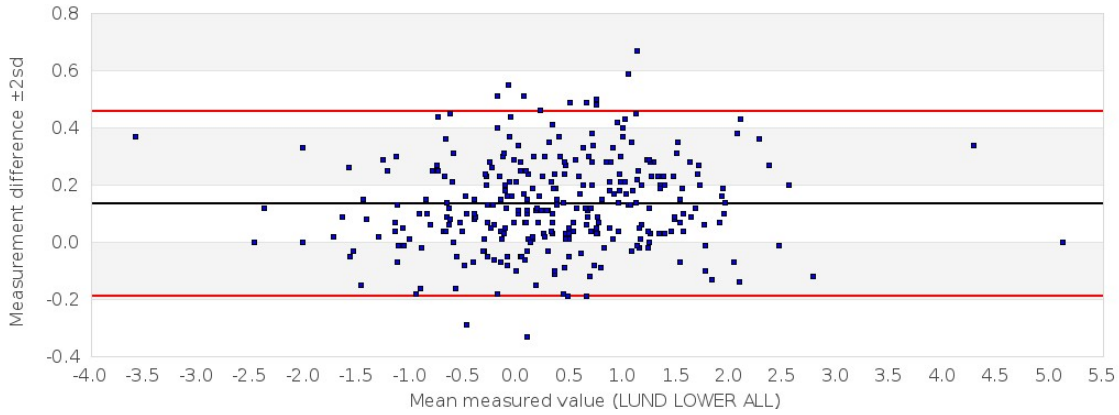


Figure 22 - Bland-Altman plot for Lundstrom analysis for all segments for mandible

7.1.2.3 SIPD

The surfactant induced protein denaturation: A molecular dynamics study
Life Sciences – Area A

Surfactant/protein complexes have been intensively examined due to their large applications in medicine. The mentioned complexes are an interesting system in terms of practical point of view as the ability of surfactant to denature/renature proteins.

Purpose of this study is to understand the binding features of sulfonate to lysozyme. Studies of sulfonate based surfactant/lysozyme complexes at low and high surfactant concentrations. A longscale parallel MD simulation was done (more than 1500ns). The main point is to check whether denaturation of lysozyme is induced by the association (condensation or hydrophobic binding) of alkyl sulfonates to protein or is provoked by protein-ions interactions.

Overall, a synergy between electrostatic and hydrophobic interactions in the presence of SPDS is observed, where the positively charged residues on protein's surface interact with the head groups of surfactants, followed by additional hydrophobic interactions with the surfactants' tails, which results the final denaturation. It is argued that both electrostatic and hydrophobic interactions play important roles in denaturation of proteins induced by anionic surfactants.

Summarizing the results obtained from large-scale MD simulations of protein denaturation, it is argued that:

- At high surfactant concentration, a larger disruption of protein's secondary structure elements occurs. The large disruptions (i.e. high mobility) are mainly in helical region, although at low surfactant concentration the protein also has fluctuations especially in helix structure.

- At low surfactant concentration, a decrease of a helical content (~20-21%) and an increase of beta-strands (up to 11%) is observed.
- The radius of gyration of protein in presence of surfactant is increased up to 1.7-1.8nm.
- Analyzing the trajectories of simulations, it was found that the disruption of tertiary contacts occurs.
- The MD simulations and related experimental measurements indicate that the presence of surfactant molecules lead to the loss of protein's tertiary structure.

The usage of fast supermachines makes it possible to see the mentioned phenomena by allowing to gain deeper molecular insights into the nature of folding/unfolding events and mechanisms.

In the future, it is planned to continue the investigations in this direction and currently the different surfactant/protein systems are considered aimed to elucidate how the surfactant operates at molecular level and reveal the mechanism of protein folding in terms of atomic interactions.

7.1.2.4 Surf_prop

Physicochemical characterization of the natural and synthetic surfactant systems with unique properties caused by specific interactions
Life Sciences – Area A

Classical molecular dynamics (MD) simulations are employed to monitor the behavior of synthetic (LAS and SLES) and natural (Escin) anionic surfactants in the bulk and at the water/vacuum interface, respectively. In the case of the natural surfactant escin, it demonstrated self-assembly behavior on the water surface, as a result of which the molecules were well organized via different types of interactions. The main intermolecular forces governing the escin self-assembly are: (1) long-range attraction due to the inhomogeneous charge distribution in the aglycone and the dispersion (London) van der Waals forces between the aglycone fragments, (2) intermediate-range dipole-dipole interaction, (3) short-range classical H-bonds, and (4) electrostatic repulsion between the charged carboxyl groups in the ionized escin molecules. These types of interaction are the most probable causes for the unique rheological characteristics of saponin adsorption layers. The objective of the current project is to check whether the types and balance of interactions will be preserved also in condensed adsorption layers.

In previous studies on the behaviour of synthetic surfactants, specific interaction between the benzene ring in the LAS molecule and Ca2+ was detected, which is absent in the system with SLES. The experimental data about surface area per molecule was reproduced by monitoring of the adsorption process. To investigate the effect of electrolyte, which is observed in the experiment, it is necessary to simulate model systems in the presence of electrolyte. To explain the experimentally observed phenomena, it is reasonable to carry out molecular dynamics simulations on models mimicking the experimental conditions, which will verify the mechanisms of the specific interactions that lead to experimentally observed

non-trivial behavior of the studied molecules. Furthermore, the more extensive computing power will allow investigation of more realistic molecular systems. Elucidation of the origin of the unique characteristics of the surfactants at the molecular level will have both a fundamental and a practical impact.

The monitoring of the spontaneous adsorption processes of two synthetic surfactants (LAS and SLES) is performed by atomistic MD simulations of large molecular systems which are constructed as close as possible to mimic the experimental conditions. The effect of calcium ions is studied as well as the molar ratio between surfactant and calcium is fixed to the experimental one. Each model system consists of 648 surfactant molecules and the same number of sodium cations to neutralize the negative charge of the amphiphiles. The number of the water molecules is ~20000. In some of the models, 948 of the water molecules are replaced with CaCl2 dissociated into ions. The average total number of atoms in all four studied systems (LAS or SLES with or without Ca2+, respectively) is ca. 100000. For each model, energy minimization, heating, relaxation, and 300 ns production run are applied and statistical data analysis of the whole trajectories is performed.

To estimate the adsorption kinetics of the surfactants in the presence and in the absence of Ca2+, the number of molecules at the surfaces is calculated from the density profiles over the production interval of the simulation time to estimate approximate rate of adsorption. The density profiles of all components in the systems are also analyzed, which provides information about the spatial distribution of the counterions in the simulation box relative to the surfactants.

It is found that the adsorption kinetics of both surfactants is very fast in the initial part of the simulation (tens of ns) when more monomers are present in the bulk. However, at longer times (up to 300 ns), an aggregation process takes place intensively in the bulk and slows down the adsorption process, which hinders the formation of dense adsorption layers. With respect to the effect of calcium on the rate of adsorption, it is found that it accelerates the process more in the initial stage and reduces the surface area per molecule. On the other hand, it leads to faster formation of large lamella in the bulk at the longer times. These results about the effect of calcium ions are in good agreement with the experimentally obtained ones where less area per molecule and precipitation of LAS in the presence of calcium were observed.

In order to explain the obtained effects in more detail, as a next step of the task it is planned to do additional data analysis about the interactions between the surfactants and the counterions, respectively. The results will provide more insight into the behavior of LAS and SLES at the air-water interface and in the bulk as well, which will enable their more efficient application in different washing formulations or will facilitate the design of their substituents with natural origin.

With respect to the part of the project devoted to natural surfactants, the ordering and the interactions between lipids ordered into a lipid bilayer at the air-water interface are

investigated. The atomistic MD computational protocol is similar to the one described above. The models consist of 370 lipid molecules from 35 lipid types, 1 embedded protein receptor, 150 mmol NaCl, and ca. 50000 hydrating waters. The total size of the models is ca. 185 000 atoms. A barostat with constant surface tension, corresponding to the experimentally measured one, is applied. So far, production runs of total length 400 ns are run on Aris, which are to be extended up to 1000 ns. The data are in the process of statistical analysis. The following properties are monitored: interfacial density profiles of the different model constituents, areas per molecule and number of neighbours determined by Voronoi analysis, order parameter of the lipid tails, lateral and normal distribution and diffusion coefficients of the various lipids, secondary structure of the protein. They will provide information of the structuring of the bilayer and its phase state, which can be compared to experimental data. This will allow quantification of the effect of the lipid composition on the physico-chemical characteristics of the bilayer and on the structuring of the cellular receptor. This model may then be used as a reference system to study the behavior of various active substances at bio-relevant interfaces.

The availability of the allocated computational resources within the application Surf_prop was key to performing the research described above. The chance to construct and simulate this much larger and much more realistic model of LAS and SLES layers enabled us to verify the conclusions made at the smaller scale and observe the competitive processes of adsorption and aggregation, which take place in parallel. The simulations of the lipid bilayers allowed estimation of the surface behaviour of a heterogeneous experiment-mimicking natural surfactants assembly, which has not been done so far at such an atomistic level of detail.

The plans for future work on the synthetic surfactants are to study new molecules in terms of adsorption and aggregation processes. The bilayers will be used as a model membrane system to study the interactions of various bioactive substances with the lipid molecules and/or with the protein receptor.

The images in Figure 23 below represent the initial/final structure of the simulated SLES system without calcium. The snapshots in Figure 24 are the respective initial and final structures of the lipid bilayer.

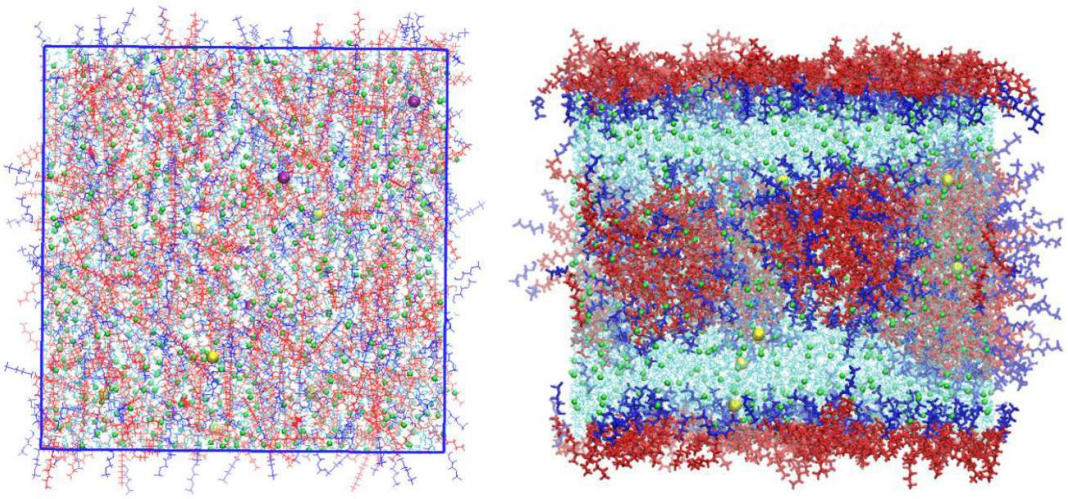


Figure 23 - Initial (left) and final (right) snapshot from the MD simulations of the model of SLES in water; the heads of the surfactants are coloured in blue, the tails – in red, sodium ions are green, chloride ones – yellow, calcium ones – magenta, and water molecules – cyan

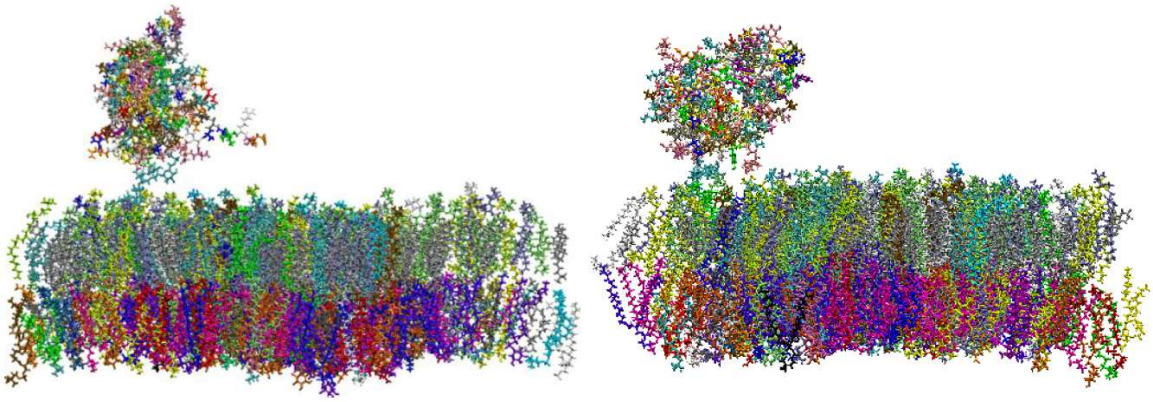


Figure 24 - Initial (left) and final (right) snapshot from the MD simulations of the heterogeneous lipid bilayer with an embedded cellular receptor; the lipids and the protein are coloured according to residue name; water molecules and inorganic ions are omitted for clarity

7.1.2.5 CCCSCMRR

Using the CCC software to compare the result with mammography radiologist results
Life Sciences – Area E

The aim for this project is to test the ability of computer data processing for image in order to detect abnormalities (mainly the cancer cells), and could detect hidden lesions that could not be detected by radiologist eyes follow up of the patient to see the result. The software

used is CCC developed by Synchrotron-light for Experimental Science and Applications in the Middle East (SESAME).

The scientists participating in the project have not yet used the granted services from the 2nd open call. However, it is planned to use the service by the end of the associated open call.

7.1.2.6 OPERA-P

OPportunistically, Elastically Resource Allocation and Provisioning task scheduler
Life Sciences – Area E

Resource management is always a significant challenge to avail all the demanded services of the scientific applications consumers besides the associated difficulty to set up, maintain, and operate these platforms at various computing areas. Aiming at dynamically exploiting SESAME's existing infrastructure, a container-based cluster was prototyped that would be deployed on top of any installed infrastructure. The proposed solution supports tasks execution in a pilot-job abstraction and dynamically shares VI-SEEM capabilities, i.e., computing power, network, and storage. Therefore, it unifies and facilitates access to increasingly heterogeneous resources over Geo-distributed locations in a single pool of resources.

This project attempts to serve as an interoperability layer for allocating and managing resources, besides enhancing the overall utilization of the VI-SEEM infrastructure and regional resources. It also couples new data analytics frameworks and algorithms for the VI-SEEM bag of solutions. In addition, it ports new methods of processing the generated data from SESAME scientific experiments. As for using MapReduce, Apache Spark, and the Hadoop stack technologies in solving protein structure, μ images reconstruction, and 3D volume μ image data analysis for both macromolecular crystallography (MX) and μ-Tomography beam-lines, it is shortly expected to contribute to the project under an open source policy in the Clowder repository.

After designing and prototyping the system, proof-of-concept experiments were performed. It is expected soon to try a real data sets produced from SESAME's scientific investigations. Besides, solving new research fields as for the climate research.
https://vi-seem.eu/2018/07/06/empowering-pilot-abstractions-of-sesame-scientific-applications-over-data-intensive-clouds-and-cyber-infrastructure/

7.1.2.7 DICOM Network

DICOM Network
Life Sciences – Area E

Modern medical information systems integrate various types of medical equipment. This project should solve the actual problems for optimizing the processing and storage of

medical radiography investigations. The standard for working with medical images is the DICOM format, which allows storing studies in good quality with the patient's personal data included. The main problem in storing data in DICOM format are caused due to the fact that one study can take more than 1 gigabyte and consist of thousands of images.

The "DICOM Network" project goal is to provide access to investigations for medical staff with the appropriate access rights and as well as patients to the personal radiography investigations. At current stage eleven medical equipment's are already connected to the system and now "DICOM Network" collects and processes more than 500 gigabytes of data per month. Flexible architecture of the system offers possibility to distribute Data Storage and Data Processing components between different processing units and storages, which could be customized using specific interfaces.

The problem of storing medical investigations archive on national level can be considered as Big Data issue. Solution for this issue should take into account the different data access levels. On the one hand, a medical investigation contains personal patient data, which means that data access should be restricted and secured. This could be reached by permission based categories of users and individual investigations access on supervised approval. On the other hand data should be accessible from any user, like patient or doctor, from any location. One of the main priority is system performance that should allow high speed access of the huge amount of data.

This system covers all necessary workflows for processing and documentation of medical investigations - from collecting images directly from equipment to archiving investigation in the patient medical record. "DICOM Network" offers extended functionality for enhancing quality of medical management and secured access to investigations. This helps doctors, specialists and penitents to gain access to structured database of medical images, allows documenting images that are collecting from various medical apparatus. At institutional level, the system helps to reduce costs of investigation, rise the quality of provided services.

"DICOM Network" is actively developing, but still far from realizing the potential built into the system. Taking into account the growing number of medical equipment and the trend towards modernization and computerization of health facilities, the system will be able to receive and will have to process dozens of terabytes of source information. The storage and subsequent transfer of such large amounts of data is an expensive process, impossible without optimization. On the other hand, for the successful development of the system it is necessary to provide the archive not for 3 years, as provided by law, but for tens of years to monitor the patient's condition and maintain a full medical record. It is also necessary to take into account the need for backup copies of such important information. It is easy to calculate that even for such a small country as Moldova, the data volumes are too large to store them in an unprocessed form. Thus, the issue of data optimization and archiving is a key factor for the development of such systems.

The system initially was deployed at the National Centre of Ambulance of Moldova and during the first year of functioning has shown it effectiveness and attractiveness for personnel that is resulting in:

- Two DICOM Portals were set up: http://dicom.md/ – national portal for Moldavian medical community and http://viseem.dicom.md/ for research community.
- Three DICOM Servers installed
- Eleven types of medical equipment were connected to the "DICOM Network".
- Sixteen medical doctors from radiology section are using "DICOM Network" system in 24/7 regime.
- About 250 investigations per day are collecting by the system.
- Over 500 doctors have access to their patients investigations from their working place.
- Over 12 776 investigations were searched and downloaded by doctors and radiology specialists during 12 months period.
- More than 133 950 investigations stored by the system.
- Additional budget savings ensured for hospitals due to refusing of printing investigations using expensive consumables.

The benefits of using archiving algorithms are opening four perspective directions:

- Reduce the costs of medical data storage and maintenance;
- Reduce internet traffic for data access and in such way reduce the costs for data transfer;
- Increase the quality of radiology services for patients.
- Solve the main problem for DICOM images database – insufficient space for quickly increased amount of data.

As far as radiology medical investigations services are offered by majority of medical institutions starting from small villages to the huge laboratories centers there algorithms should be in a great demand, because on the one hand it reduces the costs and saves organization budget and on other hand increases quality offered services and opens huge opportunities for research and collaboration with other institutes. Of course the most perspective market are huge diagnostically centers in governmental and private sector that have modern equipment and huge number of investigations that should be archived and transferred to other medical institutions. But these solutions will be also interesting for small hospitals that do not have their own equipment but anyway need to have access to the radiology investigations for their patients. Using the proposed algorithms will have possibility to have access to the investigations that were done in the external institution, like private or governmental diagnostics centers.

Algorithms and solutions discussed in this article can be applied in any medical information system and are not tied to a specific project, since the developed approach involves tight integration with the DICOM standard and rather complements it rather than modifies it.

Datasets collecting in the "DICOM Network" system will provide new opportunities for researchers. Although the system is now in production stage, functionality of the "DICOM Network" is rather flexible and permanently enhancing. During the process of the system implementation beneficiaries have ability to specify their specific necessities for providing additional features and services, such as:

- Studying and realization of new methods for optimization of data transfer and archiving.
- Image preprocessing and detection of anomalies.
- Incorporation of expert systems to help making diagnoses for doctors.
- Development of open APIs for "DICOM Network" to collect, archive, access and jointly process medical images by different teams at national, regional and international levels using distributed computing infrastructure.

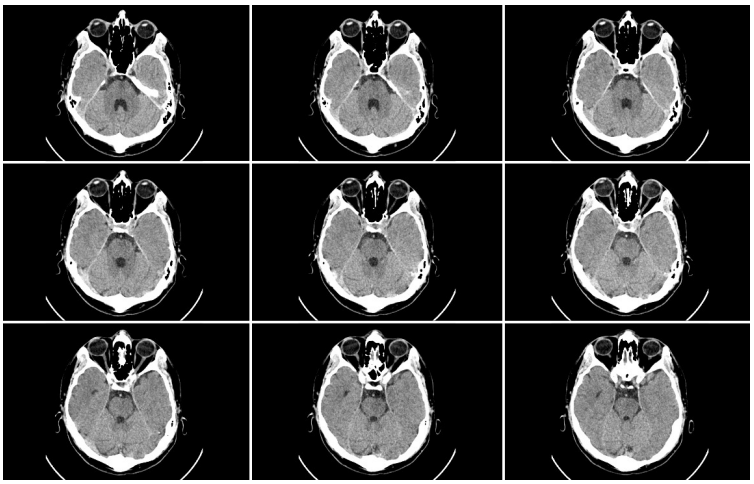


Figure 25 – Medical radiography investigations stored in DICOM Network

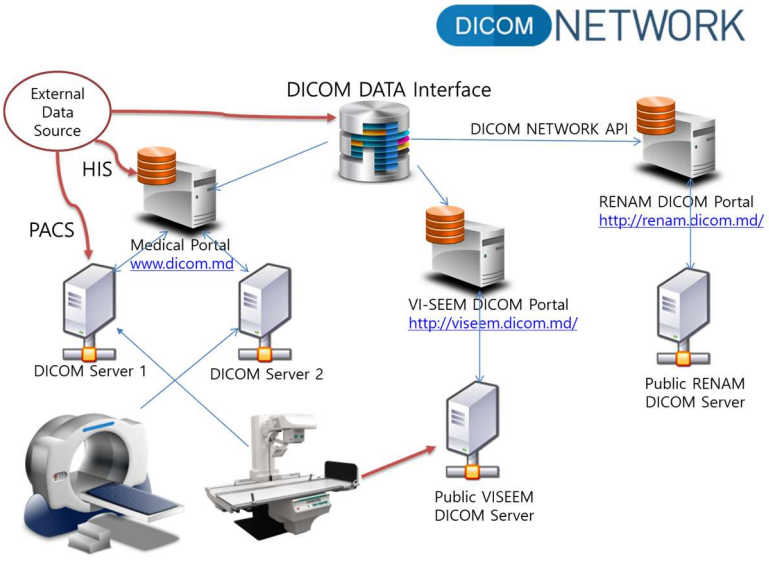


Figure 26 – The "DICOM Network" topology

7.1.3 Applications in Digital Cultural Heritage

7.1.3.1 DCH

The "Aharoni" Digitized Collections: Past, present and future of the southern Levant biodiversity
Digital Cultural Heritage – Area A

The "Aharoni" Online Digitized Collection is an innovative project aimed at creating a suitable platform for presenting and preserving the greatest Levantine faunal collection from the beginning of the 20th century. Natural History collections are rich repositories that document our planet's past and present ecosystems and represent a monumental societal investment in research and applied environmental science. The National Natural History collections at the Hebrew University curate sole collections, The "Aharoni" collections that are comprised of unique fauna (avian, amphibian, reptiles and mammalian) collections and archive material, are the sole direct evidence of the species richness and biodiversity of the Levant region at the beginning of the 20th century. The archival material includes historical documents describing Aharoni's numerous zoological expeditions through the Levant, his comments on specimens, ecological description of habitat that are gone, buried under cities and roads, and other valuable information associated with his studies. The southern Levant, a continental corridor between Europe, Africa and Asia, is a biodiversity hotspot and ideal natural laboratory for measuring historical evolutionary changes influenced by human civilizations in a faunal assemblage over time. Natural history collections represent a crucial contribution to the study of taxonomy, systematics, invasive species, biological conservation, land management and biotic responses to climate changes and human rapid alteration of the ecosystems in the region during the last 300 years. Unfortunately, these collections are significantly under-utilized due to difficulties in access to the specimens by the scientific community. Online digitization including 360-degree images of the specimens as well as the historical and current information on the species will serve as a high-quality database of the southern Levant fauna both for the academic community as a key biological resource and for the general public as a repository of knowledge on this unique region.

As a first stage of the DCH project the scientist in charge used the Clowder repository as a storage space. A number of datasets have been uploaded in Clowder, however, they are "private" and, thus, cannot be seen by anyone else apart from the scientist in charge. However, soon, and once the data analysis is being performed, the permission of the datasets will change and the "Aharoni" collection will be publicly available.

7.1.3.2 mGeoAI

Massive Georeferencing of Aerial Images
Digital Cultural Heritage – Area C

Large number of images often require repetitive and time consuming operation to be georeferenced. With the ability to run AutoGR-SIFT on the Cloud, users can speed-up considerably the entire process of aerial images and have their photographs georeferenced (and with great accuracy) just few minutes after the upload to the server of service provider.

Photogrammetry is quite an old discipline and even its digital version has a long tradition of studies and applications. Recently, thanks to the advent of free/affordable and easy to use amateur or professional applications, a number of scholars from various disciplines started to adopt photogrammetry on large-scale artifacts in Cultural Heritage: vases and statues, historical buildings, arriving to document entire landscapes.

The specific topic of photogrammetric applications to landscape studies constitutes probably one of the most important and promising research branch, i.e. giving archaeologists the possibility to get a digital measurable model of the object of their studies. Recent developments in the Computer Vision community offer new automated procedures for both image orientation and 3D reconstruction purposes at different scales. Despite the availability of a number of easy to use software specifically optimized for end-user computers (Graphic Processing Units or GPU), most of the image-matching processing is still a time consuming operation, and the required time exponentially increases with the number of images (and their sizes) to be processed.

The advent of drones (also known as Unmanned Aerial Vehicles or Remotely Piloted Aerial Systems) for low altitude aerial photography makes it even easier to collect endless number of overlapping photographs for different purposes and final ground resolution of the 3D model. These large amounts of data need proper processing especially when the required output is a set of georeferenced images.

A specific software for automated image georeferencing has been developed at the GeoSat ReSeArch Lab IMS/FORTH by Gianluca Cantoro, allowing large image-set to be automatically matched with a given ortho-photo and geo-referenced in a completely automated way, without human input and completely controllable output. The software, AutoGR-Toolkit (now at version 3.5 - http://www.ims.forth.gr/AutoGR) is already freely available online and it is consistently used by Universities, Research Centers, Commercial companies and private users all over the world (at the moment, more than 2000 users are counted). The possibility to have such a code on a much more powerful computer (or grid of computers) through ViSEEM was of great importance, since it gives user the possibility to easily and quickly georeference large datasets without dedicated computer and expensive hardware requirements for the end user.

Further development of the code is necessary to allow for higher resolution images to be matched and georeferenced, given the provided resources.

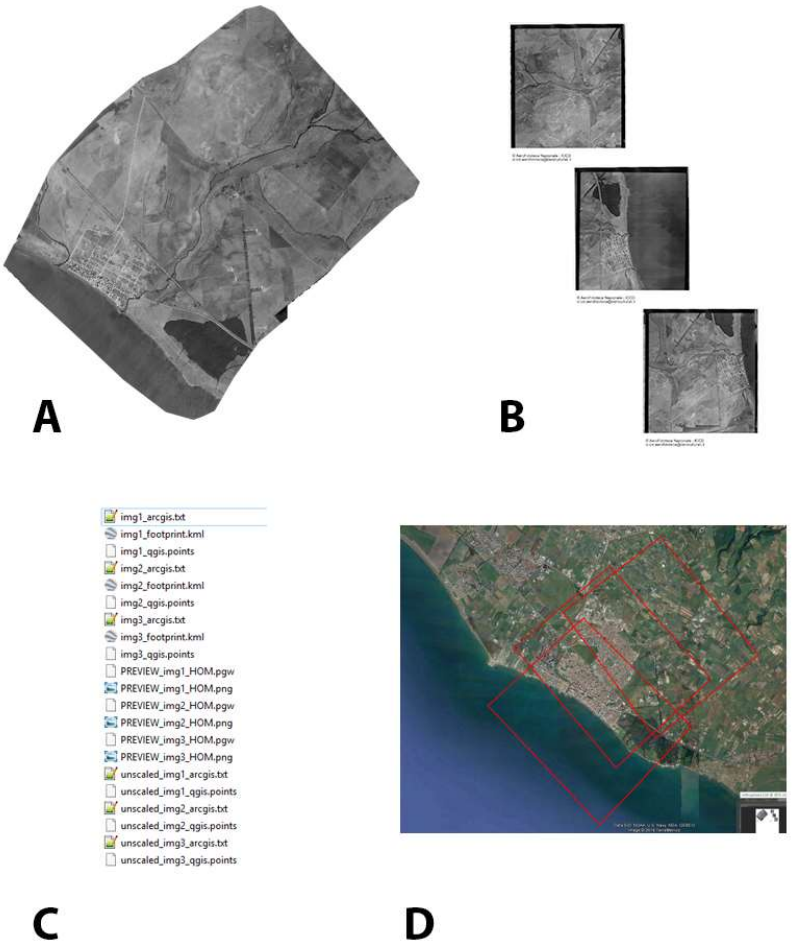


Figure 27 - Example of the usual output of image processing after few seconds from uploading to the platform: A) Reference orthophoto; B) Historical aerial images to be georeferenced; C) Files automatically generated by AutoGR-SIFT; D) Footprints (in red) of photographs in Google Earth.

7.1.3.3 3DGEOEX

3D non-invasive geophysical "excavation"
Digital Cultural Heritage – Area B

Ground based archaeological prospection methods have shown an exponential advance in Greece during the last 30 years in order to meet the needs regarding the efficient management of the Greek cultural heritage. All these years a significant experience has been gradually and systematically built based on the efforts of Greek universities and research institutes as well as foreign expeditions. Especially the research in archaeological

prospection the last decade exhibited a slight deviation from the traditional mapping techniques to more sophisticated imaging and tomographic approaches to address specific problems.

Electrical Resistivity Tomography (ERT) is one of the most important modern techniques of near surface applied geophysics (environmental, geological, archaeological). The last twenty years the wide use of ERT in numerous cases has been facilitated by the development of the technology related with fully automatically multiplexed electrode arrangements and automatic measuring systems. These technological advancements have given the ability to acquire a large volume of data in limited time periods. Further, the advent of fast computers allowed the development of automated resistivity modeling and inversion schemes that aim to construct an estimate of a subsurface resistivity distribution which is consistent with the experimental data. Modeling resistivity algorithms are based on well-established numerical methods such as Finite Element while the reconstruction of the subsurface resistivity can be made feasible though iterative optimization least-squares approaches. Despite the relatively routine employment of the ERT for reconstructing two-dimensional vertical sections of the subsurface, data processing and inversion can be extremely time consuming in cases of truly three-dimensional (3-D) approaches. Such cases involve millions of numerical calculations that also require extensive computer memory resources.

The project was focused on on re-processing ERT data from past archaeological geophysical explorations in Greece. Data from different geophysical case studies were used trying to emphasize in this way the multi-dimensionality and applicability of the ERT archaeological prospection method. These included past ERT surveys from rural and urban archaeological sites.

The data from past ERT surveys falling in the above categories were subject of re-processing with the modern inversion services provided by VI-SEEM and the updated reconstructed ERT models formed the basis to plan future policies for the promotion and management of the archaeological sites.

Future work will include incorporation of time constraints for the time lapse (4-D) of 3D tomographic data, incorporation of structural and a-priory information to constrain the inversion and minimize the ambiguities and development of joint inversion algorithms (e.g. seismic tomography and ERT).

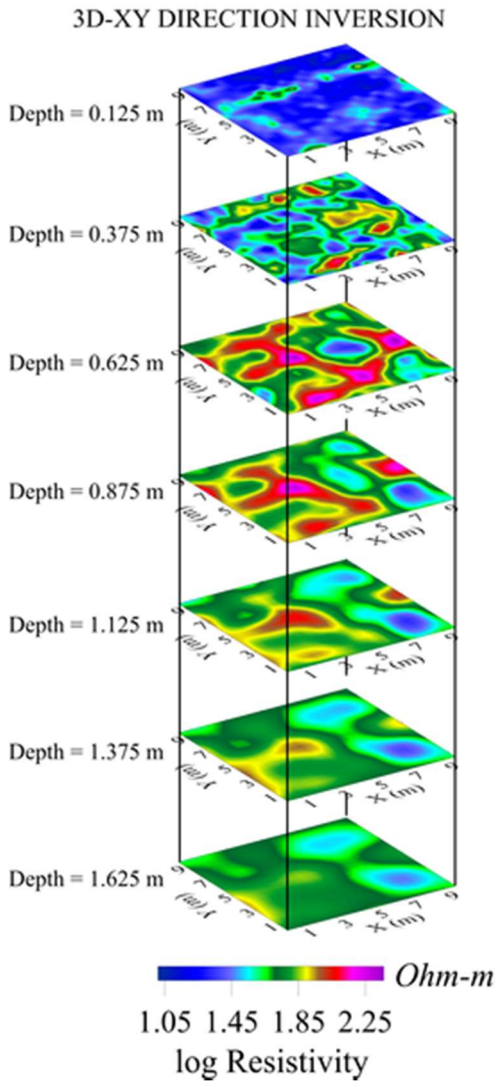


Figure 28 - Horizontal resistivity slices extracted from the 3D inversion model resulted by the processing of 3D tomographic data from the archaeological site of Europos.

7.1.3.4 VirMuF

Virtual Museum Framework
Digital Cultural Heritage – Area C

The Virtual Museum Framework (VirMuf) project is a set tools designed to work under the Unity game engine to help cultural heritage community create virtual museums in a fast and easy way. With cultural heritage digitization technologies becoming more robust and affordable, many museums worldwide are actively digitizing their collections. VirMuF aims at how these digitized collections can be presented in an easy, meaningful and useful way for both professional and casual users. VirMuF makes it easy to create entire virtual museums with so many useful tools without having to do any programming. VirMuf is open-source,

and hence, users with programming knowledge can further extend it so as to fit their requirements.

Currently, all the datasets are stored in the Clowder repository where people can access them and experience the Virtual Museum Framework. However, due to backup reasons the developers of the VirMuf are currently in the process of uploading their datasets into the simple storage as this was granted in the review process of the 2nd open call.

7.1.3.5 DataCrowds

Data-Driven Crowds
Digital Cultural Heritage – Area A

Today more people are living in urban environments than in rural areas. It is forecasted that 70% of the global population will be living in cities by 2050. This intense urbanisation poses huge challenges in overcrowding, segregation, demographics and use of resources. The main goal of this project is to innovate in the unified area of research that is occupied with the transdisciplinary study of crowds in built environment. We envision a web-accessible, social platform that allows researchers from very diverse fields, such as Crowd Simulation, Urban Modeling and Simulation, Pedestrian Dynamics, Computer Graphics, Social Dynamics and Architecture to collaborate, share data and take advantage of each field's breakthroughs in order to contribute more accurate crowd simulations for the future sustainability of urban environments. As a first step in the implementation of this project, we will share tracked data of crowds from various sources in the Clowder platform of the VI-Seem project.

The computer resources allow for the management of the data. These data will be shared to the interested communities through the publications. These data will be used in part by data-driven methods to analyze and simulate crowds.

7.1.3.6 HaPPEn

High Performing Photogrammetry
Digital Cultural Heritage – Area C

Image matching has a history of more than 50 years, with the first experiments performed with analogue procedures for cartographic and mapping purposes. The recent integration of computer vision algorithms and photogrammetric methods is leading to interesting procedures which have increasingly automated the entire image-based 3D modeling process. Image matching is one of the key steps in 3D modeling and mapping. Due to the availability of a number of different low-cost and open-source software systems, automated 3D reconstruction methods are becoming very popular.

However due to the massive dataset collected on sites, the lack of computing resources, such as CPUs, GPUs, RAM, can represent an important limitation. This phenomenon is particularly evident in the cultural heritage domain. Budget constraints for hardware upgrade are indeed often tight and, at the same time, high resolution reconstruction are more frequently requested in a reasonable amount of time. Usually the outputs are then represented by a compromise between the number and quality of the input images and the available quantity and quality of hardware available.

The HaPPEn project aims to setup a set of photogrammetric tools, both open source and commercial, assessing their performance on a HPC infrastructure, and freeing the user for any installation and configuration issue.

Massive Aerial and terrestrial data set concerning the cultural heritage domain (Digital Terrain Model, Built heritage, Archeological sites) will be used for the benchmark.
The photogrammetric steps of the MicMac open source software which may largely benefit from parallel computing are namely:
- the tie point detection and matching (Tapioca command);
- Dense Image Matching (Malt, Tawny and NuagetoPly).

Historically, MicMac code hasn't been written for parallel computing. However, exploiting the pymicmac python script (https://github.com/ImproPhoto/pymicmac) the aforementioned steps can be executed exploiting different nodes on a HPC infrastructure.

However, to date, the installation of the code has faced major issues and it hasn't been totally implemented yet. It has been possible to fully and successfully run in parallel only the tie point detection and matching step, using, as test a small dataset of 70 images running on 5 nodes. The analysis and monitoring of the results of this task has highlighted significant timewise improvement. The Tapioca command run on Cytera infrastructure completed in 40 minutes (5 nodes) whereas the same job on a middle level workstation (cores) finished in 5 hours.

During the execution of the Dense Image Matching step (Malt, Tawny and NuagetoPly) on Cytera, the task could not be completed due to execution errors that, to date, haven't been solved yet.

In conclusion, it can be stated that although showing promising results in term of computing time, the implementation of the code cannot be fully evaluated until the full pipeline can be executed.

In future work, it is planned to test of the full pipeline on the HPC infrastructure and make use of large datasets.

7.2 List of publications from the 2nd Open Call applications

1. A Molecular Dynamics Study of Protein Denaturation Induced by Sulfonate Based Surfactants// In Preparation.
2. A. Gevorgyan, A case study of low-level jets in Yerevan simulated by the WRF model. Journal of Geophysical Research: Atmospheres, 2018, DOI: 10.1002/2017JD027629.
3. Hrachya Astsatryan, Hayk Grogoryan, Eliza Gyulgyulyan, Anush Hakobyan, Aram Kocharyan, Wahi Narsisian, Vladimir Sahakyan, Yuri Shoukourian, Artur Mkoyan, Rita Abrahamyan, Zarmandukht Petrosyan, Julien Aligon, Weather Data Visualization and Analytical Platform, Scalable Computing: Practice and Experience, Special Issue 19(2), pp. 149-156, doi: 10.12694/scpe.v19i2.1351.
4. Hamlet Melkonyan, Artur Gevorgyan, Sona Sargsyan, Vladimir Sahakyan, Zarmandukht Petrosyan, Hasmik Panyan, Rita Abrahamyan, Hrachya Astsatryan, Yuri Shoukorian, An analysis of wintertime cold-air pool in Armenia using climatological observations and WRF model data, IEEE Proceeding Computer Science and Information Technologies (CSIT), 2017, DOI: 10.1109/CSITechnol.2017.8312156.
5. M Alvanos, T Christoudias, GPU-accelerated atmospheric chemical kinetics in the ECHAM/MESSy (EMAC) Earth system model (version 2.52), Geoscientific Model Development 10 (10), 3679-369.
6. M Alvanos, T Christoudias, MEDINA: MECCA Development in Accelerators–KPP Fortran to CUDA source-to-source Pre-processor, Journal of Open Research Software 5 (1).
7. Nikolaou, Z. M., Chen, J.-Y., Proestos, Y., Lelieveld, J. and Sander, R. "Accelerating simulations using Direct Relation Graphs for atmospheric chemistry mechanism reduction". Geoscientific Model Development Discussions, 2018 (doi: 10.5194/gmd-2018-106) – Accepted for publication.
8. J. Kushta, G. Georgiou, Y. Proestos, T. Christoudias, P. Thunis and J. Lelieveld, "Evaluation of EU air quality standards through modelling and the FAIRMODE benchmarking methodology". – Submitted for publication.
9. M. Arapović-Savić, M. Savić, A. Arbutina, M. Umićević-Davidović, V. Mirjanić, System for Measurements of 3D Scanned Orthodontic Study Models, CONTEMPORARY MATERIALS, Vol. 8, No. 2, pp. 172-179, Feb, 2018.
10. M. Arapović-Savić, M. Savić, M. Umićević-Davidović, A. Arbutina, N. Nedeljković, B. Glišić, Photogrammetry based space analysis measurements in orthodontic diagnosis, Stomatološki glasnik Srbije , Vol. 65, No. 2, pp. 78-88, Aug, 2018.
11. A Molecular Dynamics Study of Protein Denaturation Induced by Sulfonate Based Surfactants// In Preparation.
12. Empowering Pilot-abstractions of Scientific Applications over Data-intensive Clouds and Cyber-infrastructure// In Preparation.
13. P Bogatencov, N Iliuha, G Secrieru, A Golubev, DICOM Network for Medical Imagistic Investigations Storage, Access and Processing, Networking in Education and Research"", Proceedings of the 11th RoEduNet IEEE International Conference, Sinaia, Romania, p. 38-42, January 2013.

14. Alexandr Golubev, Peter Bogatencov, Grigore Secrieru, Nicolai Iliuha, DICOM network-solution for medical imagistic investigations exchange, International Workshop on Intelligent Information Systems. Proceedings IIS p 13-14. 2016.
15. Alexandr Golubev, Peter Bogatencov, Grigore Secrieru, Optimal Methods of Storage, Transfer and Processing of DICOM Data in Medical Information Systems, International Conference on Distributed Computer and Communication Networks, p 269-280, September 2017.
16. Alexandr Golubev, Nicolai Iliuha, Peter Bogatencov, "DICOM Network" services DICOM data exchange solution integrated in the regional VI-SEEM infrastructure, Smart Technologies, IEEE EUROCON 2017-17th International Conference on, p 558-563, July 2017.
17. Alexandr Golubev, Peter Bogatencov, Grigore Secrieru, DICOM data processing optimization in medical information systems, Scalable Computing: Practice and Experience, p 189-201, May 2018.
18. A Golubev, P Bogatencov, G Secrieru, Updating DICOM network architecture for its integration at international level, RoEduNet Conference: Networking in Education and Research, 2016 15th, p 1-5, September 2016.
19. Alexandr GOLUBEV, Peter BOGATENCOV, Grigore SECRIERU, Исследование способов хранения, передачи и обработки больших объёмов данных в DICOM формате для медико-информационных систем, The 9th International Conference "Microelectronics and Computer Science" and the 6th Conference of Physicists of Moldova, October 2017.
20. Nikos Papadopoulos, 3DINV version 2: 3-D inversion of surface ERT data, USER's Manual, Rethymno, May 2017.
21. Gianluca Cantoro, AutoGR-SIFT Manual, Rethymno, September 2017.

7.3 Measurable data from the calls

7.3.1 Usage of resources in the 2nd Open Call

The VI-SEEM consortium accepted 17 applications out of 19 submitted in the 2nd call for production use of the VI-SEEM resources and services. As it was the case in the previous call, the consortium selected projects based on the results of their technical and scientific evaluation. Eight projects in total requested access to the HPC resources, six of them asked for access to the Cloud resources, while one of them (OPERA-P application from Jordan) needed the Grid infrastructure. Seven projects required access to the VI-SEEM Simple storage service, three to the Data analysis service, two to the Live access service, three to the Clouder, and one project to each of the Repository, Archival, ChemBioServer, and AFMM service.

In total, 2.2 million CPU-core hours, 1.1 million GPU-card hours, and 74 virtual machines were allocated for the 2nd call projects. Figure 29 illustrates utilization of the allocated HPC

CPU and GPU resources, and Figure 30 the numbers of VM-cores per project. Blue bars represent the amounts of allocated resources, while red bars show the corresponding usage.

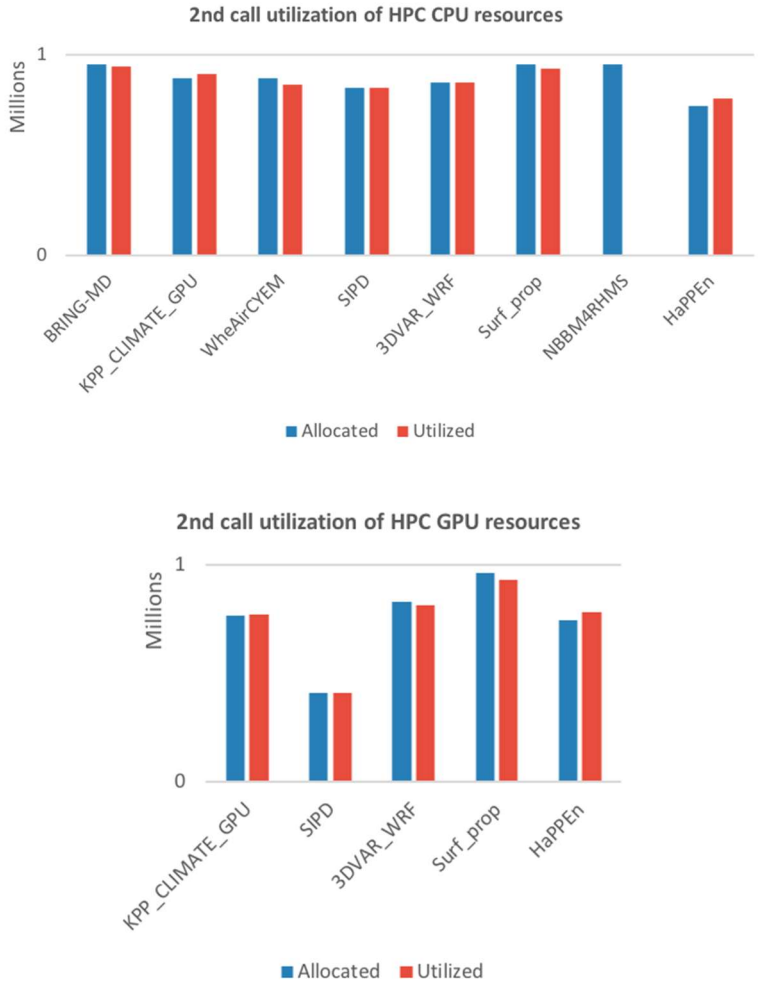


Figure 29 - Utilization of the allocated HPC CPU (top) and GPU (bottom) resources by the 2nd call projects.

To date (September 2018), the total usage of the allocated HPC CPU resources is at 70%, for the HPC GPU resources it is at 75%, and for the Cloud VMs at 85%. In absolute numbers, the 2nd call projects consumed 1.5 million CPU-core hours, 0.8 million GPU-hours, and 63 virtual machines. The call is still active, and we expect additional utilization by November 2018, when it will end. One project (NBBM4RHMS from FYRoM) has faced problems with the model deployment, which required additional porting and development effort, and resulted in a delay of production use. The principal investigator of the NBBM4RHMS project informed us about this issue, which has been resolved by now, and in the upcoming three months we expect the full utilization of the allocated resources to this project.

One project from the second call for production use of the VI-SEEM resources and services used the Grid infrastructure (OPERA-P project from Jordan). The project consumed around 34,000 of CPU-core hours.

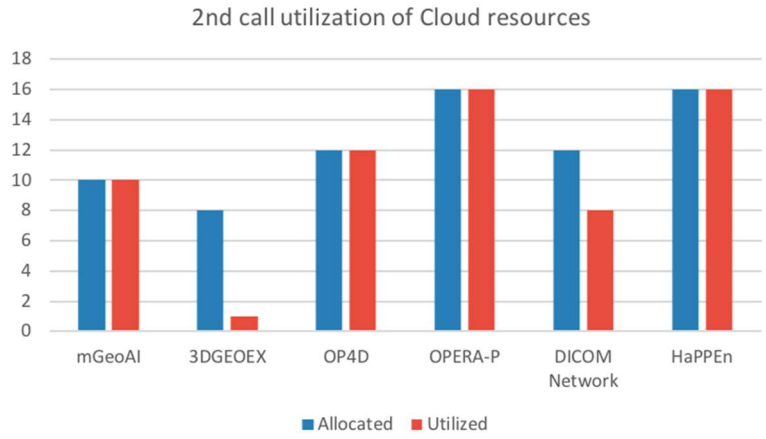


Figure 30 - Utilization of the allocated Cloud resources by the 2nd call projects.

7.3.2 Usage of resources in the SME Call

Through the SME call **Error! Reference source not found.**, the VI-SEEM consortium opened up possibilities for SMEs and regional scientists from the selected scientific fields to have access, via joint projects, to the resources and services we offer. The following projects have been accepted and granted access to the infrastructure:

- MAGIA (Mutation cAlling algorithm IntegrAtion, LS area), submitted by the HybridStat Predictive Analytics company from Greece, is granted access to the ARIS VI-SEEM HPC facility from May to July 2018. The allocated resources for this purpose are 100,000 CPU-core hours and 4 TB of storage space. Additionally, the project is granted 50GB of storage space at the VI-SEEM Simple Storage service.
- IntelliRent (Assessing Impact of Air Quality on Rental Prices, CR area), submitted by the Rental 2000 company from Bulgaria, is granted access to the Avitohol VI-SEEM HPC facility from May to July 2018. The allocated resources for this purpose are 100,000 CPU-hours and 100,000 Xeon Phi-core hours and 1 TB of storage space. Access to the VI-SEEM Simple Storage service (50 GB of storage space) and the VI-SEEM Data Analysis Service is provided as well.
- AOGD (Assessment of Impact of Fleet Management on Air Quality, CR Area), submitted by the ICOM COMPUTERS company from Bulgaria, is granted access to the Avitohol VI-SEEM HPC facility from June to August 2018. In total, 20,000 CPU-core hours and 200 GB of storage space are allocated for this project.
- BRFAA-NGS (BRFAA_NextGenSequencing, LS area), submitted by the IBET company from Greece, is granted access to the ARES VI-SEEM HPC facility from July to

September 2018. Allocated resources for this purpose are 10,000 CPU-core hours and 5 TB of storage space.

- NPssDNA (Atomistic simulations of immune stimulatory single stranded bacterial DNA, LS area), submitted by the Ingredio company from Greece, is granted access to the Leo VI-SEEM HPC facility from August to October 2018. Allocated resources for this project are 100,000 CPU-core hours, 6,500 GPU node hours, and 2 TB of storage space. Additionally, the project is granted access to the VI-SEEM Simple Storage service (50 GB of storage space), and to the VI-SEEM Life Sciences Application Specific Service - Nanocrystal.
- COMP-SS-COMP (Computational exploration of the solid-state compatibility between active pharmaceutical ingredients and excipients by quantum chemistry, multivariate statistics and machine learning techniques, LS area), submitted by the Research & Development Alkaloid AD company from FYRoM, is granted access to the Finki VI-SEEM HPC facility from September to November 2018. Allocated resources for this purpose are 100,000 CPU-core hours and 1 TB of storage space. Access to the VI-SEEM Simple Storage service (50 GB of storage space) is provided as well.

In summary, the total amount of resources used in the SME call is 430,000 CPU core-hours, 6,500 GPU node-hours, 100,000 Xeon Phi-core hours, and 13 TB of storage space. This is a continuous call and will be open until the end of the project, October 2018.

7.3.3 PhD/MSc theses and talks/posters resulting from the 2nd Open Call

Actions	Number
PhD/MSc thesis (taking advantage of / collaborating in the developed project)	4
Talks given/Posters presented	16

7.3.3.1 PhD/MSc Thesis

The following PhD thesis taking advantage of / collaborating in the developed project are the following:

1. Hayk Grigoryan, Ph.D. thesis title - "Weather data visualization and analytical platform''
2. Feras Awaysheh, An enhanced Big Data environment with adaptive, hybrid, and heterogeneous task scheduling architecture
3. Alexandr Golubev
4. Mihail Matenco

7.3.3.2 Talks/Posters

The following Talks/Posters have been presented:

1. T Christoudias, M Alvanos, Accelerated chemical kinetics in the EMAC chemistry-climate model, International Conference on High Performance Computing & Simulation (HPCS), 886-889, IEEE, 2016.
2. Goran Petrevski, Vlado Spiridovnov, Numerical weather prediction system - tool for environmental hazard forecasting, "e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean" The Institute of Information and Communication Technologies (IICT) at the Bulgarian Academy of Sciences, Sofia, Bulgaria, May, 2018.
3. M. Arapović-Savić, A System for Measurements of 3D Scanned Orthodontic Study Models, Contemporary Materials Conference, Feb. 2018, Banja Luka.
4. M. Arapović-Savić, M. Savić, Measurements of Orthodontic Study Models in OP4D, International conference "e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean", The Institute of Information and Communication Technologies (IICT) at the Bulgarian Academy of Sciences, Sofia, Bulgaria, May, 2018.
5. Oral presentation "Adsorption process of LAS molecules studied by molecular dynamics" delivered at "e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean", Sofia, Bulgaria, May 15-16, 2018
6. Oral presentation "Structural characteristics of a solvated multilipid bilayer containing a folate receptor – effect of pressure scaling" delivered at "e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean", Sofia, Bulgaria, May 15-16, 2018
7. Feras Mahmoud Awaysheh, Empowering Pilot-abstractions of Scientific Applications over Data-intensive Clouds and Cyber-infrastructure, Opportunistic Broker for Elastic Resource Allocation (OBERA), https://nubeusc-my.sharepoint.com/:p:/g/personal/feras_awaysheh_rai_usc_es/Ed9xdsgi2RFPg0cITmlBwDcBpiDu09iquMYFCnCNp2wo9A?e=OmW3MZ
8. DICOM Training event AITT, http://grid.md/training-events/dicom-training-event-aitt/
9. Gianluca Cantoro & Nikos Papadopoulos, 2016, GRID computing for archaeology: AutoGR-SIFT and 3DINV applications for non-invasive landscape studies, 4th PANCRETAN SCIENTIFIC CONFERENCE FOR THE ARCHAEOLOGICAL WORK IN CRETE, Rethymno, Greece.
10. Papadopoulos, N. and Cantoro, G., (2018). Image and Signal Processing for Cultural Heritage: AutoGR SIFT and 3DINV Cloud Applications, INTERNATIONAL CONFERENCE "e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean", Sofia Bulgaria, 15-16 May 2018.
11. November 2017: CityScapes Training School, Nicosia, Cyprus: Simulation, Analysis and Visualization of Human Crowds.

12. May 2018: Nonuments Event, Nicosia, Cyprus: DataCrowds - Data-Driven Methods for the Simulation and Analysis of Human Crowds.
13. May 2018: Vi-SEEM Regional Event (Sofia, Bulgaria): DataCrowds - Data-Driven Methods for the Simulation and Analysis of Human Crowds.
14. Gianluca Cantoro & Nikos Papadopoulos, 2016, GRID computing for archaeology: AutoGR-SIFT and 3DINV applications for non-invasive landscape studies, 4th PANCRETAN SCIENTIFIC CONFERENCE FOR THE ARCHAEOLOGICAL WORK IN CRETE, Rethymno, Greece.
15. Papadopoulos, N. and Cantoro, G., (2018), Image and Signal Processing for Cultural Heritage: AutoGR SIFT and 3DINV Cloud Applications, NTERNATIONAL CONFERENCE "E-INFRASTRUCTURES FOR EXCELLENT SCIENCE IN SOUTHEAST EUROPE AND EASTERN MEDITERRANEAN", Sofia Bulgaria, 15-16 May 2018.
16. Gianluca Cantoro, 2017, From the Clouds to the Cloud: automatic aerial image georeferencing through network computing system. AARG 2017 International Conference, September 13 – 15th – Pula, Croatia.

8 Conclusion

This document provides the report on the 3rd and the SME calls for proposals for projects accessing VI-SEEM resources and services, the results of the review process of the applications, the applications which have been given access to computational resources, storage resources and VI-SEEM services as well as the efforts of supporting the chosen applications. WP6 oversees capacity building and, therefore, supports open access calls for VRE partners using existing data and codes but also providing new ones, enriching further the services of the VRE. In total three open calls and an SME call were organised and this deliverable focusses on the two final calls. A pay-per-use call for proposals was also launched before the end of the VI-SEEM project.

The 3rd open call (similarly to the two previous calls) has been designed and implemented according to the outcome of deliverable D6.1 "Framework for VRE resource and service provision". The call was addressed to scientists and researchers that work in academic and research institutions in the region of South Eastern Europe and the Eastern Mediterranean and focuses on research topics in the specific fields of Life Sciences, Climate research, and Digital Cultural Heritage.

The 3rd open call enabled researchers from SEEM countries to access VI-SEEM resources and services. In total 22 applications were received, underwent technical review and where all the requirements were clarified they were sent for the scientific review. 22 applications appeared to match the evaluation criteria, were supported and made use of VI-SEEM resources and services namely, 6 in Climate Research, 8 in Life Sciences, 5 in Digital Cultural Heritage and 3 in Cross-Disciplinary research areas.

The projects which are awarded resources in this call will have access to VI-SEEM services, infrastructure and the overall VRE well beyond the end of the project: the computational resources will be awarded for a period of 12 months, and access to the VI-SEEM data repository service for up to 2 years, beginning July 2018 (for the 3rd call) and immediately after technical evaluation (for the SME call). Provision of guaranteed service to the selected communities beyond the timeline of the project is ensured by a Memorandum of Understanding (and thus also partially funded by the local providers/governments) which was signed in the spring of 2018. This also supports the maintenance of the core data services (PID, VI-SEEM repository) to serve the 3rd call applications beyond project lifetime and also ensures long term preservation of data, with the guarantees and quality of service offered by the local resources providers. This demonstrates a baseline level of VI-SEEM sustainability.

VI-SEEM also opened a specific continuous Call for Proposals for preparatory/development projects accessing the VI-SEEM Virtual Research Environment and services and the associated infrastructure - specifically for partnerships involving industrial partners (SMEs) and academic/research institutions. The call was available since the second half of 2018. In

effect, this was a "try-before-you-buy" program, created in order to attract SMEs to the VI-SEEM VRE services and resources. The "try-before-you-buy" program was based on the idea to provide possibility for SMEs to test the VI-SEEM VRE services free of charge for their purposes before making any decision about possible purchase. After the test period, the SME / industrial players can buy a full access and professional support for the offered service via the pay-per-use call, choosing among the bundles offered in the bundle catalogue.

Access through the SME call was provided for free for research purposes. If applicants required extra support for services other than those that are specified in this call, and they were interested for paid services, they were encouraged to contact the project. In this way the VI-SEEM project was exploring the possibility of supporting its sustainability through per-pay-use by industry.

The call was addressed to consortia of scientists and researchers that work in academic/research institutions and SMEs. Academic/research institutions should be located in the region of South Eastern Europe and Eastern Mediterranean, while the SME partners could be from anywhere in Europe. Collaborators in proposals might reside in any country provided that no specific geographical restrictions apply for access by the corresponding centers that offer resources in the various resource-providing countries. The project proposals were expected to address non-proprietary/open research topics in the specific fields of Life Sciences, Climate research, and Digital Cultural Heritage. Via this call VI-SEEM opened possibilities for SMEs and regional scientists from the selected scientific fields to have access, via joint projects, to the advanced resources and services that it offered.

The SME call received in total 6 applications, they underwent technical review, appeared to match the evaluation criteria and were supported. Hence, 6 applications got access to VI-SEEM resources and services namely, 2 in Climate Research and 4 in Life Sciences.

Finally, VI-SEEM has also built its bundle catalogue to attract customers coming from the private market/industrial sector. The organizations involved in the project opened a pay-per-use call, which will last beyond the lifetime of the VI-SEEM project, offering the possibility for industrial players to buy a full access and professional support for the offered service.

