H2020-EINFRA-2015-1

VI-SEEM

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

Deliverable D5.1

Detailed technical implementation plan for VRE services and tools

Author(s): Constantinos Lazarou (ed), VI-SEEM consortium
Status –Version: Final – f
Date: January 31, 2016
Distribution - Type: Internal

Abstract: Deliverable D5.1 –"Detailed technical implementation plan for VRE services and tools", provides the plan for the development and integration of services, tools and datasets into the VRE platform. It presents the scientific applications, which will be involved for the deployment of these services and tools; it defines the VRE service portfolio; and presents the integration plan.

© Copyright by the VI-SEEM Consortium

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.

This document contains material, which is the copyright of certain VI-SEEM beneficiaries and the European Commission, and may not be reproduced or copied without permission. The information herein does not express the opinion of the European Commission. The European Commission is not responsible for any use that might be made of data appearing herein. The VI-SEEM beneficiaries do not warrant that the information contained herein is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
01/12/2015	a	Constantinos Lazarou	Initial ToC.
30/12/2015	b	Theodoros Christoudias, George Artopoulos, Zoe Cournia, Constantinos Lazarou	First draft, analysis of the user survey results.
5/1/2016	c	Theodoros Christoudias, George Artopoulos, Zoe Cournia	Second draft, analysis of Life Sciences, Climate and Digital Cultural Heritage applications.
14/01/2016	d	Constantinos Lazarou	Third draft
15/01/2016	e	Constantinos Lazarou, Ioannis Liabotis, Dusan Vudragovic, Anastas Mishev, Ognjen Prnjat, Zoe Cournia, Theodoros Christoudias, George Artopoulos, Ivan Marton, Aneta Karaivanova, Emanouil Atanassov	Fourth draft, definition of services and integration plan
31/01/2016	f	Constantinos Lazarou	Final, executive summary, introduction and conclusions.

© VI-SEEM consortium

Preface

In the last decade, a number of initiatives were crucial for enabling high-quality research - by providing e-Infrastructure resources, application support and training - in both South East Europe (SEE) and Eastern Mediterranean (EM). They helped reduce the digital divide and brain drain in Europe, by ensuring access to regional e-Infrastructures to new member states, states on path to ascension, and states in European Neighborhood Policy area – in total 14 countries in SEE and 6 in EM.

This VI-SEEM project brings together these e-Infrastructures to build capacity and better utilize synergies, for an improved service provision within a unified Virtual Research Environment (VRE) for the inter-disciplinary scientific user communities in the combined SEE and EM regions (SEEM). The overall objective is to provide user-friendly integrated e-Infrastructure platform for regional cross-border Scientific Communities in Climatology, Life Sciences, and Digital Cultural Heritage for the SEEM region; by linking compute, data, and visualization resources, as well as services, models, software and tools. This VRE aspires to provide the scientists and researchers with the support in full lifecycle of collaborative research: accessing and sharing relevant research data, using it with provided codes and tools to carry out new experiments and simulations on large-scale e-Infrastructures, and producing new knowledge and data - which can be stored and shared in the same VRE. Climatology and Life Science communities are directly relevant for Societal Challenges.

The driving ambition of this proposal is to maintain leadership in enabling e-Infrastructure based research and innovation in the region for the 3 strategic regional user communities: supporting multidisciplinary solutions, advancing their research, and bridging the development gap with the rest of Europe. The VI-SEEM consortium brings together e-Infrastructure operators and Scientific Communities in a common endeavor.

The overall objective is to provide user-friendly integrated e-Infrastructure platform for Scientific Communities in Climatology, Life Sciences, and Digital Cultural Heritage for the SEEM region; by linking compute, data, and visualization resources, as well as services, software and tools.

The detailed objectives of the VI-SEEM project are:

1. Provide scientists with access to state of the art e-Infrastructure - computing, storage and connectivity resources - available in the region; and promote additional resources across the region.
2. Integrate the underlying e-Infrastructure layers with generic/standardized as well as domain-specific services for the region. The latter are leveraging on existing tools (including visualization) with additional features being co-developed and co-operated by the Scientific Communities and the e-Infrastructure providers, thus proving integrated VRE environments.
3. Promote capacity building in the region and foster interdisciplinary approaches.
4. Provide functions allowing for data management for the selected Scientific Communities, engage the full data management lifecycle, link data across the region, provide data interoperability across disciplines.
5. Provide adequate user support and training programs for the user communities in the SEEM region.

 © VI-SEEM consortium

6. Bring high level expertise in e-Infrastructure utilization to enable research activities of international standing in the selected fields of Climatology, Life Sciences and Digital Cultural Heritage.

The VI-SEEM project kicked-off in October 2015 and is planned to be completed by September 2018. It is coordinated by GRNET with 15 contractors from Cyprus, Bulgaria, Serbia, Hungary, Romania, Albania, Bosnia-Herzegovina, FYR of Macedonia, Montenegro, Moldova (Republic of), Armenia, Georgia, Egypt, Israel, Jordan. The total budget is 3.300.000 €. The project is funded by the European Commission's Horizon 2020 Programme for Excellence in Science, e-Infrastructure.

The project plans to issue the following deliverables:

Del. no.	Deliverable name	Nature	Security	Planned Delivery
D1.1	Project management information system and "grant agreement" relationships	R	CO	M01
D1.2	3-Monthly progress report	R	CO	M03n *
D1.3a	First period progress reports	R	CO	M18
D1.3b	Final period progress reports	R	CO	M36
D2.1	Internal and external communication platform, docs repository and mailing lists	DEC	PU	M02
D2.2	Promotional package	DEC	PU	M04
D2.3	Dissemination and marketing plan	R	PU	M05
D2.4	Training plan	R	PU	M06
D2.5	Promotional package with updates	R	PU	M16
D2.6	1st Dissemination, training and marketing report	DEC	PU	M18
D2.7	2nd Dissemination, training and marketing report	R	PU	M35
D3.1	Infrastructure and services deployment plan	R	PU	M04
D3.2	Service registry, operational and service level monitoring	R	PU	M12
D3.3	Infrastructure overview, assessment and refinement plan	R	PU	M18
D3.4	VRE AAI Model and compatibility with other eInfrastructures	R	PU	M27
D3.5	Final infrastructure overview and assessment report	R	PU	M36
D4.1	Data sources and services deployment plan	R	PU	M06
D4.2	Description of the initial deployed data services	R	PU	M11
D4.3	Description of the final data platform available to VRE users	R	PU	M23
D4.4	Final report on data, services, availability and usage	R	PU	M35
D5.1	Detailed technical implementation plan for VRE services and tools	R	PU	M04
D5.2	Data management plans	R	PU	M06
D5.3	User-oriented documentation and training material for VRE services	R	PU	M13
D5.4	Report on integrated services and the VRE platform	R	PU	M14
D5.5	Final report on integrated services and the VRE	R	PU	M36

 © VI-SEEM consortium

	platform			
D6.1	Framework for VRE resource and service provision	R	PU	M09
D6.2	1st Report of open calls and integration support	R	PU	M20
D6.3	Sustainability and business model	R	PU	M24
D6.4	2nd Report of open calls and integration support	R	PU	M36

Legend: R = Document, report, DEC = Websites, patent fillings, videos, etc., PU = Public, CO = Confidential, only for members of the consortium (including the Commission Services).

**n=1,2,3,…12*

 © VI-SEEM consortium

Table of contents

 © VI-SEEM consortium

References

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

[2] VI-SEEM User Survey http://goo.gl/forms/x6q8IKU9fk

 © VI-SEEM consortium

List of Figures

 © VI-SEEM consortium

List of Tables

 © VI-SEEM consortium

Glossary

AMBER	Assisted Model Building with Energy Refinement molecular simulation programs
CBIR	Content-Based Image Retrieval
COSMO	Consortium for Small-scale Modeling
CPU	Central Processing Unit
CRYSTAL	Computational tool for solid state chemistry and physics
DICOM	Digital Imaging and Communications in Medicine
DREAM	The Dust REgional Atmosphere Model
ECHAM	Global Climate Model developed by the Max Planck Institute for Meteorology
EMAC	ECHAM/MESSy Atmospheric Chemistry
ERT	Electrical Resistivity Tomography
FERRET	Interactive computer visualization and analysis environment
FFTW	Fastest Fourier Transform in the West, library for computing the discrete Fourier transform
FIRFLY	Ab initio and density functional theory chemistry program.
GAMESS	General Atomic and Molecular Electronic Structure System is a general ab initio quantum chemistry package
GATK	Genome Analysis Toolkit
GIS	Geographic Information System
GPU	Graphics Processing Unit
GrADS	Grid Analysis and Display System
GROMACS	Molecular Dynamics Software Toolkit
GUI	Graphical User Interface
HPC	High Performance Computing
IDL	Interactive Data Language, a programming language used for data analysis
LAS	Live Access Server
MEDICI	A multimedia content management system
MESSY	Modular Earth Sub-model System
MM5	The PSU/NCAR mesoscale model
NAMD	Scalable Molecular Dynamics Toolkit
NCL	NCAR Command Language
NetCDF	Network Common Data Form)
NWCHEM	High Performance Computational Chemistry Software
OPENFOAM	Open source Field Operation And Manipulation toolbox for continuum mechanics
PIDs	Persistent Identifiers

 © VI-SEEM consortium

RegCM	The Regional Climate Model system
SEEM	South East Europe and Eastern Mediterranean region
SOL	Soft Ontology Layer
VI-SEEM	VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean
VRE	Virtual Research Environment
WEST	Wind Energy Simulation Toolkit
WRF	Weather Research and Forecasting Model

 © VI-SEEM consortium

Executive summary

What is the focus of this Deliverable?

The focus of this deliverable is to assess the needs of researchers from the three scientific communities - Climate, Digital Cultural Heritage and Life Sciences - and generate a detailed implementation plan, which will describe the domain-specific services, tools, codes, and datasets which will be integrated into the VI-SEEM Virtual Research Environment (VRE), and the timelines for their integration. The results from the questionnaire, which was used to survey the users' needs, is also used to assess the needs in terms e-Infrastructure resources, data services and training.

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

This deliverable provides the work plan for the development and integration of domain-specific services, tools, codes and datasets in the VRE platform. Based on this, WP5 will work in collaboration with WP3 and WP4 to develop the VI-SEEM service portfolio and integrate it on the e-Infrastructure provided by WP3 and the generic storage services provided by WP4. Furthermore, the outputs from this deliverable will be used as a guideline for developing the web portal – the Virtual Research Environment platform - which will entail a user-tailored interface, connected to the e-Infrastructure provided by WP3, the data services provided by the WP4 and the domain specific services provided by WP5. The results and conclusions from this deliverable will be used by the following activities:

- WP3.1-Service Registry and Service Level Definitions
- WP3.2-Implementation of resource and service provisioning
- WP3.3-Operations and resource management of the e-Infrastructure for Scientific Communities
- WP4.1-Data services design
- WP4.2-Data access, preservation and re-use
- WP4.3-Data collection and provisioning
- WP4.4-Data analysis
- WP5.2-Data management plans
- WP5.3-Development of the VRE platform

The complete description of the activities and dependencies can be found in the VI-SEEM DoA[1].

What are the deliverable contents?

After a short introduction, an overview of the user survey and its main results are presented. Next, the analysis of the integration plan for the domain-specific services tools, codes and datasets is performed. Finally, a short overview for each scientific community is provided, along with a presentation of the scientific applications selected from the survey to be part of the VI-SEEM Development Access scheme, which will enable the development of the VI-SEEM domain-specific services.

Conclusions and recommendations

The presented implementation plan for the VI-SEEM services, in line with the VI-SEEM DoA [1], provides the detailed work plan for the development and integration of domain-specific services into the VRE platform.

 © VI-SEEM consortium

The scientific applications, selected with the aid of the user survey, will enable the development of three generic services and one application service category that will contain domain-specific application services. Through these services, the VI-SEEM VRE will be offering more than 25 codes and tools for simulations and post-processing of data, more than 20 datasets, and 10 domain-specific application services. More data sets are expected to become available through the integration of the project results to be achieved during the open calls. The development and integration of these services and service modules will be facilitated through a continuous access scheme, which will allow developers access to the VI-SEEM infrastructure, in order to develop and integrate their codes tools, datasets and domain specific application services to the platform.

The user survey, which was the main tool for preparing this implementation plan, revealed a great deal of interest from the three scientific communities, both in developing and sharing services, datasets, workflows and best practices, but also in using the actual VI-SEEM infrastructure and the VRE. In-line with the VI-SEEM DoA, applications surveyed covered most of the use cases and scientific applications envisaged to be targeted by the project.

Usage but also further integration of new services into the VRE platform will be also facilitated through the open calls. Via the open calls, the production users will be invited to offer the output of their applications in the form of data sets to future users of the platform.

 © VI-SEEM consortium

1 Introduction

Work Package 5 has as main objective to integrate the top level-domain-specific services and applications into the e-Infrastructure and data service platforms of WP3 and WP4, and is also responsible for providing the final VRE platform. This will be carried out jointly with the three scientific communities, i.e. Climate, Digital Cultural Heritage and Life Sciences.

This deliverable defines the implementation plan for the development and integration of domain-specific services for the three scientific communities. The deliverable is the main output generated from Task 5.1 "Refinement of service requirements and technical assessment for integration". The refinement of services and the preparation of the implementation plan was derived from the results of a user survey which was prepared by the three scientific communities' leaders, and representatives from WP2, WP3, and WP4. The survey was covering not only the users' needs in terms of domain-specific services, but also their needs in terms of e-Infrastructures, data services and training.

The domain-specific services were based on use-cases identified in the DoA [1]. The user survey was aiming at assessing the importance such services and datasets (input and/or output) have among members of the three scientific communities, the technical requirements, and the effort needed for developing and integrating these domain-specific services, codes, tools and datasets. From analyzing the results of the survey, the implementation priorities were identified based on the following two criteria: a) the level of readiness and maturity for each service, and b) the effort needed for integrating these services. According to the implementation priorities, services will be integrated into the VRE platform through three integration phases, which will be part of the VI-SEEM Development Access scheme. This access scheme will provide access for a six month period to the developers of 37 applications which will be developing and integrating domain-specific services, codes, tools and datasets to the VI-SEEM VRE. In addition to the timelines for the development and integration of the domain-specific services, the implementation plan provides a summary for the applications selected through the user survey. The survey was very successful, since most of the use-cases targeted by the three scientific communities were covered by more than one application.

The next section provides an overview of the user survey and the main results from the survey. Next, an analysis of the integration plan including timelines is presented, followed by a presentation of all domain-specific services for each scientific community in section 4. Conclusions are summarized in section 5, whereas the appendix contains the questionnaire used in the survey.

 © VI-SEEM consortium

2 User survey

2.1 Purpose and structure of the survey

In order to assess the user needs and for preparing the implementation plan for the domain-specific services, a survey was conducted in December 2015. The survey tool is based on the freely available Google Forms application [2], and it was prepared in cooperation with representatives from WP2, WP3, and WP4.

The survey's main objective was to identify domain-specific services, and assess the technical requirements for the development and integration of the services into the VRE platform. Furthermore, the survey was also used to assess the three scientific communities' needs in terms of e-Infrastructure, data services, and training in the use of e-Infrastructures and domain-specific services. The survey included nine sections with the following setup:

1. **General Information:** Basic identification information regarding the application and its developers i.e. application's name and acronym, the scientific community it belongs to, contact person (scientific contact), data manager (contact details), main developers, present and foreseen co-developers.
2. **Application Details:** A description of the application, the problems it solves, the motivation, the current developing stage and the estimated time when it will be ready for production, foreseen scientific and social impact, expected number of users, and expected services to be offered by this application to the VI-SEEM platform.
3. **Collaborations:** Short description for any scientific collaborations planned or foreseen as part of the VI-SEEM project and the particular application.
4. **Licensing and Availability:** Information about any software license needed, specific copyright requirements, and ethical issues that need to be taken into account.
5. **System Requirements:** Description of the computing resources, i.e. HPC, Cloud or Grid, needed for the deployment of the application, along with other information regarding programing tools, libraries and operating systems.
6. **Data Services:** Specification of the data services needed, i.e. data storage, data processing tools, use of the persistent identifiers (PIDs), semantics and metadata. Furthermore this part of the survey address ethical and policy issues that relate to collection and curation of sensitive data.
7. **Interactions and communications:** Type of interaction between the service developed as part of this application and the environment i.e. the type of user interface.
8. **Training and Support:** Training and support needs for the use of e-Infrastructures i.e. HPC, Grid, Cloud, for the use of data services and big data tools, and for the use of parallel programming methods.
9. **Domain Specific Questions:** For each scientific community, ten additional domain-specific questions assessed specific requirements in terms of tools, codes, training and support, and for identifying foreseen collaborations and publications within VI-SEEM.

 © VI-SEEM consortium

2.2 Overview of the results

The survey was conducted in December 2015. A total of 48 responses were submitted from 17 users in the climate community, 16 in Life Sciences and 15 from Digital Cultural Heritage. From the analysis of the results the following general trends for all three scientific communities were observed:

- Most of the use-cases identified for the three scientific communities, were well surveyed (Table 1). In particular, for the Climate community, most of the applications are in the areas of regional climate modelling, air pollution and quality, model development and weather forecasting. For the Life Sciences community there was interest in the areas of modelling and molecular dynamics study of identified drug targets, computer-aided drug design, analysis of next generation DNA sequencing data and RNA profiling data and image processing for biological applications. For the Digital Culture Heritage there was strong interest in digital libraries, but also in image classification and interactive visualization.

	Application Category	Applications
Climate	Regional Climate Modelling	12
	Air Pollution/Quality	8
	Visualization	8
	Model Development	7
	Weather Forecasting	6
	Global Climate Modelling	3
Life Sciences	Modelling and Molecular Dynamics (MD) study of identified drug targets	10
	Computer-aided drug design	8
	Analysis of Next Generation DNA sequencing data and RNA profiling data	6
	Image processing for biological applications	6
	Visualization	5
	Computational simulation of DNA and RNA	5
	Synchrotron data analysis	5
	Data mining to identify prevalent diseases/mutations in the SEEM region	1
Digital Culture Heritage	Digital Libraries	12
	Image Classification	7
	Interactive Visualization Tools	6
	Visualization	4
	Semantic Referencing	3
	Scientific simulation of materiality and systems' properties	3
	Modelling of Built Environments and Advanced Representation Techniques	1
	Geo-referencing Tools	1

Table 1 - Number of applications by scientific application category.

 © VI-SEEM consortium

- The majority of the applications require the use of the HPC infrastructure. In detail: 37 applications require HPC resources, 12 cloud, and 3 the Grid infrastructure (Figure 1). Six applications are planning to use HPC and Cloud in conjunction, 3 are planning to use HPC and Grid in conjunction, and one is planning to use all three infrastructure types in conjunction.

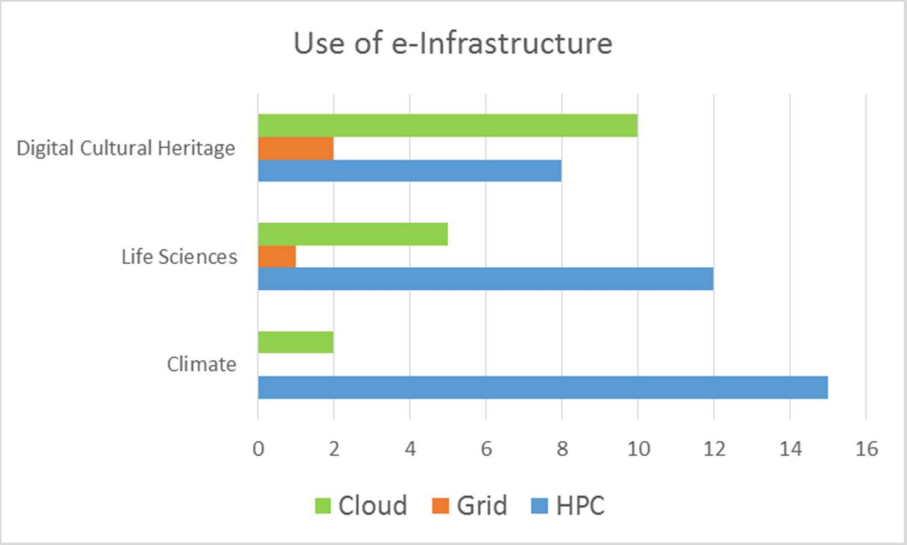

Figure 1-Use of Infrastructure for the scientific communities

- With respect to the maturity of each application, defined as the amount of time required before the application is production ready, for Climate most of the applications will be production ready within two months and a year (Figure 2). For the Digital Cultural Heritage and Life Sciences communities, most of the applications will take longer to be production ready, with the time required to reach maturity ranging from six months to two years.

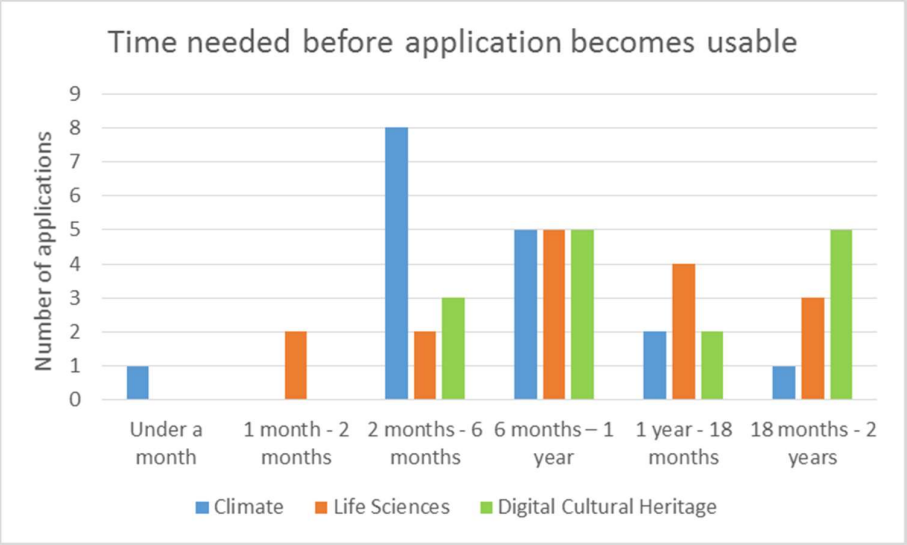

Figure 2-Time needed the applications to become usable.

 © VI-SEEM consortium

- The applications for each community will be targeting communities of direct users with different sizes. In particular, in the case of the Climate community most of the communities are sized in the range of 1-50 users, whereas for Life Sciences the direct users community size ranges from small, to medium size, up to very large (Figure 3). For Digital Culture Heritage the survey results show that roughly half of the applications are targeting large communities of users.

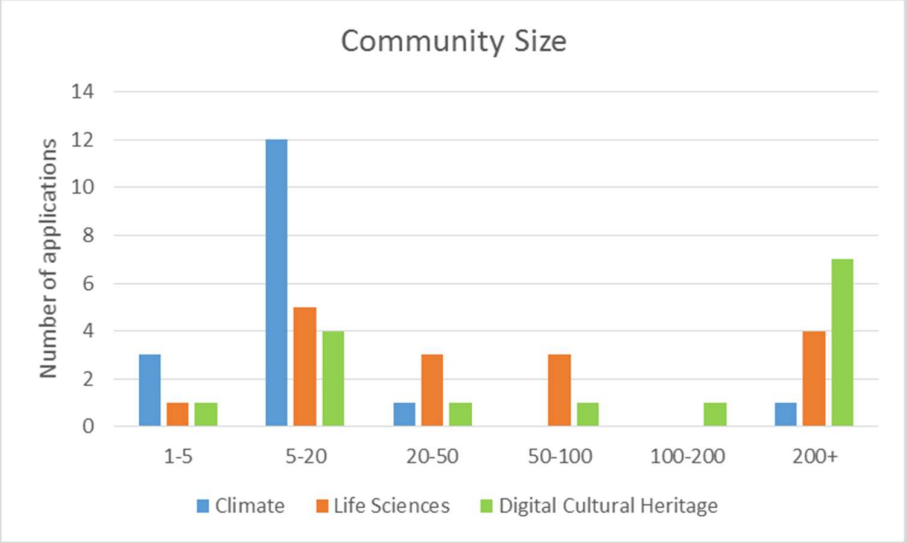

Figure 3-Users community size applications target.

 © VI-SEEM consortium

3 Services integration plan

3.1 Selection of applications and services offered

In order to build the VI-SEEM Services portfolio, the scientific community leaders, and representatives from WP3 and WP4 evaluated the applications gathered from the user survey. In order to select the most appropriate applications for developing and integrating services and datasets to the VRE platform, applications had to fulfill certain requirements. In particular the evaluation procedure was based on the following criteria:

1. **Scientific scope of the application:** Applications scientific scope and expected impact were evaluated in order to assess the appropriateness of the application for the scope of the project. An important requirement was that applications need to be aligned with the use cases identified in the DoA [1], but this did not exclude any applications that were considering other use cases of regional importance
2. **Services offered by the application:** Important condition for an application to be selected was the offering of outputs in the form of services and datasets for the future users of the VI-SEEM platform. Services and datasets need to be unique in the sense that they are not already offered by other projects, and are expected to be used by other scientists in the SEEM region.
3. **Technical specifications of the application:** The technical specifications were also evaluated, where for each application it was important to have completely defined technical specifications that are appropriate for the VI-SEEM infrastructure.
4. **Data services requirements:** Data services requirements for each application were evaluated, and it was important for an application to have well-defined requirements that are appropriate for the VI-SEEM infrastructure.
5. **Maturity and readiness of the application:** Finally, an assessment of the maturity and readiness for each application selected was made in order to facilitate the preparation of the Services Integration Plan (Section 3.2).

Out of the 48 applications, 37 were selected during the evaluation procedure. In particular 13 applications come from the Climate community, 11 from the Digital Cultural Heritage community, and 13 from the Life Science community. The major reason for not accepting the remaining 11 applications was either because there technical specifications or scientific scope were not well defined, or were not appropriate for the scope of the project.

From each application, it is expected a contribution of one or more outputs in the form of a service or data set for the users of the VI-SEEM platform. After the analysis of the responses to the relevant questions of the questionnaire, the services that the select applications will offer to the three scientific communities are divided into 3 generic services and one service category that will contain domain specific application services. More specifically:

1. **VI-SEEM scientific application environment:** This service provides several modules such as optimized applications and libraries, VM images and list of codes, relevant for the work of the regional scientific communities of interest.

 © VI-SEEM consortium

Examples of such modules include, for example, software for calculating Air Quality and Discomfort indices (Climate), and an OCR tool for digitizing Hebrew manuscripts (Digital Culture Heritage).

2. **Workflow, pipeline and software tools repository:** This service provides several modules such as documents containing best practice procedures and workflows for the production of scientific results relevant to the application categories identified in the SEEM region. Examples of such documents are the workflows for simulating a protein with NAMD or with GROMACS (Life Sciences), and the MECCA atmospheric chemistry module on accelerator technology (Climate).
3. **Regional Community Datasets:** This service provides datasets of regional importance for the scientific communities of interest. Examples of such datasets are the RTi dataset of ancient Cypriot coinage (Digital Culture Heritage) and Datasets with data for thermodynamic stability of RNA/DNA and DNA/DNA duplexes (Life Sciences).
4. **Application-Level Services for the regional communities:** This category contains Web-based or visualization services providing easy access to underlying workflows, applications, and resources. Examples of such services are the ChemBioServer (Life Sciences), and the Live Access Server (Climate).

Tables 2, 3 and 4 contain all the applications and expected services, service modules and/or datasets from each application, for the Climate, Digital Culture Heritage, and Life Science scientific communities

3.2 Services integration plan

The development and integration of the services offered by the applications selected from the user survey will be facilitated through the VI-SEEM Development Access scheme. This is a continuous process that will start in M5 of the project. During this process, applications will be given access to the VI-SEEM infrastructure and necessary resources for a period of six months, during which the application developers are expected to, develop and integrate the relevant services to the VRE platform.

The integration plan has been divided into three phases. The timelines for these three phases were selected so that they are in accordance with major milestones of the project. In particular the three integration phases will be implemented as follows:

1. The first integration phase will start in M5 of the project where 10 applications will be given access to the infrastructure and they will be expected to complete the development and integration of services by M12, when the VRE platform will be made available online, and the 1st call for production access will be opened.
2. Sixteen applications that were selected to be part of the second integration phase will be given access to the infrastructure at any time between M5 and M12 of the project and will be expected to develop and integrate services by M18 of the project when the 2nd call for production access will open.
3. Finally, 11 applications which will be offered access during the third integration phase will be given access to the system at any time between M12

 © VI-SEEM consortium

and M22 of the project, and will be expected to develop and integrate their services by M28 when the 3rd call for production access is open.

The integration phase allocation for each application, along with the expected timeline of completion, and the responsible partners and national institutes/communities are provided in table 5. In case that the development of a service is not completed within the six months period, then an appropriate extension of access to the infrastructure will be provided to the developers of the service.

 © VI-SEEM consortium

Table 2 - Services and service modules offered from different Climate applications

Application Acronym	VI-SEEM scientific application environment	Workflow, pipeline and software tools repository	Regional Community Datasets	Application Level Services for the regional communities
EMAC	WRF, WRF-CHEM, EMAC, Paraview, FERRET, NCL	MECCA atmospheric chemistry on accelerator technology		Live Access Server for post-processing of data
WRF-ARW	WRF, NCL		Daily model output (raw output data as well as post-processed output)	
VINE	WRF, WRF-CHEM, EMAC, RegCM, GRADS, NCL		Observation dataset on dust particles in ambient air available from Georgian National Environmental Agency	
RCM MENA-CORDEX	WRF		Gridded datasets of temperature and rainfall for the MENA, via the CORDEX data portals	
ACIQLife	WRF, WRF-CHEM, CMAQ, SMOKE, Software for calculating Air Quality and Discomfort Indices			
HIRECLIMS	CCLM, GRADS, IDL		ROCADA (Romanian Climatic Dataset)	
WRF-Chem (NOA)	WRF, WRF-CHEM, NCL GRADS		WRF-Chem dust aerosol concentrations and various meteorological parameters	
DREAMCLIMATE	DREAM, GRADS, NCL, R, IDL		Downscaled atmospheric-dust DREAM covering wide North Africa, Southern Europe and Middle East regions.	

© VI-SEEM consortium

DRS-ACS	WRF-CHEM		Characteristic constants describing the kinetics of atmospherically relevant processes and spectroscopic properties of the involved species.	
ClimStudyArmenia	WRF, WRF-CHEM, RegCM, GRADS, Paraview, GIS	Methods and methodologies for accurate weather prediction and climate change based on series of experiments		
OPENFOAM	OPENFOAM, Paraview			Software for calculations of physical fields, flows and turbulences
ENB-RCM	WRF, RegCM, MM5, GRADS, FERRET, R		Rainfall records at rain gauge stations in the Eastern Nile Basin	
TVRegCM	RegCM, GRADS, NCL, IDL	RegCM behavior for different model configurations.		

© VI-SEEM consortium

Table 3 - Services and service modules offered from different Digital Culture Heritage applications.

Application Acronym	VI-SEEM scientific application environment	Workflow, pipeline and software tools repository	Regional Community Datasets	Application Level Services for the regional communities
Dioptra	Medici	Introductory level Semantic description of datasets documentation, meta-data definition, Reflectance Transformation imaging process description, structure from motion workflow.	RTi dataset of ancient Cypriot coinage	Online visualization, viewers for: x3D, 3D pdf, RTi ptm and Giga-pan image files
CSAD			Hundreds of RTi files of Ptolemaic inscriptions	Online access to epigraphy and papyrology collections of Eastern Mediterranean
3DINV	3DINV		Existing and new datasets of geoelectrical tomographic data collected from field campaigns.	Web UI for tool for modelling of geoelectrical tomographic data, Subsurface reconstruction and imaging
AutoGR	AutoGR		A demo dataset will be used for demonstration purposes share with the VI-SEEM community for research purposes.	Image based 3D reconstruction tool/ Structure-from-Motion
CHERE	Easy to use modules for individual operations that form the application. The plan is to provide docker containers for every module that can be shared and/or reused, e.g., 3D model cleaning and generating streaming model.			

© VI-SEEM consortium

ELKA			Word, .pdf files of Karamanlidika texts	ORACLE database that enables access to Karamanlidika texts for searching and identifying of phonetic varieties of the indexed lexemes, as well as the finding of grammatical and formational suffixes.
Manuscript	Providing tool for automatic processing of historical documents / OCR tool of digitising Hebrew manuscripts		Datasets of digitized handwritten documents in Arabic or Hebrew	
PETRA		The output can be provided as service and consultation. This includes: databank of surveyed materials, objects, facades, technical datasets of analyzed materials; consultation services of technical details and data analysis methodologies	MEGA Jordan GPS/ geo-referenced data	
BVL	SOL		Banatica database: 1000 books and 200 digitized books	
VirMuf	Medici	Object digitization pipeline, Interactive 3D Museum tours, web UI, consultation services.	3D models, CH Multimedia content and geo-spatial data	
CH-CBIR			Aerial images with annotated land cover types. Learned image representations for imagery in the areas of cultural heritage and remote sensing.	Learned image representations, image search engine, trained classifiers.
IMC4CH	PostgreSQL relational database development packages		Digital material for archaeological monuments	Web interface based of the mapping tools: Arches

© VI-SEEM consortium

Table 4 - Services and service modules offered from different Life Sciences applications.

Application Acronym	VI-SEEM scientific application environment	Workflow, pipeline and software tools repository	Regional Community Datasets	Application Level Services for the regional communities
MD-Sim	NAMD, GROMACS, AMBER, Maestro, Desmond, VMD, Pymol, FFTW, R bbb	Workflows for simulating a protein with NAMD and GROMACS, Workflow for force field parameterization for ligand using AMBER force field, workflow for CADD with Glide	MD trajectories of oncogenic proteins with mutations relevant to the SEEM area	ChemBioServer for post-processing computer-aided drug design results
DICOMNetwork[1]		Workflow for image exchange using different PACS technologies, based on DICOM (Digital Imaging and Communications in Medicine).	Generalized statistical datasets. Patient dataset available after special permission or relevant anonymization of data.	DICOM Web portal with functionalities for doctors, specialists, patients, foreign specialists and legal institutions
CNCADD	NAMD	Share scripts, tools and own codes	Produce and share parameter sets relevant to the community	
PSOMI	NAMD, GROMACS	Workflow for the synthesis and spectroscopic synthesis of molecules	Datasets with molecule synthesis results.	
CCC	OpenCV	Workflow for cancer cells classification		
SQP-IRS	Pymol, Rasmol, ImageJ, Image-Pro, LabView, WinCoot, VMD, JMV, BioCORE, CellProfiler, PDB	Workflows for protein sequence prediction VS. IR spectroscopic data Workflow for CADD for selected targets.	Biological dataset. Computational vs. Experimental database for proteins secondary structures. Crystallographic vs. Spectroscopic database for selected targeted proteins	Web-based portal for SQP-IRS database.
SEMaCD	NAMD, GROMACS, AMBER, Maestro, Maestro, VMD, Pymol, Rasmol, R , ImageJ			
MS4DD	GAMESS, FIREFLY, CRYSTAL, Molekel, Avogadro	Script to submit short workflow of jobs with different input files to use application software as GAMESS.		

[1] The services offered from this application, will be further evaluated before they are integrated into the VRE platform.

© VI-SEEM consortium

THERMOGENOME	GROMACS, Vaa3D, OpenCV	Workflow for measurement of thermodynamic stability of RNA/DNA and DNA/DNA duplexes. Perl and Python Software scripts for measurement of thermodynamic stability of RNA/DNA and DNA/DNA duplexes.	Datasets with data for thermodynamic stability of RNA/DNA and DNA/DNA duplexes for all transcripts, exons, introns, 5-UTRs, 3-UTRs for Homo sapiens (human), A. thaliana, C. elegans, D. melanogaster, D. rerio.	Under development a Web-based application for measurement and visualization of thermodynamic stability of RNA/DNA and DNA/DNA duplexes of genomic regions of interest.
BioMoFS	NAMD, GROMACS, VMD, Pymol, Rasmol	Input files concerning GROMACS simulations of the double - stranded DNA hybridization		
MDSMS	NAMD, GROMACS, VMD, Rasmol, Tcl, FFTW	Analysis tools written in Tcl		
NGS1	R	Documents with best practices and workflows for analysis of NGS data and list of pipelines using free libraries and software		
NGS2	R	List of pipelines using free libraries and software for the discovery of mutations		

© VI-SEEM consortium

Table 5 - Integration phase for the Climate applications.

	Application Acronym	Integration Phase	Integration Completion	Country	Responsible Partner	Institute	Scientific Contact
Climate	EMAC	**1st**	M12	Cyprus	CYI	The Cyprus Institute	Theodoros Christoudias
	WRF-ARW	**1st**	M12	Cyprus	CYI	Department of Meteorology	Demetris Charalambous
	VINE	**1st**	M12	Georgia	GRENA	Tbilisi State University Vekua Institute of Applied Mathematics	Teimuraz Davitashvili
	WRF-Chem (NOA)	**1st**	M12	Greece	GRNET	National Observatory of Athens Institute for Environmental Research	Vassiliki Kotroni
	DREAMCLIMATE	**1st**	M12	Serbia	IPB	Institute of Physics Belgrade	Slobodan Nickovic
	ACIQLife	**1st**	M12	Bulgaria	IICT-BAS	National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Science	Kostadin Ganev
	RCM MENA-CORDEX	**2nd**	M18	Cyprus	CYI	The Cyprus Institute	Panos Hadjinicolaou
	DRS-ACS	**2nd**	M18	FYR of Macedonia	UKIM	Institute of Chemistry, UKIM	Ljupco Pejov
	HIRECLIMS	**2nd**	M18	Romania	UVT	National Meteorological Administration Bucharest	Velea Liliana
	ClimStudyArmenia	**2nd**	M18	Armenia	IIAP-NAS-RA	Armenian State Hydrometeorological and Monitoring Service	Rita Abrahamyan
	TVRegCM	**2nd**	M18	Bulgaria	IICT-BAS	National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Science	Kostadin Ganev
	OPENFOAM	**3rd**	M28	Albania	UPT	Polytechnic University of Tirana	Neki Frasheri
	ENB-RCM	**3rd**	M28	Egypt	BA	Cairo University, Faculty of Engineering, Department of Irrigation and Hydraulics	Mohammed Abouelhaggag

© VI-SEEM consortium

Table 6 - Integration phase for the Digital Culture Heritage applications.

	Application Acronym	Integration Phase	Integration Completion	Country	Responsible Partner	Institute	Scientific Contact
Digital Culture Heritage	BVL	**1st**	M12	Romania	UVT	Central University Library "Eugen Todoran" Timisoara	Delia Pârșan
	CHERE	**2nd**	M18	Bosnia and Herzegovina	UNI BL	University of Banja Luka Faculty of Electrical Engineering	Mihajlo Savic
	Dioptra	**2nd**	M18	Cyprus	CYI	The Cyprus Institute	Georgios Artopoulos
	3DINV	**2nd**	M18	Greece	GRNET	Foundation for Research & Technology, Lab of Geophysical Satellite Remote Sensing and Archaeoenvironment	Nikos Papadopoulos
	AutoGR	**2nd**	M18	Greece	GRNET	Institute for Mediterranean Studies Foundation for Research and Technology	Gianluca Cantoro
	Manuscript	**2nd**	M18	Israel	IUCC	Ben-Gurion University of the Negev	Jihad El-Sana
	VirMuf	**2nd**	M18	Egypt	BA	Bibliotheca Alexandrina, International School For Information Science	Mohammed Elfarargy
	CH-CBIR	**3rd**	M28	Bosnia and Herzegovina	UNI BL	University of Banja Luka, Faculty of Electrical Engineering	Vladimir Risojević
	CSAD	**3rd**	M28	Cyprus	CYI	Centre for the Study of Ancient Documents, Oxford University	Kyriakos Savvopoulos
	ELKA	**3rd**	M28	Cyprus	CYI	Ca Foscari Venice	Matthias Kappler
	PETRA	**3rd**	M28	Jordan	SESAME	Technical University Berlin, Department of Optics and Atomic Physics	Maram Na'es

© VI-SEEM consortium

Table 7 - Integration phase for the Life Sciences applications.

	Application Acronym	Integration Phase	Integration Completion	Country	Responsible Partner	Institute	Scientific Contact
Life Sciences	MD-Sim	**1st**	M12	Greece	GRNET	Biomedical Research Foundation Academy of Athens	Zoe Cournia
	PSOMI	**1st**	M12	Montenegro	UOM	Faculty of Natural Sciences, University of Montenegro	Miljan Bigovic
	CCC	**2nd**	M18	Jordan	SESAME	Computer and Information Technology, Tafila Technical University	Mohammad Alfraheed
	SEMaCD	**2nd**	M18	Serbia	IPB	Institute of molecular genetics and genetic engineering Belgrade	Djordje Francuski
	THERMOGENOME	**2nd**	M18	Bulgaria	IICT-BAS	Institute of Molecular Biology Bulgarian Academy of Sciences	Stoyno Stoynov
	MS4DD	**2nd**	M18	Bulgaria	IICT-BAS	Institute of Molecular Biology Bulgarian Academy of Sciences	Nicolay Dodoff
	DICOMNetwork	**2nd**	M18	Moldova	RENAM	Institute of Emergency Medicine,	Diana Zagadailo Natalia Golubev
	CNCADD	**2nd**	M18	FYR of Macedonia	UKIM	Faculty of pharmacy, UKIM	Marija Glavash Dodov
	SQP-IRS	**3rd**	M28	Jordan	SESAME	SESAME	Gihan Kamel
	BioMoFS	**3rd**	M28	Armenia	IIAP-NAS-RA	Yerevan State University, Armenia	Yevgeni Mamasakhlisov
	MDSMS	**3rd**	M28	Armenia	IIAP-NAS-RA	The International Scientific-Educational Centre of NAS RA, Bioinformatics Department	Armen Poghosyan
	NGS1	**3rd**	M28	Cyprus	CYI	Cyprus Institute of Neurology and Genetics	Athina Theodosiou
	NGS2	**3rd**	M28	Cyprus	CYI	Cyprus Institute of Neurology and Genetics	Petros Mina

© VI-SEEM consortium

4 Applications and related services

4.1 Climate scientific community

4.1.1 Overview of the community

The VI-SEEM Climate scientific community has very strong computational needs, encompassing primarily HPC infrastructure. There are moderate short-term storage needs, large long-term storage requirements and limited adoption of accelerator architectures (potential utilization of Intel Xeon Phi and/or GPUs processors only for certain use cases).

The projected effort focuses on:

- Regional climate modelling and weather forecasting.
- Complementary global climate modelling.
- Study of air pollution and the influence on the climate and human health.

The major underlying scientific objectives are:

- To jointly enable the assessment of the impact of climate change on regional climate.
- Impact studies of the upcoming change on humans, the environment and society.
- Results crucial to predict regional extreme weather and understand future climate trends.

The climate and atmospheric modelling activities pursued through VI-SEEM:

- Have strong geographic and thematic synergies.
- All require neat integration of data and computing resources.

VI-SEEM Climate scientific community objectives:

- Integration of computational resources (HPC, Grid, data infrastructure) to support regional research and operational activity.
- Complement research with code development i) to improve simulation methods ii) for visualization.
- Analysis of the enormous amount of data created in simulations.
- Communication of results to policy makers in the wider public.

4.1.2 Climate applications

4.1.2.1 ECHAM/MESSy Global Chemistry-Climate Model

ACRONYM	EMAC
APPLICATION NAME	ECHAM/MESSy Global Chemistry-Climate Model
MAIN DEVELOPER	Cyprus Institute CaSToRC Climate Modelling Group
SCIENTIFIC CONTACT	Theodoros Christoudias
SCIENTIFIC SCOPE	Assessment of climate change effects on pollution transport in support of Air Quality Policy formulation, pollution source

 © VI-SEEM consortium

	apportionment and advise for impacts

4.1.2.2 Operational Weather Research and Forecast Model

ACRONYM	WRF-ARW
APPLICATION NAME	Operational Weather Research and Forecast Model
MAIN DEVELOPER	Department of Meteorology, Cyprus (CYMET)
SCIENTIFIC CONTACT	Demetris Charalambous
SCIENTIFIC SCOPE	Provision of weather forecasting data and products to operational forecasters, researchers and the public as well as the development and fine-tuning of the model for the Cyprus region in order to achieve the best possible weather forecast accuracy.

4.1.2.3 VIAM/NEA Regional Chemistry-Climate Model

ACRONYM	VINE
APPLICATION NAME	VIAM/NEA Regional Chemistry-Climate Model
MAIN DEVELOPER	Tbilisi State University, Vekua Institute of Applied Mathematics National Environmental Agency, Department of Hydrometeorology
SCIENTIFIC CONTACT	Teimuraz Davitashvili and Georgi Mikuchadze
SCIENTIFIC SCOPE	Improving research in process-level understanding considering to the coupling and feedbacks above the territory of Georgia: dust emission (influence of climate, land surface state, etc.); dust ageing (cloud processing, physical and chemical interactions with other aerosols and gases); dust deposition (land use, lifecycles).

4.1.2.4 Regional Climate Modelling MENA/CORDEX

ACRONYM	RCM MENA/CORDEX
APPLICATION NAME	Regional Climate Modelling (Middle East, North Africa) - WRF
MAIN DEVELOPER	Atmosphere and Climate Division, Energy Environment & Water Research Center (EEWRC), The Cyprus Institute
SCIENTIFIC CONTACT	Panos Hadjinicolaou and George Zittis
SCIENTIFIC SCOPE	Very high horizontal resolution climate projections for the Middle East, North Africa and the eastern Mediterranean, improved climate change projections that will drive important vulnerability, impact and adaptation studies for the region, and resolution of smaller scale meteorological features critical for the realistic simulation of regional climate

 © VI-SEEM consortium

4.1.2.5 WRF-CHEM Regional Climate Modelling (NOA)

ACRONYM	WRF-Chem (NOA)
APPLICATION NAME	WRF-Chem Regional Climate Modelling
MAIN DEVELOPER	National Observatory of Athens, Institute for Environmental Research
SCIENTIFIC CONTACT	Vassiliki Kotroni
SCIENTIFIC SCOPE	Evaluation of the impact of windblown mineral dust on the climate of the Mediterranean region. Quantification of the direct, semi-direct and indirect effects of dust aerosols on the climate of the Mediterranean. Construction of a high-resolution dust aerosols load hindcast for the study region.

4.1.2.6 Dust Regional Atmospheric Model Climatology

ACRONYM	DREAMCLIMATE
APPLICATION NAME	Dust Regional Atmospheric Model Climatology
MAIN DEVELOPER	Institute of Physics Belgrade
SCIENTIFIC CONTACT	Slobodan Nickovic and Luka Ilic
SCIENTIFIC SCOPE	Plan to downscale the atmospheric-dust DREAM model to fine horizontal resolution of 5-10 km, covering wide North Africa, Southern Europe and Middle East regions. Such model setup should be used to perform a decadal model execution and to produce corresponding dust concentration climatology.

4.1.2.7 Wind simulation over rugged terrain

ACRONYM	OPENFOAM
APPLICATION NAME	Wind simulation over rugged terrain
MAIN DEVELOPER	Polytechnic University of Tirana
SCIENTIFIC CONTACT	Neki Frasheri
SCIENTIFIC SCOPE	Simulation of wind over terrain of Albania, with perspective to extend simulations for other atmospheric parameters and weather/climate problems. Results would help to improvement of urban and rural development, redistribution of economic activities, and planning of wind farms.

 © VI-SEEM consortium

4.1.2.8 Eastern Nile Basin Regional Climate Model

ACRONYM	ENB-RCM
APPLICATION NAME	Eastern Nile Basin Regional Climate Model (WRF)
MAIN DEVELOPER	Cairo University, Faculty of Engineering, Department of Irrigation and Hydraulics
SCIENTIFIC CONTACT	Mohammed Abouelhaggag
SCIENTIFIC SCOPE	The regional climate model will be the core model for an integrated regional climate, hydrology, and marine simulator for the Nile basin that can be applied in different research themes. This will fill the gap in the availability of high resolution climate data in the Nile region.

4.1.2.9 Tuning and Validation of the RegCM

ACRONYM	TVRegCM
APPLICATION NAME	Tuning and Validation of the RegCM
MAIN DEVELOPER	National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences (NIGGG-BAS)
SCIENTIFIC CONTACT	Kostadin Ganev and Georgi Gadjhev
SCIENTIFIC SCOPE	Adaptation and tuning of the RegCM model for the Balkan Peninsula and Bulgaria and thus development of a methodology able to predict possible changes of the regional climate for different global climate change scenarios and their impact on spatial/temporal distribution of precipitation, hence the global water budgets, to changes of the characteristics and spatial/temporal distribution of extreme, unfavorable and catastrophic events (drought, storms, hail, floods, fires, sea waves, soil erosion, etc.). All these changes will have influence on the ecosystems and on practically all sectors of the economy and human activity and consequently on the quality of life.

4.1.2.10 Atmospheric Composition Impact on Quality of Life and Human Health

ACRONYM	ACIQLife
APPLICATION NAME	Atmospheric Composition Impact on Quality of Life and Human Health (WRF/CMAQ)
MAIN DEVELOPER	National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences (NIGGG-BAS)
SCIENTIFIC CONTACT	Kostadin Ganev and Georgi Gadjhev
SCIENTIFIC SCOPE	Development of a methodology and performing reliable, comprehensive and detailed studies of the impact of lower atmosphere parameters and characteristics on the quality of life (QL) and health risks (HR) for the population in our country.

 © VI-SEEM consortium

4.1.2.11 *Dynamics, reactivity and spectroscopy of atmospheric chemical species*

ACRONYM	DRS-ACS
APPLICATION NAME	Dynamics, reactivity and spectroscopy of atmospheric chemical species
MAIN DEVELOPER	Institute of Chemistry, FNCS, UKIM
SCIENTIFIC CONTACT	Ljupco Pejov
SCIENTIFIC SCOPE	Development of novel computational and theoretical techniques for modeling of atmospherically relevant molecular species and processes. Study the processes and species affecting the air quality could potentially improve the quality of living.

4.1.2.12 *High Resolution Climate Services for Southern Europe using Regional Climate Modelling*

ACRONYM	HIRECLIMS
APPLICATION NAME	High Resolution Climate Services for Southern Europe using regional climate modelling - COSMO
MAIN DEVELOPER	National Meteorological Administration (NMA), Bucharest, Romania; Climate Department
SCIENTIFIC CONTACT	Velea Liliana
SCIENTIFIC SCOPE	Advance knowledge on the impact of lower boundary specification (soil type, soil moisture, vegetation cover) on the evolution of main atmospheric parameters in the SE-Europe. Improving the skills of climate estimations at monthly/seasonal time scales and for specific regions.

4.1.2.13 *Accurate Prediction and Investigation of Weather and Climate in Armenia and South Caucasus*

ACRONYM	ClimStudyArmenia
APPLICATION NAME	Accurate Prediction and Investigation of Weather and Climate in Armenia and South Caucasus (WRF)
MAIN DEVELOPER	Armenian State Hydrometeorological and Monitoring Service Institute for Informatics and Automation Problems of NAS RA
CONTACT PERSON	Rita Abrahamyan and Anna Shahnazaryan
SCIENTIFIC SCOPE	Methods and methodologies for accurate weather prediction and climate change based on series of experiments, as mountainous terrain of the country, the apparent ruggedness of the terrain, the big difference between relative altitudes, as well as atmospheric general circulation features make it challenge.

 © VI-SEEM consortium

4.2 Digital Cultural Heritage scientific community

4.2.1 Overview of the community

The SEEM region is renowned for its ancient civilizations. It is also an area of major socioeconomic and cultural developments during the medieval and early modern periods. In this context, the cultural heritage of the region is of central importance for the whole of humanity. Yet, these rich traditions and their study are under threat by contemporary political developments and conflicts.

Cultural heritage methodologies deal to a large extent with the storing and analyzing of artifacts and past knowledge. Information technology has enabled electronically preserved heritage to a global audience. The development of sophisticated interfaces and human-computer interaction techniques have only recently allowed researchers in the Humanities to interact meaningfully with the qualitatively complex conditions and parameters that shape human artifacts and affect cultural heritage - an empirical environment which cannot be simulated, and modelled, via simplified formulas.

This Scientific Community aspires to invest in advanced computational tools in order to build an appropriate VRE for the collection and study of the historical artefacts and build scientific interaction that transcends ethnic boundaries and conflict. The use of cloud and grid-computing and HPC capabilities, will enable for the first time researchers, scholars and users from various disciplines to integrate interactive visualization tools with image classification algorithms and semantic referencing, geo-referencing tools, advanced representation and modelling techniques. In particular, the project effort focuses on:

- Strengthening links among key players in the field bringing users currently working autonomously together. Large potential is identified for research groups that have not used large scale computing before. Linking these to experienced groups will significantly improve productivity.
- Facilitate access to the immense CH data.
- Permit the creative use, experimentation, visualization and sharing of these valuable data in a secure VRE context.
- Enable interdisciplinary research.

The biggest e-Infrastructure needs in the VI-SEEM Digital Culture Heritage community, as identified through the user survey, lie in the following applications:

- Content/data management and shared access. The most popular type of application is that of a Digital Library, and communities requested support in semantic organization of data in order to make data available, discoverable, intelligible and searchable, as well as meta-data generation opportunities.
- Visualization, especially for RTi data and 3D models via online integrated players. Two thrusts of needs were identified: a) the development of live (online) viewers for high resolution images, panoramic photographs, RTi files and 3D models, and b) the integration of dynamic generation of 3d models from sequences of specialized image files (e.g., RTi).
- Data processing, mainly regarding Image Classification/Analysis.

 © VI-SEEM consortium

4.2.2 Digital Cultural Heritage applications

4.2.2.1 Dioptra: The Edmee Leventis Digital Library for Cypriot Culture

ACRONYM	Dioptra
APPLICATION NAME	Dioptra: The Edmee Leventis Digital Library for Cypriot Culture
MAIN DEVELOPER	The Cyprus Institute, STARC
SCIENTIFIC CONTACT	Augoustinos Augousti
SCIENTIFIC SCOPE	To this purpose an endeavor that would greatly benefit from Grid & Cloud computing is the CyI DIOPTRA platform (http://dioptra.cyi.ac.cy/index.html), which among others hosts the Collections of Cyprus Antiquities in Foreign Museums (CCAFM) project (http://ccafm.cyi.ac.cy) that aims at assembling in a digital form all the collections of Cypriot antiquities and artifacts expatriated since the 19th century (a program conducted with the support of more than 10 international museums) so as to facilitate research in the archaeology of Cyprus. The digital registration, which has been initiated by the Cyprus Institute, will be enriched continuously, with the assistance of curators of foreign museums. The deployment plan proposes the support of DIOPTRA at the CyI, and in particular the History and Culture of Cypriot Medieval Coins component of the platform. This collaborative project with the Bank of Cyprus Cultural Foundation (BOCCF) aims at the creation of a novel digital library that will promote the study, promotion and dissemination of the history of Medieval Cypriot Coinage (12th-16th centuries). This effort involves the implementation of Reflectance Transformation Imaging (RTI) photography and other advanced imaging applications on coins from the BOCCF's Museum of the History of Cypriot Coinage. RTI imaging, developed at CyI's ICACH, offers tremendous documentation and image analysis possibilities for coins. Results will be appropriately contextualized with explanatory texts and information offering an overview of the historical and cultural framework of medieval coinage in Cyprus.

 © VI-SEEM consortium

4.2.2.2 Banat Virtual Library

ACRONYM	BVL
APPLICATION NAME	Banat Virtual Library
MAIN DEVELOPER	West University of Timisoara
SCIENTIFIC CONTACT	Ion Faget
SCIENTIFIC SCOPE	The motivation of the CUL is that Banat region is characterized by the most diverse cultural, linguistic, national and religious in Romania. Thus there are now in Banat region the following national communities: Romanian, Hungarian, German, Serbian, Italian, Slovak, Hebrew, Bulgarian, Russian, Ukrainian, Gypsy, Arabic, Persian, Indian, Chinese, Greeks, Africans, Asians, etc. All those communities left a footprint in the cultural common heritage. For a better understanding between different communities is mandatory to identify the common heritage, to know the history and the events one community can identify with, and a digital archive of the evolution of language and dialects in the region would enable mapping identities and stories of the different communities that populate the area.

4.2.2.3 Electronic Corpus of Karamanlidika Texts

ACRONYM	ELKA
APPLICATION NAME	Electronic Corpus of Karamanlidika Texts
MAIN DEVELOPER	University of Cyprus
SCIENTIFIC CONTACT	Stelios Irakleous
SCIENTIFIC SCOPE	The final product of the Electronic Corpus of Karamanlidika (ELKA) will offer access to graphic, phonetic and morphemic varieties search, as well as ceramics and pottery study and restoration data management, visualization and presentation to wider communities.

4.2.2.4 Characterization and Conservation of Paintings on walls and sculptures from Nabataean Petra

ACRONYM	PETRA
APPLICATION NAME	Characterization and Conservation of Paintings on walls and sculptures from Nabataean Petra
MAIN DEVELOPER	Department of Antiquities of Jordan
SCIENTIFIC CONTACT	Maram Na'es
SCIENTIFIC SCOPE	Characterization of newly excavated painted marble sculptures from Petra and of gilded wall paintings from Petra. The application will develop a method for confocal μXRF and μXANES, and experimental conservation material for gold on painted surfaces. Moreover the application will be surveying, documenting and condition assessment of wall and sculpture paintings in Petra.

 © VI-SEEM consortium

4.2.2.5 *Centre for the Study of Ancient Documents, Oxford University*

ACRONYM	CSAD
APPLICATION NAME	Centre for the Study of Ancient Documents, Oxford University
MAIN DEVELOPER	Centre for the Study of Ancient Documents, Oxford University
SCIENTIFIC CONTACT	Kyriakos Savvopoulos
SCIENTIFIC SCOPE	Provide a focus for the study of ancient documents, while holding one of the largest epigraphical archive in the world. A wealth of archive materials that were previously unavailable online will be accessed by a worldwide audience.

4.2.2.6 *Virtual Museum Framework*

ACRONYM	VirMuf
APPLICATION NAME	Virtual Museum Framework
MAIN DEVELOPER	Bibliotheca Alexandrina, International School For Information Science (Amr Rizq, Amal Shedeed, Marwa Rashwan, Mariam Seleem)
SCIENTIFIC CONTACT	Mohammed Elfarargy
SCIENTIFIC SCOPE	Popularize the museum experience by adding many ways to simplify and present archaeological and historical data. By utilizing advance programming and graphics techniques, we aim to make a tool that complements and enriches the real museum, substitutes for the inability to visit it, adds new layers and means of storytelling, connects multiple museums worldwide, and that can be used remotely by anyone in the world.

4.2.2.7 *Cultural Heritage Repository*

ACRONYM	CHERE
APPLICATION NAME	Cultural Heritage Repository
MAIN DEVELOPER	University of Banja Luka Faculty of Electrical Engineering
SCIENTIFIC CONTACT	Mihajlo Savic
SCIENTIFIC SCOPE	Documenting and streamlining operational procedures for management and use of cultural heritage data. Application will enable efficient data creation and management for Institutions of culture as well as easier access to said data for general population.

 © VI-SEEM consortium

4.2.2.8 *Image matching with GRID system*

ACRONYM	AutoGR
APPLICATION NAME	Image matching with GRID system
MAIN DEVELOPER	Institute for Mediterranean Studies (Greece): Gianluca Cantoro, Aris Kidonakis, Angelos Chliaoutakis
SCIENTIFIC CONTACT	Aris Kidonakis
SCIENTIFIC SCOPE	The specific application is particularly suited for large image datasets, such as the aerial photographs collected with UAVs or during systematic aerial surveys. The GRID system is going to speed up the georeferencing process. The purpose of AutoGR is to be used as online service for image georeferencing.

4.2.2.9 *Content-Based Image Retrieval and Classification in Cultural Heritage Applications*

ACRONYM	CH-CBIR
APPLICATION NAME	Content-Based Image Retrieval and Classification in Cultural Heritage Applications
MAIN DEVELOPER	University of Banja Luka, Faculty of Electrical Engineering: Digital Signal Processing Group, Computer Science Group
SCIENTIFIC CONTACT	Vladimir Risojević
SCIENTIFIC SCOPE	Development of discriminative and compact image representations for CH applications. Effective and efficient algorithms for image retrieval and classification in large collections. Efficient usage of HPC and Grid infrastructure in CBIR.

4.2.2.10 *Manuscript*

ACRONYM	Manuscript
APPLICATION NAME	Manuscript
MAIN DEVELOPER	Ben-Gurion University of the Negev
SCIENTIFIC CONTACT	Jihad El-Sana
SCIENTIFIC SCOPE	Provide a tool for automatic processing of historical documents, enabling scholars to access and process historical documents efficiently

4.2.2.11 *Electrical Resistivity Tomography*

ACRONYM	3DINV
APPLICATION NAME	Electrical Resistivity Tomography
MAIN DEVELOPER	Foundation for Research & Technology, Lab of Geophysical Satellite Remote Sensing and Archaeoenvironment (Greece)
SCIENTIFIC CONTACT	Nikos Papadopoulos
SCIENTIFIC SCOPE	Electrical Resistivity Tomography (ERT) comprises one of the most important modern techniques of near surface applied geophysics. The method has met an increasing interest

 © VI-SEEM consortium

	within the geophysical community due to its robustness and applicability in solving diverse problems related to: a) the mapping of geological formations and stratigraphy; b) the identification of underground zones related to water and mineral resources and geothermal activity; c) the detection of pollution zones and pollutants that flow in the earth's subsurface; d) the extraction of quantitative information for buried archaeological relics; and, e) the spatial and temporal monitoring of the subsurface resistivity. 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 make possible to acquire a large volume of data in limited time periods, the analysis of and access to these large scale data repositories can benefit greatly by the use of HPC. Further, a HPC facility would enable the integration of cross-discipline data and facilitate automated resistivity modeling and inversion schemes that aim to construct an estimate of a subsurface resistivity distribution, which is consistent with the experimental data.

 © VI-SEEM consortium

4.3 Life Sciences scientific community

4.3.1 Overview of the community

The Life Sciences (LS) research community in the SEEM region deals with research topics that play a central role in achieving a higher quality of life in the region. It has very strong computational needs requiring mainly HPC resources. There are moderate short-term storage needs, moderate long-term storage requirements and limited adoption of accelerator architectures (potential utilization of GPUs).

The projected effort focuses on:

- Characterization of special genotypes/phenotypes of human sub-populations of the SEEM area.
- Analysis of dominant cancerous mutations or overexpression of oncogenic proteins in Eastern European patients that have been already genotyped.
- Primary, secondary and tertiary analysis of Next Generation DNA sequencing data for the detection of aneuploidies.
- Using DNA sequencing alignments of Next Generation Sequencing data.
- Modelling and classification of pathologies by means of image processing.
- Molecular dynamics simulations for characterization of proteins involved in disease.
- Computer-aided drug design targeting proteins involved in disease.

The major underlying scientific objectives are to:

- Enable the description of the medical needs in this particular target area.
- Identify disease mechanism pathways.
- Target oncogenes using computer-aided drug design.
- Provide patients with timely diagnosis, assessment of risk for developing the disease.
- Prediction of mutation effect based on phenotype.
- Targeted and efficient therapy.
- Ultimately lead to regional characteristics that would assist the effort for developing personalized medicine in the SEEM region.

VI-SEEM Life Sciences Scientific Community Objectives:

- Integration of computational resources (HPC, Grid, data infrastructure) to support regional research and operational activity.
- Complement research with code development i) to improve simulation methods ii) for visualization.
- Analysis of the enormous amount of data created in simulations.
- Communication of results to policy makers in the wider public.

 © VI-SEEM consortium

4.3.2 Life Sciences applications

4.3.2.1 MS4DD – Molecular Structure for Drug Design

ACRONYM	MS4DD
APPLICATION NAME	Molecular Structure for Drug Design
MAIN DEVELOPER	Bulgarian Academy of Sciences, Institute of Molecular Biology "Roumen Tsanev", Department of Molecular Design and Biochemical Pharmacology (BAS-IMB-MDBP)
SCIENTIFIC CONTACT	Nicolay Dodoff
SCIENTIFIC SCOPE	Modeling and molecular dynamics for studying and understanding of molecular structures and properties in the light of drug design. The obtained results could contribute to the development of novel citystatic agents, and the development of non-conventional metal-based anticancer drugs.

4.3.2.2 BioMoFS – Biological molecules folding simulation

ACRONYM	BioMoFS
APPLICATION NAME	Biological molecules folding simulation
MAIN DEVELOPER	Yerevan State University, Armenia
SCIENTIFIC CONTACT	Yevgeni Mamasakhlisov
SCIENTIFIC SCOPE	Modeling and molecular dynamics targeting the deeper understanding of the mechanisms of biological molecules self - organization and in particularly to the folding - misfolding of proteins. The project will contribute to the understanding of molecular mechanisms of the neuro degenerative diseases and, thus to the improvement of the health care.

4.3.2.3 The Molecular Dynamics Simulation of Mixed Systems

ACRONYM	MDSMS
APPLICATION NAME	The Molecular Dynamics Simulation of Mixed Systems
MAIN DEVELOPER	The International Scientific-Educational Center of NAS RA, Bioinformatics Department The Institute for Informatics and Automation Problems of NAS RA
SCIENTIFIC CONTACT	Armen Poghosyan and Wahi Narsisian
SCIENTIFIC SCOPE	Modeling and molecular dynamics to study the mechanism of interaction of various proteins and surfactants and to receive the pathways on the dynamical peculiarities of mentioned mixed systems. The results of the study will make an important contribution to basic researches in Biophysics.

 © VI-SEEM consortium

4.3.2.4 DICOM Network

ACRONYM	DICOM
APPLICATION NAME	DICOM Network
MAIN DEVELOPER	Institute of Emergency Medicine, Section Imagistics, Russia Institute of Emergency Medicine IT Departament, Moldova
SCIENTIFIC CONTACT	Diana Zagadailo, Natalia Golubev
SCIENTIFIC SCOPE	Application aims on distributing medical image sets. It is expected that "DICOM Network" will help doctors and medical researched with processing and comparing investigation results. As a result research community will have access to a large number of investigation archive databases, that will offer possibility for making decisions based on experience database.

4.3.2.5 Conventional vs Novel Computer Assisted Drug Delivery

ACRONYM	CNCADD
APPLICATION NAME	Conventional vs Novel Computer Assisted Drug Delivery
MAIN DEVELOPER	Institute of pharmaceutical technology Center of pharmaceutical nanotechnology Faculty of pharmacy, UKIM
SCIENTIFIC CONTACT	Marija Glavash Dodov
SCIENTIFIC SCOPE	Comparing conventional with novel models for computer assisted drug delivery simulation, using methodologies including molecular dynamics, statistical physics, and Monte Carlo.

4.3.2.6 Protein-Small-Organic-Molecules-Interaction

ACRONYM	PSOMI
APPLICATION NAME	Protein-Small-Organic-Molecules-Interaction
MAIN DEVELOPER	Faculty of Natural Sciences, University of Montenegro
SCIENTIFIC CONTACT	Miljan Bigovic
SCIENTIFIC SCOPE	Connect pure theoretical and practical organic chemistry research with practical application and usage of newly synthesized organic molecules. So far newly synthesized molecules or group of molecules have never been tested for biological activity. The results of research will be of great importance for the understanding of ligand-receptor in simulated “live” system

 © VI-SEEM consortium

4.3.2.7 *Bioinformatics Sequence prediction Algorithms Compared With Infrared Spectroscopy Experimental Data for Proteins Secondary Structure Optimization"*

ACRONYM	SQP-IRS
APPLICATION NAME	"Bioinformatics Sequence prediction Algorithms Compared With Infrared Spectroscopy Experimental Data for Proteins Secondary Structure Optimization"
MAIN DEVELOPER	SESAME (Synchrotron–light for Experimental Science and Application in the Middle East)
SCIENTIFIC CONTACT	Gihan Kamel, and Salman Matalgah
SCIENTIFIC SCOPE	More than 80% of the known protein structures have been solved using Macromolecular Crystallography technique that can be used in any Synchrotron light facility. Since the protein structure determination process is a long and complicated one, another approach is mandatory to help solving structures in a faster but still a very reliable way, using computational algorithms coupled with simpler and feasible experimental techniques to double check the ultimate result. For this, having a comparative and/or complementary study of Infrared microspectroscopy that is a label-free technique requiring almost no prior sample preparation, together with bioinformatics software will shorten and simplify the existing procedures.

4.3.2.8 *SEE monogenic and complex disease database*

ACRONYM	SEMaCD
APPLICATION NAME	SEE monogenic and complex disease database
MAIN DEVELOPER	Institute of molecular genetics and genetic engeneering Belgrade
SCIENTIFIC CONTACT	Djordje Francuski
SCIENTIFIC SCOPE	Develop a database with genotype and phenotype data about patients with specific disorders which would benefit both researchers and MDs facilitating diagnostics and disease management and therapy

4.3.2.9 *MD simulations of biomolecules*

ACRONYM	MD-Sim
APPLICATION NAME	MD simulations of biomolecules
MAIN DEVELOPER	Biomedical Research Foundation, Academy of Athens
SCIENTIFIC CONTACT	Zoe Cournia
SCIENTIFIC SCOPE	Modeling and Molecular Dynamics (MD) study of identified drug targets and computer-aided drug design. Simulations of biomolecules important in health and disease.

 © VI-SEEM consortium

4.3.2.10 *Thermodynamic stability of DNA/DNA and RNA/DNA duplexes of entire genomes of eukaryotic organisms*

ACRONYM	THERMOGENOME
APPLICATION NAME	Thermodynamic stability of DNA/DNA and RNA/DNA duplexes of entire genomes of eukaryotic organisms
MAIN DEVELOPER	Institute of Molecular Biology, Bulgaria
SCIENTIFIC CONTACT	Stoyno Stoynov
SCIENTIFIC SCOPE	Study on how thermodynamic pattern of the genome influence fundamental processes of the cell such as transcription and RNA processing. Results could show why particular mutations alter RNA processing and lead to genetic diseases.

4.3.2.11 *Classification of cancer cells*

ACRONYM	CCC
APPLICATION NAME	Classification of cancer cells
MAIN DEVELOPER	Computer and Information Technology, Jordan
SCIENTIFIC CONTACT	Mohammad Alfraheed
SCIENTIFIC SCOPE	Develop an advanced approach for x-ray images in the context of signal processing, enabling medical users to diagnosis automatically cancer cells.

4.3.2.12 *Next Generation Sequencing Pipeline*

ACRONYM	NGS1
APPLICATION NAME	Next Generation Sequencing pileline
MAIN DEVELOPER	Cyprus Institute of Neurology and Genetics
SCIENTIFIC CONTACT	Athina Theodosiou
SCIENTIFIC SCOPE	Apply NGS technology and pipeline to address the identification of genetic mutations that cause rare diseases in families and of genetic variants that contribute to complex diseases such as autism and cancer. Moreover potential users can use the pipeline with other non-human data-sets to identify variants.

4.3.2.13 *Next Generation Sequencing Data Analysis*

ACRONYM	NGS2
APPLICATION NAME	Next Generation Sequencing Data Analysis
MAIN DEVELOPER	Translational Genetics Team - Cyprus Institute of Neurology and Genetics
SCIENTIFIC CONTACT	Petros Mina
SCIENTIFIC SCOPE	Characterization of biomarkers using NGS data targeting improvement of non-invasive prenatal diagnosis methods and detection methods of genetic abnormalities in a non-invasive assay.

 © VI-SEEM consortium

5 Conclusions

This deliverable is the main output of Task 5.1 "Refinement of service requirements and technical assessment for integration", and provides the implementation plan for the VI-SEEM VRE platform services. It entails a detailed work plan for the development and integration of domain-specific services, datasets, tools and codes into the VRE platform.

The scientific applications selected with the aid of the user survey, will enable the development and integration to the VRE platform of three generic services and a service category that will contain domain-specific application services. In particular, the VRE platform will offer a scientific application environment, a repository for workflows, pipelines and software tools, regional community datasets and application level services for the regional communities. Through these services, the VI-SEEM VRE will be offering more than 25 codes and tools for simulations and post-processing of data, more than 20 datasets, and more than 10 domain specific application services.

The development and integration of these services will be facilitated through the continuous VI-SEEM Development Access scheme, which will allow developers access to the VI-SEEM infrastructure, in order to develop and integrate their services to the VRE platform. Through the open calls, the production users will be invited to offer the output of their applications in the form of datasets, for use from future production users.

 © VI-SEEM consortium

6 Appendix: The User Survey

VI-SEEM User Survey
* Required

A. General Information
1. Application's Name*

2. Acronym*

3. The scientific research community it belongs*
- ☐ Climate
- ☐ Life Sciences
- ☐ Cultural Heritage

4. Contact person (scientific contact)
- 4.a Name and Surname *
- 4.b Affiliated Institution and Department *
- 4.c E-mail *

5. Dataset manager (data curator)
- 5.a Name and Surname
- 5.b Affiliated Institution and Department:
- 5.c E-mail

6. Main developers, present and foreseen co-developers

B. Application Details
1. Please specify in which of the following categories your application belongs to (Climate) *
- ☐ Regional Climate Modelling
- ☐ Global Climate Modelling
- ☐ Weather Forecasting
- ☐ Air Pollution/Quality
- ☐ Model Development
- ☐ Visualization, Datasets, etc.
- ☐ Modelling and Molecular Dynamics (MD) study of important drug targets
- ☐ Computer-aided drug design
- ☐ Analysis of Next Generation DNA sequencing data and RNA profiling data
- ☐ Data mining to identify prevalent diseases/mutations in the SEEM region
- ☐ Image processing for biological applications
- ☐ Computational simulation of DNA and RNA
- ☐ Synchrotron data analysis
- ☐ Digital Libraries
- ☐ Interactive Visualization Tools
- ☐ Semantic Referencing
- ☐ Image Classification
- ☐ Modelling of Built Environments and Advanced Representation Techniques
- ☐ Scientific simulation of materiality and systems' properties
- ☐ Geo-referencing Tools
- ☐ Bioarchaeology

 © VI-SEEM consortium

☐ Other:

2. Expected scientific impact *
Effect on the scientific methods, methodologies, research and results, e.g., investigation of novel physical phenomena, processing big damaged images, heritage management, art characterization

3. Expected societal impact *
Effect on the social structure of the community and well-being of the individuals, e.g., improved social services, new or better daily-life products, increased welfare, tourism

4. Problem solved by application *
Please describe the problems solved in contrast with the existing state of the art. What is the novelty of the expected results

5. Do you have existing datasets that you will use with your application and are you willing to share with the VI-SEEM community for research purposes? If yes, please specify. *

6. Do you expect to produce datasets to be shared with the VI-SEEM community for research purposes? If yes, please specify. *

7. What output of your application can be provided as service to the VI-SEEM community? *
For example workflow engine, web UI, consultation services, datasets etc. Please provide as much detail as possible.

8. Expected user community size *
Those who interact with the service by generating inputs, submitting, interacting with the application, or retrieving results; not those who use the result published on the web or in media

☐ 1-5
☐ 5-20
☐ 20-50
☐ 50-100
☐ 100-200
☐ 200+

9. Expected user community *
Example given EMAC model users community, archaeologists, historians, librarians, educators, engineers and programmers working in DCH, etc.

10. How long is needed before application becomes usable? *

☐ Under a month
☐ 1 month - 2 months
☐ 2 months - 6 months
☐ 6 months – 1 year
☐ 1 year - 18 months
☐ 18 months - 2 years

C. Collaborations
Ideally, each application is expected to collaborate with other partners of the VI-SEEM project in various phases. These phases especially include idea development, testing, using and producing scientific results and datasets.

1. Collaborations already set (specify if within VI-SEEM or external)

 © VI-SEEM consortium

Please specify the collaborator, scope and means of collaboration for each

2. Collaborations planned for the future (specify if within VI-SEEM or external)
Please specify the collaborator, scope and means of collaboration for each

3. Please describe potential cross-disciplinary collaborations and links to other user-communities

D. Licensing and Availability
1. Information regarding licensed software (libraries, tools, etc.) the application requires. If any please indicate whether the license is already obtained or not.

2. Please specify if there are any specific copyright requirements related to the application and the data used or produced, as well as of the services involved. Specifically comment regarding your own developed source code, input data, output data and any licensing or IP restrictions you might implement.

3. Please specify if there are any ethical issues pertaining to the application as defined by the EU regulations (e.g., removal of personal identification information from life science datasets).

E. System Requirements
1. System of preference (you can select more than one option if needed) *
- ☐ HPC infrastructure cluster
- ☐ Grid infrastructure cluster
- ☐ Cloud infrastructure VMs

2. HPC infrastructure cluster requirements

2.a Number of CPU cores needed in single job run
e.g. 128 CPU cores

2.b Amount of memory required per CPU core
e.g. 2GBs of RAM per CPU core

2.c Expected wall clock time for a single job run
e.g. 12 hours

2.d Expected number of job runs
e.g. approximately 100 runs

2.e Total number of core hours needed e.g. 700,000 CPU-hours. This number is calculated as explained in the following example: For one job run that requires 512 cores (i.e. MPI, OpenMP, or hybrid models), with a wall clock time of 10 Hours and assuming you need 100 such jobs runs, then your total requirements in core hours are: 512 x 10 x 100 = 512.000 (0.5 Million) core hours.

2.f Required accelerator type
- ☐ None
- ☐ GP-GPU
- ☐ IBM PowerXCell
- ☐ Intel Xeon Phi
- ☐ Other:

2.g Amount of hard disk space needed for single job run
e.g. 10 GB of temporary storage space

 © VI-SEEM consortium

2.h Amount of long-term data storage
e.g. 1TB of longer term (for the duration of the project) storage space

2.i Interconnection, latency, bandwidth requirements
e.g. Application performance over InfiniBand scales as cluster size.

3. Grid infrastructure cluster requirements
3.a Number of CPU cores needed in single job run
e.g. 8 CPU cores

3.b Amount of memory required per CPU core
e.g. 2GBs of RAM per CPU core

3.c Expected wall clock time for a single run
e.g. 12 hours

3.d Expected number of job runs
e.g. approximately 100 runs
3.e Amount of hard disk space needed for single run
e.g. 10 GB of temporary storage space

3.f Amount of long-term data storage
e.g. 1TB of permanent storage space

4. Cloud infrastructure requirements
4.a Number of VMs needed for application deployment
e.g. 5 VMs

4.b Number of CPU cores per VM
e.g. 2 CPU cores per VM

4.c Amount of memory per VM
e.g. 4GBs of RAM per VM

4.d Amount of storage space per VM
e.g. 200 GB

5. Internet connection requirements
e.g. 200 GB

6. Operating systems compatible with your application .
☐ Linux
☐ Unix
☐ MS Windows
☐ Sun
☐ Mac OS X
☐ Other:
7. Deployment Information
7.a Used programming languages
☐ C/C++
☐ Fortran
☐ Java
☐ Perl
☐ Python
☐ Other:

 © VI-SEEM consortium

7.b Supported and compatible compilers
- ☐ GNU compiler collection
- ☐ Intel compiler suite
- ☐ Portland group compiler
- ☐ Oracle compiler suite
- ☐ Matlab compiler
- ☐ Other:

7.c Parallel programming platform
- ☐ SMP (ex. using OpenMP, Pthreads)
- ☐ Clustered multiprocessing (ex. using MPI)
- ☐ GPU (ex. using CUDA, OpenCL)
- ☐ Multiple serial jobs (data-splitting, parametric studies)
- ☐ Other:

7.d Libraries and tools used by application
e.g. The application depends on the following libraries: NAMD, NetCDF, OpenMP, MPICH-2, and FFTW.

7.e Is there a need for an application-specific monitoring, logging, tracking or statistical analysis of application-generated events?

7.f Can the application benefit from some VO/domain-specific commonalities that could be provided as services?

7.g Other requirements, suggestions or comments

F. Data Services

1. For the data sets that you will either collect, generate or use provide a name and a description. Description should include data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful. Include information on the existence (or not) of similar data and the possibilities for integration and reuse.

2. What are the standards and metadata used in your scientific discipline? If these do not exist provide an outline on how and what metadata will be created.

3. Describe how data will be shared providing information about access procedures and technical mechanisms for dissemination and re-use of the data, and define whether data will be widely open or restricted to specific groups. Is there any embargo periods of data accessibility? If the dataset cannot be shared (for a limited time or permanently, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

4. What type of storage service would you require?
- ☐ Simple storage for file sharing i.e. documents exchange, small data sets.
- ☐ Long term archiving system
- ☐ Data repositories
- ☐ In situ interactive analysis of production data

5. What are the data preservation requirements? Do you need backup / archival storage? How much? How long would you like to preserve your data for?

6. Do you require persistent identifiers for your data?

7. When would you provide a PID to your data set? What kind of persistent ID is used for identifying data?

 © VI-SEEM consortium

8. Would you like make your data available, discoverable, intelligible and searchable? What metadata are you using?

9. Total required storage space. In order to better assess your storage requirements please respond to the following questions.

9.a What is initial amount of the temporary storage you will need during a single run?

9.b What is the full-scale amount of the temporary storage you will need during a single run?

9.c What is the initial amount of storage you will need for the installation of the application?

9.d What is the full-scale amount of storage you will need for the installation of the application?

9.e What is the initial amount of long-term storage you will need?

9.f What is the full-scale amount of long-term storage you will need?

9.g What is the minimum duration for the initial long-term storage?

9.h What is the minimum duration for the full-scale long-term storage?

10. Technical requirements of data access

10.a What is the minimum I/O ratio?

10.b What is the minimum data throughput?

11. Data security requirements

11.a In terms of access control what are you requirements?

11.b In terms of encryption what are you requirements?

11.c In terms of anonymity what are you requirements?

11.d In terms of privacy policy what are you requirements?

11.e In terms of replication what are you requirements? How many copies if any?

11.e Is off-site replication needed?

☐ Yes

☐ No

11.f Is validity checking needed?

☐ Yes

☐ No

12. Are there any data location preferences or legal constraints (e.g. in-country location, etc.) that you have? Please specify location preferences together with importance (preferred/important/mandatory).

13. Is a sophisticated file / data indexing / identification mechanism that cannot be achieved using standard directories and file naming conventions required (e.g., geographical proximity search, matching of application-dependent metadata attributes)? If yes, please describe?

 © VI-SEEM consortium

14. Is a workflow management system required? If yes, please specify.

15. Please describe the required data workflow(s) if needed. (e.g. data processing, conversion on storage side, data mining or analysis)

G. Interactions/Communication

1. What is the type of interaction between the application and the external environment?

- ☐ Using GUI in batch mode
- ☐ Using GUI in interactive mode
- ☐ Using GUI in real-time mode
- ☐ Using command line in batch mode
- ☐ Using command line in interactive mode
- ☐ Using command line in real-time mode
- ☐ Other:

H. Training and Support

1.a Mark your need for training regarding the general use of available infrastructure:

	No Training Needed	Basic training	Intermediate training	Advanced training
Grid	☐	☐	☐	☐
HPC	☐	☐	☐	☐
GPU	☐	☐	☐	☐
Other	☐	☐	☐	☐

2. Please select the type of training in general cloud and big data tools that you would like to receive.

- ☐ Not needed
- ☐ Basic general training
- ☐ Intermediate general training
- ☐ Advanced specialized training

3.a Mark your need for training for the different offered open source tools for Big Data:

	No Training Needed	Basic training	Intermediate training	Advanced training
Hadoop	☐	☐	☐	☐
MapReduce	☐	☐	☐	☐
Storm (Hadoop of real time)	☐	☐	☐	☐
Cassandra (NoSQL)	☐	☐	☐	☐
MongoDB (NoSQL)	☐	☐	☐	☐
OrientDB (NoSQL)	☐	☐	☐	☐
SpagoBI	☐	☐	☐	☐
Pentaho	☐	☐	☐	☐
Jaspersoft	☐	☐	☐	☐
RapidMiner/RapidAnalytics	☐	☐	☐	☐
Mahout	☐	☐	☐	☐
Orange	☐	☐	☐	☐
KEEL	☐	☐	☐	☐

 © VI-SEEM consortium

	No Training Needed	Basic training	Intermediate training	Advanced training
Rattle (R based)	☐	☐	☐	☐
Pig	☐	☐	☐	☐
R	☐	☐	☐	☐
ECL	☐	☐	☐	☐
Lucene	☐	☐	☐	☐
Solr	☐	☐	☐	☐
Other	☐	☐	☐	☐

3.b If your answer in 3.a was Other please specify.

4. Do you require training in:

	No Training Needed	Introductory	Advanced	Expert
MPI	☐	☐	☐	☐
OpenMP	☐	☐	☐	☐
CUDA/OpenCL	☐	☐	☐	☐
Numerical Libraries	☐	☐	☐	☐
Programming Languages	☐	☐	☐	☐
Compilers	☐	☐	☐	☐
Visualization	☐	☐	☐	☐
Access to infrastructure	☐	☐	☐	☐
Profiling/Optimization	☐	☐	☐	☐
Data	☐	☐	☐	☐
Domain specific	☐	☐	☐	☐
Molecular Dynamics simulations	☐	☐	☐	☐
Computer-aided drug design	☐	☐	☐	☐
Image analysis	☐	☐	☐	☐
Linux/Unix	☐	☐	☐	☐
Data mining	☐	☐	☐	☐
Analysis of next generation sequencing data	☐	☐	☐	☐
Statistical Analysis	☐	☐	☐	☐
Databases	☐	☐	☐	☐

5. Please specify if you need any advanced user support?

	Yes	NO
MPI	☐	☐
OpenMP	☐	☐

 © VI-SEEM consortium

	Yes	NO
CUDA/OpenCL	☐	☐
Numerical Libraries	☐	☐
Programming Languages	☐	☐
Compilers	☐	☐
Visualization	☐	☐
Access to infrastructure	☐	☐
Profiling/Optimization	☐	☐
Data	☐	☐
Linux/Unix	☐	☐

7. Please give details of any additional required support.

I. Domain Specific Questions

1. Please select your scientific research community
   - ☐ Cultural Heritage
   - ☐ Climate
   - ☐ Life Sciences

1. Cultural Heritage

1.a Describe the type of activity (e.g., digitization, processing, post-production, storage, ontology engineering, data-mining, data-analytics, visualization, dissemination/publication, etc.)

1.b Please specify your interest in the following activities (1-Strong, 6-Weak)

	1	2	3	4	5	6
Dataset design and implementation	☐	☐	☐	☐	☐	☐
Integration and interoperability of datasets	☐	☐	☐	☐	☐	☐
Multimedia management in archaeological datasets	☐	☐	☐	☐	☐	☐
Metadata mapping and interoperability	☐	☐	☐	☐	☐	☐
Design of scientific datasets	☐	☐	☐	☐	☐	☐

 © VI-SEEM consortium

	1	2	3	4	5	6
Modelling and Simulation	☐	☐	☐	☐	☐	☐
3D documentation / shape analysis	☐	☐	☐	☐	☐	☐

2.a Data access and management. Please describe legacy data from past field research you intend to re-use within VI-SEEM

2.b Please specify data management required

	Yes	No
Management of legacy data	☐	☐
New dataset creation	☐	☐
Management of datasets deriving from new research activity	☐	☐
Management of datasets deriving from new documentation methods (e.g., 3D)	☐	☐

3. Resources. Please describe resources currently used.

3.a Database analytics

3.b Figures/logistics of assets involved, like artifacts, images, videos, reports, etc.

3.c Metadata available.

4. Please describe the codes you are using

4.a Data conversion tools

4.b Mapping tools

4.c Formal languages

4.d Version control system for software development

5. Models and Methods used

6. Choose any Framework / Computer Language / Technologies required

	Yes	No
PHP	☐	☐
MySQL	☐	☐
Java Script	☐	☐
CSS	☐	☐
HTML	☐	☐
Java	☐	☐

 © VI-SEEM consortium

Python	☐	☐
Apache solr lucene	☐	☐
Content management system	☐	☐
Linux centos	☐	☐

7.a Mark using increasing numbers (1 – highest and 8 - lowest) the need for training in the following.

	1	2	3	4	5	6	7	8
Advanced content management systems	☐	☐	☐	☐	☐	☐	☐	☐
Data repositories	☐	☐	☐	☐	☐	☐	☐	☐
Visualization	☐	☐	☐	☐	☐	☐	☐	☐
Computational intensive workflows and querying	☐	☐	☐	☐	☐	☐	☐	☐
Content-based image retrieval (CBIR)	☐	☐	☐	☐	☐	☐	☐	☐
Image processing and machine learning algorithms	☐	☐	☐	☐	☐	☐	☐	☐
GPU based algorithms for image classification	☐	☐	☐	☐	☐	☐	☐	☐
Content-based image retrieval (CBIR)	☐	☐	☐	☐	☐	☐	☐	☐
Other	☐	☐	☐	☐	☐	☐	☐	☐

7.b If your answer in 7.a was Other please specify.

8.a Mark your need for training for the different offered tools, codes and models.

	No Training Needed	Basic training	Intermediate training	Advanced training
Osteoware	☐	☐	☐	☐
MEDICI	☐	☐	☐	☐
Algorithms for remote sensing image classification	☐	☐	☐	☐
Algorithm for aerial image classification	☐	☐	☐	☐
ImaNote	☐	☐	☐	☐
Soft Ontology Layer (SOL)	☐	☐	☐	☐
idPromo	☐	☐	☐	☐
InViTo	☐	☐	☐	☐
3DINV	☐	☐	☐	☐
AutoGRToolkit version 3.2	☐	☐	☐	☐
DAR	☐	☐	☐	☐

 © VI-SEEM consortium

Osteoware	☐	☐	☐	☐
Other	☐	☐	☐	☐

8.b If your answer in 8.a was Other please specify.

8. Give details of any other domain-specific training you require.

9. What domain-specific training can you provide to the community?

10. Please list all of the above-mentioned models, tools and codes for which you are able and willing to provide training for the other project members.

2. Climate

1. Which models are you using
   - ☐ WRF
   - ☐ WRF-CHEM
   - ☐ EMAC
   - ☐ RegCM
   - ☐ MM5
   - ☐ DREAM
   - ☐ Other:

2. What input/boundary conditions data do you require to be staged/hosted for simulations/runs on remote systems?
e.g. GFS boundary conditions to be downloaded twice-daily for weather forecasting with WRF & MOZART data for WRF-Chem. Data freely available from NCAR. Can provide download script.

3. Do you plan/foresee collaboration with other partners within VI-SEEM. Please identify partners and give details on proposed framework and specific outcomes.

4. Will you be conducting model code development activities? Please provide details e.g. GPU development of KPP atmospheric chemistry package. Required parallel NetCDF libraries, Totalview debuggers and CUDA compilers.

5. Please provide, to the extent possible, information on proposed or planned publications based completely or in part on the VI-SEEM participation & resources. Include title/theme, collaborators, targeted journal and submission time, VI-SEEM involvement/contribution.

6. Which post/processing and visualization tools do you require?
   - ☐ GRADS
   - ☐ Paraview
   - ☐ FERRET
   - ☐ GIS (ARGIS/ENVI/Other) (please specify)
   - ☐ NCL
   - ☐ R (please list specific packages)
   - ☐ Python
   - ☐ IDL

7. What datasets will you be contributing for the VISEEM climate community? Please provide details (Creator, model if applicable, size, format, lisence-ownership, hosting)
e.g. Self-generated, Gridded meteorological and hydrological data for Israel, and station observations for PM10 and PM2.5 measurements in 25 locations in Israel, single NetCDF file, 52 MB, CC-By lisence, data hosted in private server but available for mirroring.

 © VI-SEEM consortium

8. Do you have specific hardware and/or software requirements for your application that was not listed already in the questionnaire?

9.a Mark your need for training for the different offered climate and weather forecasting models.

	No Training Needed	Basic training	Intermediate training	Advanced training
WRF	☐	☐	☐	☐
WRF-Chem	☐	☐	☐	☐
ECHAM (EMAC)	☐	☐	☐	☐
RegCM	☐	☐	☐	☐
MM5	☐	☐	☐	☐
DREAM	☐	☐	☐	☐
Other	☐	☐	☐	☐

9.b If your answer in 9.a was Other please specify.

10.a Mark your need for training for the different offered visualization and GIS software:

	No Training Needed	Basic training	Intermediate training	Advanced training
GrADS	☐	☐	☐	☐
ParaView	☐	☐	☐	☐
FERRET	☐	☐	☐	☐
ArcGIS	☐	☐	☐	☐
ENVI	☐	☐	☐	☐
LAS (Live Access Server)	☐	☐	☐	☐
Other	☐	☐	☐	☐

10.b If your answer in 10.a was Other please specify.

11.a Mark your need for training for the different offered tools:

	No Training Needed	Basic training	Intermediate training	Advanced training
NCL (NCAR Command Language)	☒	☐	☐	☐
R	☐	☐	☐	☐
Python	☐	☐	☐	☐
IDL	☐	☐	☐	☐
Other	☐	☐	☐	☐

11.b If your answer in 11.a was Other please specify.

12. Give details of any other domain-specific training you require.

 © VI-SEEM consortium

13. What domain-specific training can you provide to the community?

14. Please list all of the above-mentioned models, tools and codes for which you are able and willing to provide training for the other project members.

15. General comments/suggestions for the climate community and climate coordinator:

3. Life Sciences

1.a Describe the type of activity (e.g., meta-analysis, data production, database management, storage, ontology engineering, data-mining, data-analytics, visualization, statistical analysis, dissemination/publication, etc.):

1.b Please specify your participation in the following activities (1-Strong, 6-Weak)

	1	2	3	4	5	6
Molecular Dynamics simulations	☐	☐	☐	☐	☐	☐
Computer-aided drug design	☐	☐	☐	☐	☐	☐
Image analysis	☐	☐	☐	☐	☐	☐
Data mining	☐	☐	☐	☐	☐	☐
Analysis of next generation sequencing data	☐	☐	☐	☐	☐	☐
Statistical Analysis	☐	☐	☐	☐	☐	☐
Analysis of synchrotron data	☐	☐	☐	☐	☐	☐

2. Do you plan to contribute new programming tools/scripts for the analysis or meta-analysis of

	Yes	No
MD simulation trajectories	☐	☐
Data mining	☐	☐
Computer-aided drug design	☐	☐
Bioimage processing and toolkits	☐	☐
Next generation sequencing or profiling data	☐	☐

3.a Which datasets will you be contributing to the VI-SEEM life sciences community? Please provide details (Creator, model if applicable, size, format, license-ownership, hosting)
e.g MD trajectories, virtual screening results, Genotyping for analysis of dominant cancerous mutations and for overexpressed oncogenic proteins, databases, in vitro results, in vivo results, bioimaging librarie, bioimaging databases, ontologies, public image repository etc.

3.b Please specify data management required

- ☐ Management of legacy data
- ☐ New dataset creation
- ☐ Management of datasets deriving from new research activity
- ☐ Management of datasets deriving from new documentation methods

 © VI-SEEM consortium

4. Do you plan/foresee collaboration with other partners within VI-SEEM. Please identify partners and give details on proposed framework and specific outcomes.

5. Will you be conducting code development activities? Please provide details.
e.g. GPU development of MD simulation package. Required parallel NetCDF libraries, Totalview debuggers and CUDA compilers.

6. Please provide, to the extent possible, information on proposed or planned publications based completely or in part on the VI-SEEM participation & resources. Include title/theme, collaborators, targeted journal and submission time, VI-SEEM involvement/contribution.

7. Which post/processing and visualization tools and libraries do you require?
Check all that apply.

- ☐ Maestro
- ☐ VMD
- ☐ Pymol
- ☐ Rasmol
- ☐ R
- ☐ Python
- ☐ Perl
- ☐ Tcl
- ☐ Matlab
- ☐ Velocity
- ☐ ImageJ
- ☐ Metamorph
- ☐ Image-Pro
- ☐ LabView
- ☐ ScanImage
- ☐ VTK
- ☐ ITK
- ☐ FFTW
- ☐ ImgLib
- ☐ Other:

8. Which production programs will you need installed

- ☐ NAMD
- ☐ GROMACS
- ☐ AMBER
- ☐ Maestro
- ☐ Desmond
- ☐ ParaView
- ☐ Vaa3D
- ☐ OpenCV
- ☐ Other:

9. Which file formats you will be using

- ☐ .netCDF
- ☐ .trr
- ☐ .dcd
- ☐ .gro
- ☐ .pdb
- ☐ .sdf
- ☐ .tif
- ☐ .xtc
- ☐ Other:

 © VI-SEEM consortium

10. Which framework / Computer Language / Technologies are required

- ☐ PHP
- ☐ MySQL
- ☐ Java Script
- ☐ CSS
- ☐ HTML
- ☐ Java
- ☐ Python
- ☐ Apache solr lucene
- ☐ Content management system
- ☐ Linux centos
- ☐ Other:

11. Do you have specific hardware and/or software requirements for your application that was not listed already in the questionnaire?

12.a Mark using increasing numbers (1 – highest and 8 - lowest) the need for training in the following data analysis methods

	1	2	3	4	5	6	7	8
Capture	☐	☐	☐	☐	☐	☐	☐	☐
Curation	☐	☐	☐	☐	☐	☐	☐	☐
Analysis	☐	☐	☐	☐	☐	☐	☐	☐
Search	☐	☐	☐	☐	☐	☐	☐	☐
Sharing	☐	☐	☐	☐	☐	☐	☐	☐
Visualization	☐	☐	☐	☐	☐	☐	☐	☐
Other	☐	☐	☐	☐	☐	☐	☐	☐

12.b If your answer in 12.a was Other please specify.

13.a Mark your need for training for the different offered tools:

	No Training Needed	Basic training	Intermediate training	Advanced training
ChemBioserver	☐	☐	☐	☐
PepServer	☐	☐	☐	☐
NAMD	☐	☐	☐	☐
GROMACS	☐	☐	☐	☐
PLOS Text Mining Collection	☐	☐	☐	☐
BWA	☐	☐	☐	☐
Picard	☐	☐	☐	☐
GATK	☐	☐	☐	☐
Other	☐	☐	☐	☐

13.b If your answer in 13.a was Other please specify.

14. Give details of any other domain-specific training you require.

15. What domain-specific training can you provide to the community?

16. Please list all of the above-mentioned models, tools and codes for which you are able and willing to provide training for the other project members.

 © VI-SEEM consortium

17. General comments/suggestions for the life sciences community and life sciences coordinator.

Closing Questions
1. Questionnaire submitted by (Name and Surname) *

2. E-mail *

3. Work telephone number *

 © VI-SEEM consortium