H2020-EINFRA-2015-1

VI-SEEM

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

Deliverable D5.5
Final report on integrated services and the VRE platform

Author(s): Mohammed Elfarargy (editor)

Status –Version: Final - h

Date: September 30th

Distribution - Type: Public

Abstract: Deliverable D5.5 – The "Final report on the integrated services and the VRE Platform" provides a detailed description of the final domain specific services and their integration for the completion of the VRE platform.

The VI-SEEM Consortium consists of:

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

IIAP-NAS-RA	Contractor	Armenia
GRENA	Contractor	Georgia
BA	Contractor	Egypt
IUCC	Contractor	Israel
SESAME	Contractor	Jordan

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

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
17-05-2018	a	Andreas Athenodorou	Initial version of ToC
03-08-2018	b	Andreas Athenodorou	Start introducing chapters
13-09-2018	c	Andreas Athenodorou, Mohammed Elfarargy, Youssef Eldakar, Stelios Erotokritou, Bojana Koteska, George Artopoulos	Edit the script
23-09-2018	d	Andreas Athenodorou	Editing
26-09-2018	e	Youssef Eldakar, Mohammed Elfarargy	Editing
27-09-2018	f	Andreas Athenodorou, George Artopoulos, Valentina Vassallo, Lidjia Milosavljevic	Additions and corrections
28-09-2018	g	Andreas Athenodorou, Valentina Vassallo, Zoe Cournia	Editing and corrections
30-09-2018	h	Andreas Athenodorou, Mohammed Elfarargy, Ognjen Prnjat	Final editing

Table of contents

References

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

[2] VI-SEEM Deliverable D3.1: Infrastructure and services deployment plan

[3] VI-SEEM Deliverable D4.2: Description of the initial deployed data services

[4] VI-SEEM Deliverable D5.1: Detailed technical implementation plan for VRE services and tools

[5] VI-SEEM Deliverable D5.2: Data management plans

[6] VI-SEEM Deliverable D5.3: User-oriented documentation and training material for VRE services

[7] Joomla framework
https://framework.joomla.org/

[8] Bootstrap library
http://getbootstrap.com

[9] VI-SEEM code repository
https://code.vi-seem.eu

[10] Live Access Server
http://las.vi-seem.eu/las

[11] Clowder
http://dchrepo.vi-seem.eu

[12] ChemBioServer
http://bioserver-3.bioacademy.gr/Bioserver/ChemBioServer/

[13] VI-SEEM training portal
https://training.vi-seem.eu/

[14] Docker
https://www.docker.com/

[15] Google Analytics
https://analytics.google.com/analytics/web/

[16] VI-SEEM data repository
https://repo.vi-seem.eu/

[17] VI-SEEM Service Catalogue and Portfolio
https://services.vi-seem.eu/ui/catalogue/services/

[18] VI-SEEM Wiki
http://wiki.vi-seem.eu/index.php/Main_Page

[19] VI-SEEM Login Service
https://services.vi-seem.eu/ui/catalogue/services/#VI-SEEM_Login

List of Figures

List of Tables

Glossary

AMBER	Assisted Model Building with Energy Refinement molecular simulation programs
CBIR	Content-Based Image Retrieval
COSMO	Consortium for Small-scale Modelling
CPU	Central Processing Unit
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
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
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
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 and the Eastern Mediterranean
VRE	Virtual Research Environment
WEST	Wind Energy Simulation Toolkit
WRF	Weather Research and Forecasting Model

Executive summary

What is the focus of this Deliverable?

This is the final deliverable submitted by the Work Package 5. The focus of this deliverable is to provide a detailed report on all the VI-SEEM integrated services as well as the final version of the VI-SEEM Virtual Research Environment as shaped after receiving comments from the reviewers.

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

This is the final deliverable of this Work Package.

What are the deliverable contents?

The deliverable contents include an overview of the main structure of the VRE platform, and how it integrates the various types of VI-SEEM services, starting with the generic VI-SEEM services. It also provides information about the domain-specific services which are dealing with the integrated domain-specific workflows, software tools and application codes. Furthermore, it provides a list of the integrated application-level services. For each service available in the VRE, information on how a user can access the existing service and contribute to updating existing service is given; as well as instructions for including new services. Finally, a detailed list of the potential future additions in the portfolio of VI-SEEM applications (workflows, software tools, application codes and datasets), as well as the integrated cross-disciplinary applications is given. In detail, the deliverable includes:

- Description of the VI-SEEM Virtual Research Environment Platform (chapter 2).
- Description of the VI-SEEM scientific application environment as a domain specific service addressing the needs of each of the three project scientific communities and their applications (chapter 3).
- Description of the VI-SEEM workflows, software tools and codes repository as a domain-specific service addressing the needs of each of the three project scientific communities (chapter 4).
- Description of the VI-SEEM regional community datasets as a domain-specific service addressing the needs of the three project scientific communities and their data (chapter 5).
- Description of the VI-SEEM application-level services (chapter 6).
- Brief description of potential new services expected to result from the three open calls for production (chapter 7).
- Description of the VI-SEEM cross-disciplinary applications as well as of the proposed ontological solution for VI-SEEM (chapter 8).

Conclusions and recommendations

The deliverable is the main output of T5.3 "Development of the VRE platform", T5.4 "Overall integration of services" as well as T5.5 "Scientific Support". It describes the design behind the VRE platform and how this enables the user to access the generic and domain-specific VI-SEEM services. The deliverable describes the different sections of the VRE platform, illustrating how each section manages to integrate the VI-SEEM offered services. Details on how the user can easily get information and use the VI-SEEM provided services are given. Detailed instructions on how the user can integrate hers/his own applications and services to the VI-SEEM Virtual Research Environment are given.

This deliverable is the final report on the development of the VRE platform, presenting its content including all the domain specific service, future applications as well as cross-disciplinary applications. This allows users to use the current services but it also enables developers and scientists participating in open calls of VI-SEEM to make their tools, datasets and services available to the wider community.

1 Introduction

VI-SEEM brought together service providers, service enablers/experts and scientists from the three scientific communities relevant to the project, to develop the platform for the provision of a Virtual Research Environment (VRE). The VRE provides the framework and the entry point for the scientific community users, where data and analysis tools can be promoted and accessed. The data, applications and services are relevant to the three scientific communities of climate, digital cultural heritage and life sciences. These were initially non-existent or scattered around the different communities' local infrastructures, while the VI-SEEM VRE integrated them in a stable and secure regional infrastructure providing a consolidated easy access point to the users.

This document describes the final version of the integration of the generic and domain specific VI-SEEM services and the VRE platform. The VRE platform is an easy accessed and user-friendly portal that connects the unified infrastructure developed in the Work Package 3 “e-Infrastructure Services”, the generic data services deployed in the Work Package 4 “Data Management Lifecycle” and the domain specific services, tools and APIs integrated through the Work Package 5 “Domain-specific Services and Support”.

We first describe the structure and the design of the VRE platform in chapter 2. The initial form of the VRE-portal, presented in D5.4 has been reshaped according to the recommendations of the reviewers as well as the cross-disciplinary needs of the VI-SEEM project. In this chapter, we provide details on how the user can get access to the generic infrastructure and data services and how she/he can get information for the three scientific communities supported by the project. The next chapter provides more details on the domain-specific services offered by the project, as this was made possible through the VI-SEEM integration phases described by D5.1 [5]. Chapter 3 describes the VI-SEEM scientific application environment where the user can get information for optimized applications and tools available and supported in the e-Infrastructure. Such applications are available in the HPC or Grid sites provided by the project partners. Additionally, the service provides ready-to-use Virtual Machine images that the user can instantiate to familiarize herself and run scientific community specific tools in the regional Cloud infrastructure. Chapter 4 describes the workflow and software tools provisioning services where the user can get access to manual or automated scientific workflows and scientific codes provided by the VI-SEEM communities. Chapter 5 describes the regional community datasets domain specific service where the user can access available scientific datasets. Chapter 6 describes all eleven services grouped together under the VI-SEEM application-level services umbrella. Such application-level services form a range of web-based or visualization services providing easy access to underlying workflows, applications and resources for each regional community. For each of the available services details on how a user can access the available material through the VRE platform are provided as well as details on how a user can contribute to the VRE by providing new software modules, workflows and datasets and new application-level services. Furthermore, in chapter 7 we describe possible forthcoming updates to the VI-SEEM domain specific services, as these could result from the output of projects participating in the three VI-SEEM open calls for production. Finally, in chapter 8 we describe all the integrated cross-disciplinary VI-SEEM applications, in other words material of domain specific services such as workflows, software tools, and datasets as well as application level services, which bridge the three distinct scientific disciplines. Furthermore, we describe the proposed ontological solution for VI-SEEM datasets in order to achieve their interoperability.

2 VI-SEEM Virtual Research Environment platform

In this section, we provide a description of the final version of the Virtual Research Environment platform, which was finalized in M36 of the project. We describe the design of the VRE platform and how it integrates the VI-SEEM domain specific services. VRE platform as this was described in D5.4 has received updates and changes after the interim review, which took place in May 2017, as well as the light review which took place in May 2018. Reviewers provided suggestions as well as recommendations for enhancing the user-friendliness of the Virtual Research Environment. By incorporating all the suggested improvements, the Virtual Research Environment, has an enhanced easy-access and user-friendly character.

2.1 Scope

The VRE portal provides access to the VRE services and applications. The portal connects the unified infrastructure developed in WP3, the generic data services deployed in WP4 and hosts the user-tailored front-end interface. It has a user-friendly structure which enables researchers, students but also SMEs' representatives to easily access and navigate throughout the portal. It also enables easy communication and data sharing between the various user groups. The VI-SEEM VRE portal is available at http://vre.vi-seem.eu/.

2.2 VRE platform structure

In the following subsections, we provide a description of the appearance of the VRE portal. This section is organized in six different subsections, namely:

- **The header** which provides access to all the domain specific services.
- **The User Zone** which is tailored in such a way in order to reflect the needs of a particular target group, more specifically, researchers, students and SMEs.
- **The Support to the three scientific disciplines** which focusses on the needs of scientists working on one of the three VI-SEEM supporting disciplines.
- **The Cross-Disciplinary fields** which focusses on the cross-disciplinary domain specific applications of the Virtual Research Environment since VI-SEEM has an enhanced cross-disciplinary character.
- **The VI-SEEM Sites** which provides an easy access to important VI-SEEM pages.
- **Infrastructure Projects** which provides links to other Infrastructure Projects which might be of the interest of the user visiting the VRE.

At the bottom of the VRE a user can find a link to a guide on the usage of the VRE-portal as well as a link for providing feedback.

A screenshot of the VRE portal is presented in Figure 1.

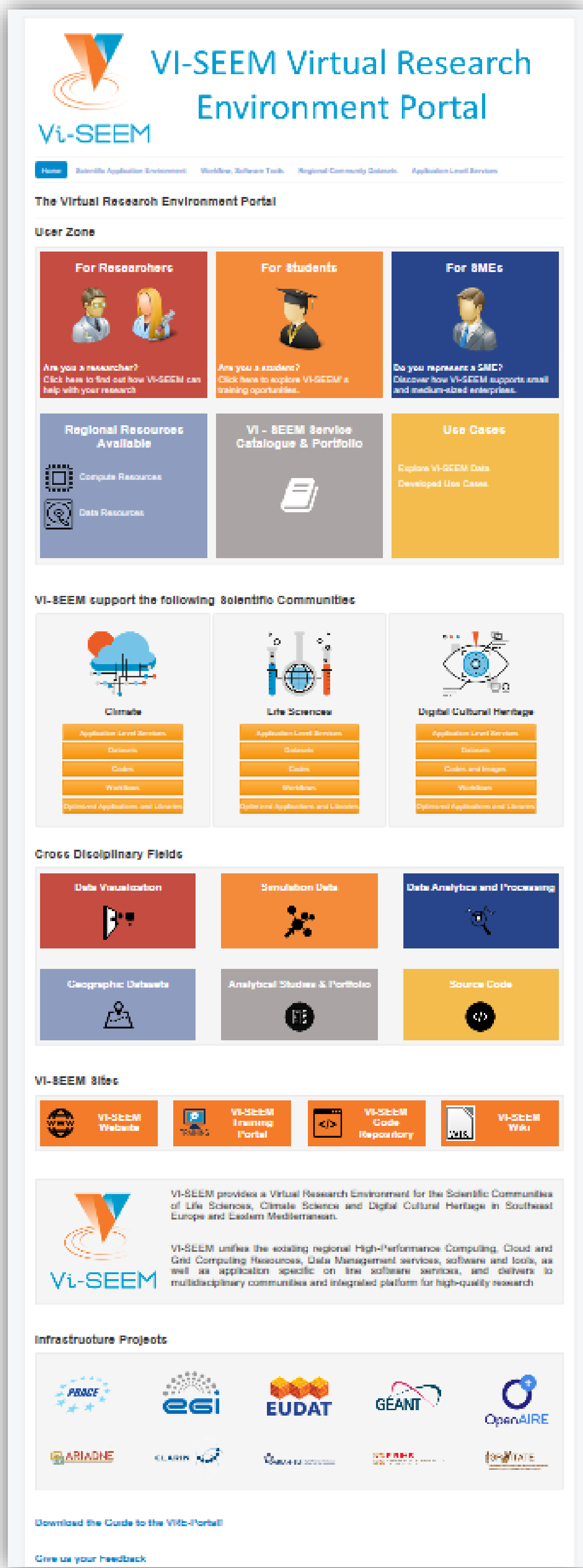


Figure 1: A scroll-down of the Virtual Research Environment portal

Header

The header of the VRE-portal as this appears in Figure 2 provides the following tabs

Figure 2: A screenshot of the VRE header

1. **Scientific Application Environment**

The scientific application environment services provides the list of **Service Enablers**, also contains links to the climate, cultural heritage, and life science **Applications and Libraries**, **Virtual Machine (VM) images** and **Software developed use cases**. Furthermore, it provides guidelines on **how to contribute** to scientific application environment. The above sections are described in more detail below:

- **How to contribute**: provides detailed information on how the user can contribute to the environment, making software, use cases and virtual machines (VMs) available to the community.
- **Optimized application and libraries**: lists the application and libraries available to the VI-SEEM integrated infrastructure services, along with documentation and training material.
- **Virtual Machine images**: provides access to virtual machine images tailor made for specific scientific communities.
- **Developed use cases**: provides information on how different communities utilize the VI-SEEM integrated environment for developing specific scientific use cases.

A screenshot of the scientific application environment is provided in Figure 3.

Figure 3: A screenshot of the Scientific Application Environment page

2. Workflows, Software Tools

Workflow and software tools repository provides several modules such as **documents containing best practice procedures** and **workflows for the production of results** relevant to the application categories identified in the SEEM region. Here a user can find links to pages containing lists with short description and access/documentation details for the scientific workflows provided by the VI-SEEM platform, and **list of the codes** that can currently be downloaded from the code repository, arranged by scientific community. Furthermore, it provides guidelines on **how to contribute** to workflows and software tools. The above sections are described in more detail below:

- **How to contribute**: provides detailed information on how the user can contribute his/her own workflows and software tools and make them available to the community.

- o **Scientific workflows**: provides access, documentation and training material for the available workflows, developed through the VI-SEEM integration phases and made available to the community.
- o **Code repository**: provides access to the available software tools developed through the VI-SEEM integration phases and made available to the community.

A screenshot of the workflow and software tools repository is given in Figure 4.

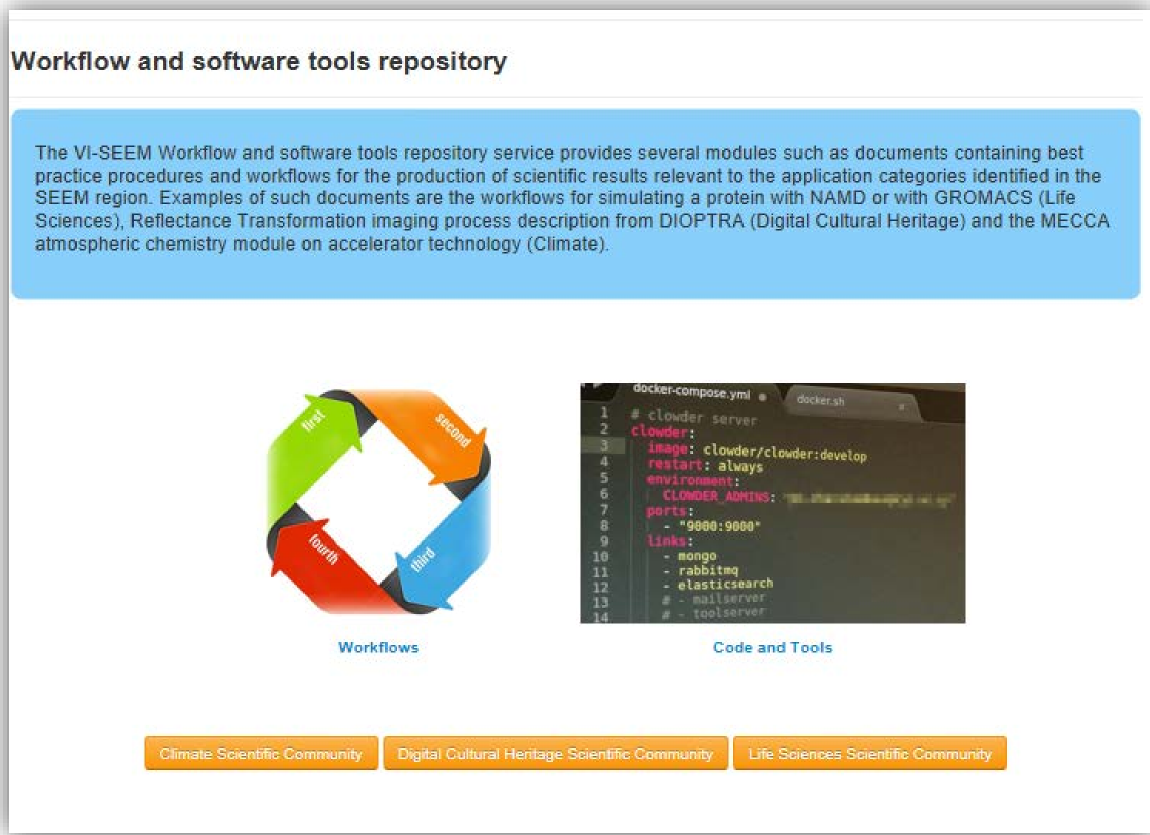


Figure 4: A Screenshot of the Workflow and software repository

3. Regional Community Datasets

Regional community datasets tab provides scientific data, publications and simplified data formats for immediate re-use organized into **climate scientific community datasets**, **digital cultural heritage scientific community datasets**, and **life sciences scientific community datasets** followed with instructions **how to contribute** to datasets.

- o **How to contribute**: provides detailed information on how the user can make his/her datasets available to the community, listing all available steps and information needed in order for the dataset to be fully utilized and lead to new and reproducible research.

- o **Climate datasets:** lists the available datasets for the climate scientific community.
- o **DCH datasets:** lists the available datasets for the digital cultural heritage scientific community.
- o **Life sciences datasets:** lists the available datasets for the life sciences scientific community.

A screenshot of the regional community datasets is provided in Figure 5.

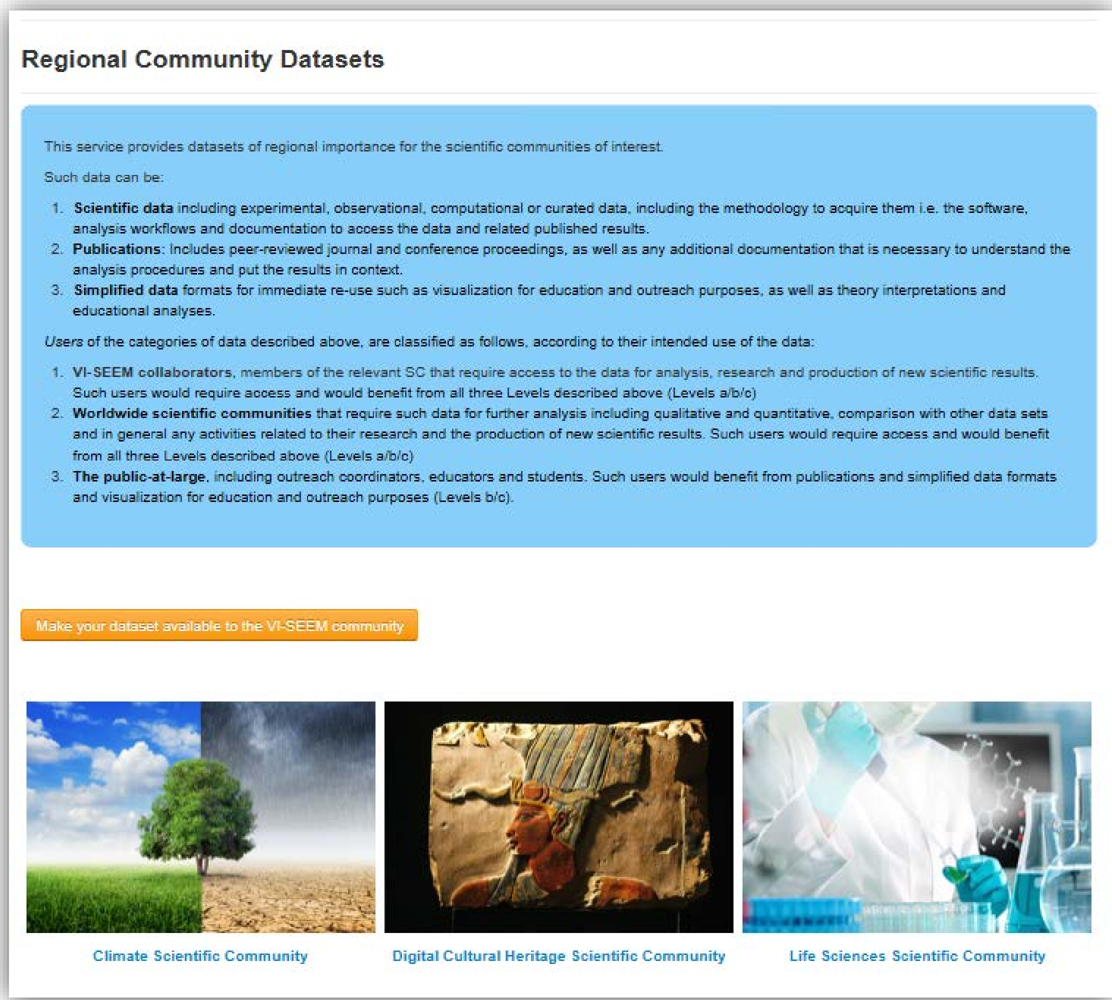


Figure 5: A Screenshot of the Regional Community Datasets.

4. Application-level Services

Application-level services tab provides links to the associated **application-level services of the three scientific communities** as well as instructions on **how to contribute to the application-level services**. All the entries under Application-level Services are described below:

- **How to contribute:** proves detailed information on how the user can make its application-level service available to the community, listing all available steps and information needed.
- **Climate:** lists the application-level services related to the climate scientific community.
- **Digital cultural heritage:** lists the application-level services related to the DCH scientific community.
- **Life sciences:** lists the application-level services related to the life sciences scientific community.

Each section is described in detail in the following sections of this deliverable. By clicking on the link "Application-level Services", the visitor is directed to a webpage the screenshot of which is presented in Figure 6.

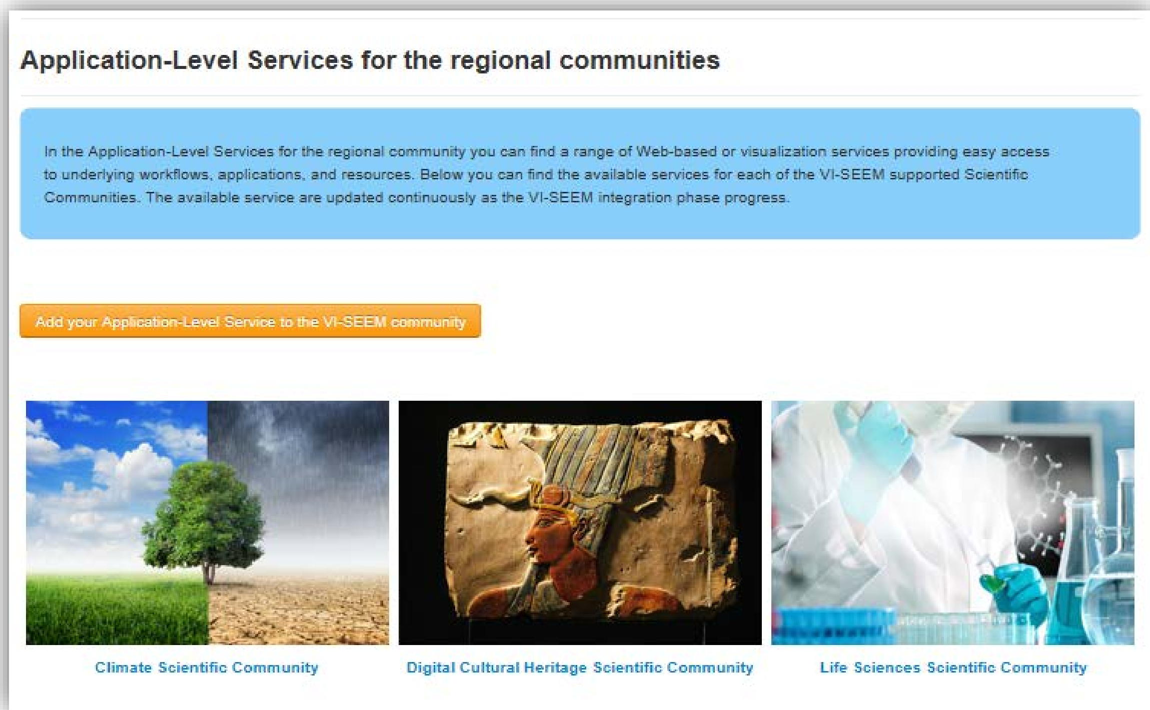


Figure 6: The Application-Level Services for the regional communities' webpage

User Zone

After the header, the "User Zone" follows. User zone is the section of the Virtual Research Environment that is dedicated to different users according to their needs. For instance, a researcher would prioritize different aspects than a student and an SME representative and vice versa. Hence, the user zone has been structured in such a way so that it can be easily used by researchers (either junior or senior), students (who are mostly interested in training material and events), SMEs and in addition provides more general aspects which would be of the interest of anyone of these three categories. A picture of the "User Zone" section is provided in Figure 7.

Figure 7: A screenshot of the User Zone section from the VRE-portal

As explained above the “User Zone” section provides tabs for linking users to the following:

- **Researchers**: by clicking on the text or the icon appearing in the square dedicated to researchers the user is directed to a webpage which is graphically organized in the next seven sections:

  - **Apply for computational resources**
  - **Access the source code repository**
  - **Access to the supported applications & libraries**
  - **Access the scientific workflows**
  - **Access the datasets**
  - **Access the application-level services**
  - **Access the VI-SEEM training portal**

The layout of the webpage dedicated to researchers is presented in Figure 8.

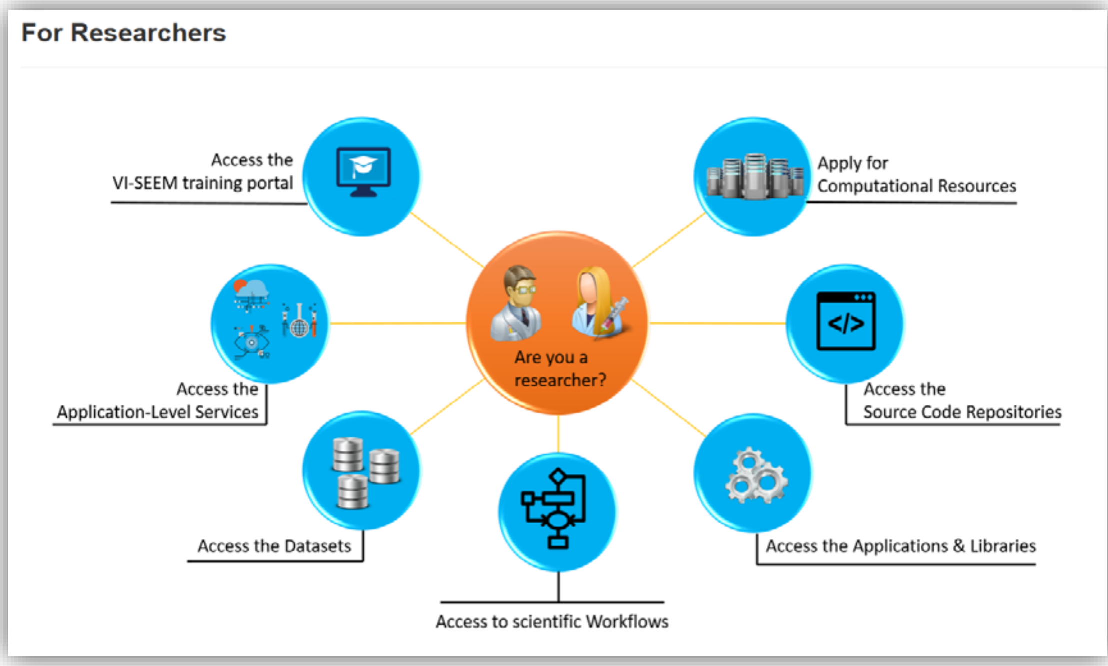


Figure 8: The VRE section dedicated to researchers

- **Students**: by clicking on the text or the icon appearing in the square dedicated to students the user is linked to a webpage which is graphically organized in the next six sections:
  - **Access the VI-SEEM training portal**
  - **Access the events calendar**
  - **Access the source code repository**
  - **Access to the supported applications & libraries**
  - **Access the scientific workflows**
  - **Access the datasets**

The layout of the webpage dedicated to students is presented in Figure 9.

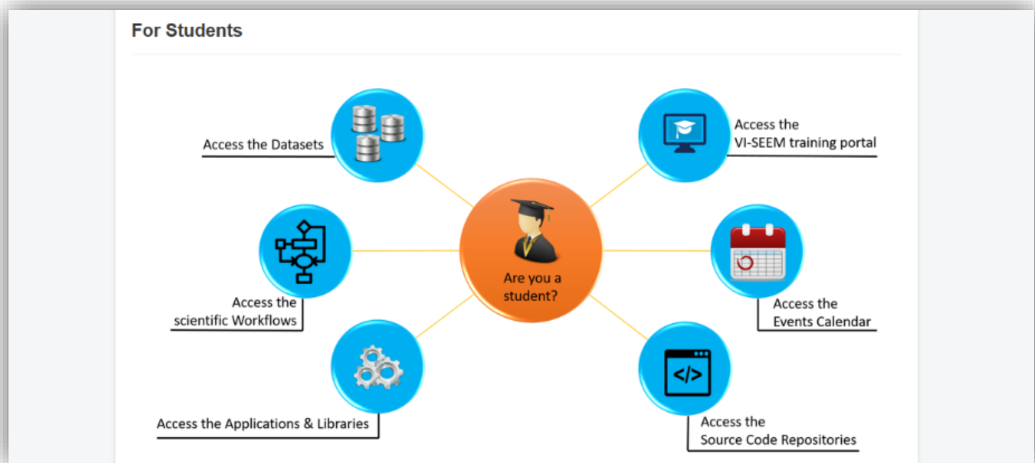


Figure 9: The VRE section dedicated to students

- **SMEs:** by clicking on the text or the icon appearing in the square dedicated to SMEs the user is linked to a webpage which is graphically organized in the next three options:

    - **Apply for computational resources**
    - **Access the datasets**
    - **Access the use cases**

The layout of the webpage dedicated to SMEs is presented in Figure 10.

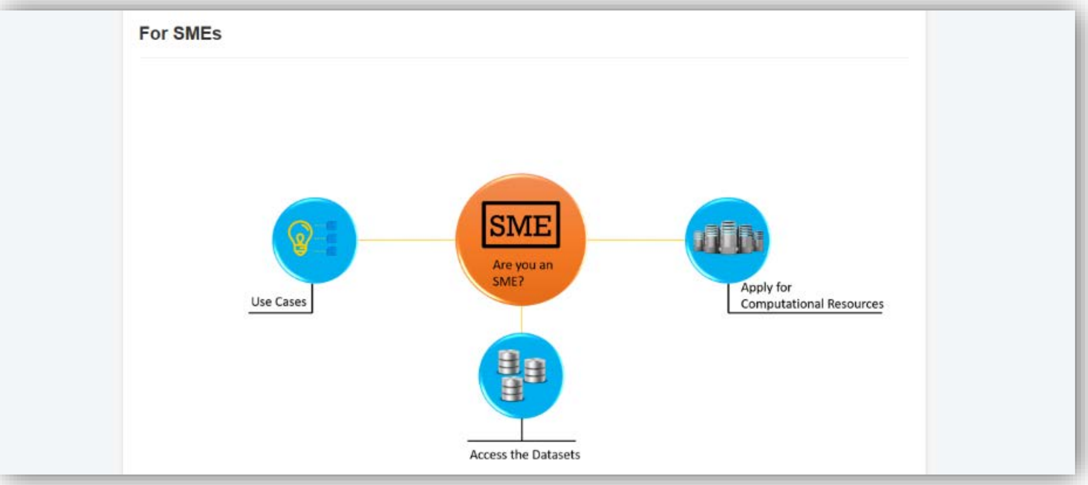


Figure 10: The VRE section dedicated to SMEs

- **Regional resources available:** this section provides access to information related to the available:

    - **VI-SEEM compute resources**
    - **VI-SEEM data storage resources**

- **VI-SEEM service catalogue & portfolio:** This section provides access to the Service Catalogue & Portfolio (https://services.vi-seem.eu/ui/catalogue/services/)

- **Use cases:** This section directs the reader to use cases developed throughout the VI-SEEM integration phases as well as the open calls. The two links direct the visitor to:

    - **Explore VI-SEEM datasets**
    - **Developed use cases**

VI-SEEM support the following Scientific Communities

This section of the VRE-portal is organized according to the scientific discipline of the visitor. Namely, it is configured in three silos one for each community (climate, life sciences, digital cultural heritage) each one containing links on related material on the following:

- **Application-Level services**
- **Datasets**
- **Codes**
- **Workflows**
- **Optimized applications and libraries**

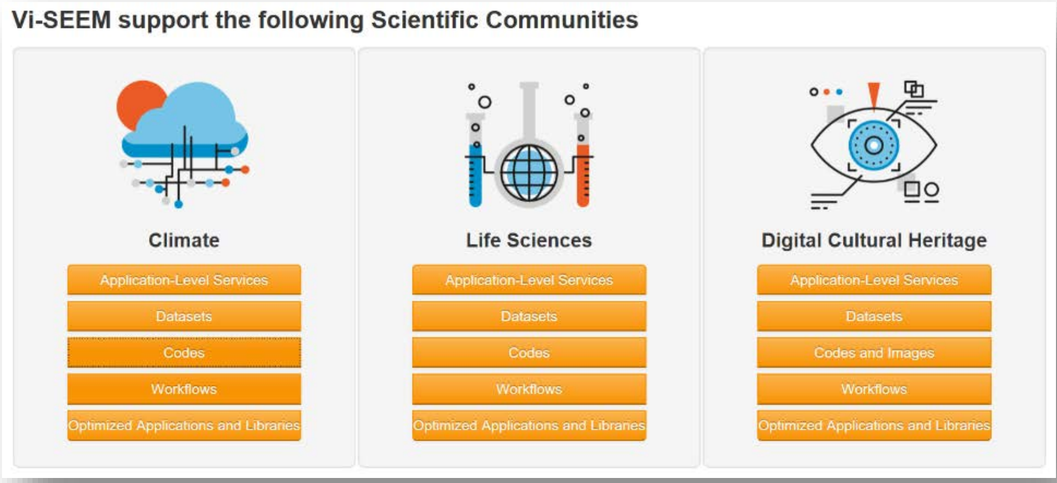


Figure 11: The "VI-SEEM support the following scientific communities" section

Cross Disciplinary Fields

This section reflects to the cross-disciplinary character of VI-SEEM. Thus, instead of the scientific discipline (silo-) orientation of an application or a dataset etc. this section provides a horizontal categorization according to the type of material, which can be used in more than one discipline. The structure of this section as this is illustrated in the next picture (Figure 12):

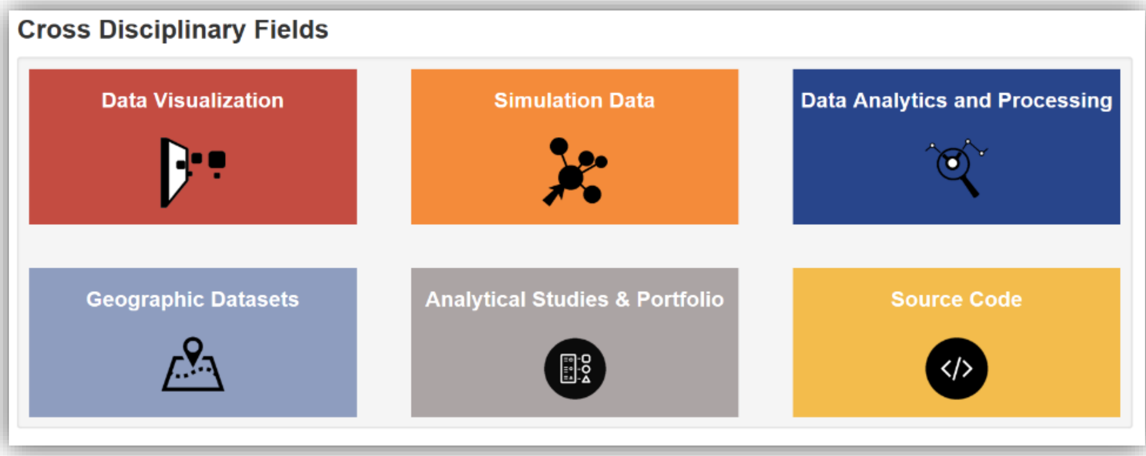


Figure 12: The "Cross Disciplinary Section" of the VRE portal

By clicking on one of the following 5 subsections:

- **Data visualization**
- **Simulation data**
- **Data analytics and processing**
- **Analytical studies & portfolio**
- **Source code**

the user is directed in a webpage which is graphically organized by the following Venn diagram (Figure 13)

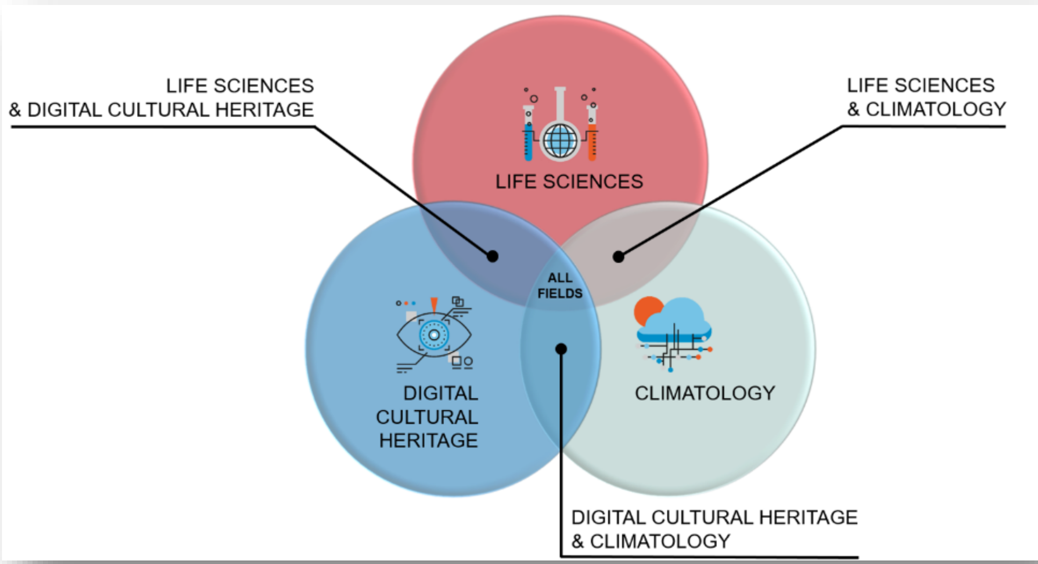


Figure 13: The Venn diagram that directs the visitors of the cross-disciplinary entries of the VRE-portal

By clicking on one of the discipline-overlapping areas the user is directed to the right section of the webpage.

By clicking on the subsection

- **Geographic datasets**

The user is directed a webpage, which is graphically organized by the following, SEEM map (Figure 14):

Figure 14: The interactive map, which directs the user to geographic datasets.

By clicking on a VI-SEEM partner country, the user is directed to a list with projects associated with geographic datasets of the given country. Such projects may involve geo-referencing, remote sensing, etc. So far only one application appears under the country of Montenegro, however more applications are expected to appear as the last two open calls

VI-SEEM Sites

This section provides links to the **VI-SEEM website**, to the VI-SEEM **training Portal**, to the VI-SEEM **code repository** as well as to the VI-SEEM **wiki**. A screenshot of the "VI-SEEM sites" section is given in Figure 15.

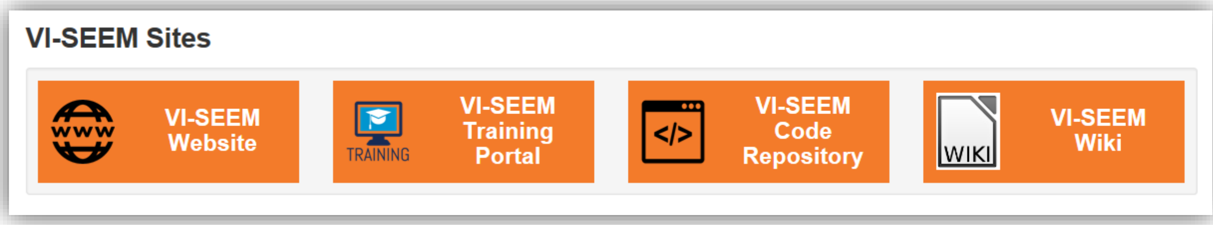


Figure 15: A screenshot of the "VI-SEEM sites" section from the VRE.

A short description of each of the VI-SEEM sites appearing in this section of the VRE is provided below:

VI-SEEM website: VI-SEEM website constitutes a core, dynamic information source, providing input on the project status, and highlights of interest to its target groups. Together with the project social media accounts (Twitter and LinkedIn), the website provides significant visibility to VI-SEEM activities, offerings and added value for the scientific excellence in the region. It is an integral part of the VI-SEEM communication infrastructure and acts as a one-stop-source as it offers links to:

- **the Virtual Research Environment**
- **the service catalogue**
- **the training portal and material**

The website structure is such that any user can easily access content and resources.

The homepage provides a menu of the main categories, links to the VI-SEEM Social media and shortcuts to:

- **the VI-SEEM scientific communities**
- **the Virtual Research Environment**
- **the VI-SEEM training portal**
- **the VI-SEEM success stories**
- **the VRE users' guide**

The homepage also features the latest news and project tweets for maximizing the visibility of the project messages to the users.

The VI-SEEM website can be accessed at https://vi-seem.eu/

VI-SEEM Training Portal: The VI-SEEM training portal can be accessed through the link http://training.vi-seem.eu/. The purpose of this portal is to collect and curate training material for the VI-SEEM services. Through the use of the training portal users have access to information for the available e-infrastructure services (HPC, Grid, Cloud and Data) as well as domain specific material for using the tools and data provided. The VI-SEEM training portal is the main source of the training material for the VI-SEEM users and it offers a number of important features that make the portal user friendly. The training-portal has an enhanced user oriented character, which provides easy access to researchers, students and users in general.

The front page of the training portal consists of a header which provides access to the following training Material:

- **Domain specific software and tools**
- **Data and visualization**
- **Common computing resources**
- **Event related training material**

In addition, the header of the training portal provides a search engine. The front page of the training portal has a high-level organization, structured in three main categories:

- **Domain specific software and tools** providing access to training material for
  - Climate software and tools
  - Digital Cultural Heritage software and tools
  - Life Science software and tools
- **Data and visualization** providing access to training material for:
  - Data
  - Storage services
  - Visualization
- **Common computing resources** providing access to training material for
  - Cloud
  - High Performance Computers
  - Grid

Therefore, the training portal is organized in nine sections each one described by a distinct icon. By clicking on the icon, the user is directed to the associated training material.

VI-SEEM code repository: VI-SEEM code repository hosts tools, software and workflows made available from the VI-SEEM scientific application environment. It is deployed on top of the git open source system. GitLab is used to provide a usable web-based UI for project management and repository control. The service is fully integrated with VI-SEEM Login service. The VI-SEEM code repository can be accessed from https://code.vi-seem.eu/.

VI-SEEM wiki: The VI-SEEM wiki (http://wiki.vi-seem.eu/) aims to support site managers for operating the integrated VRE services on top of the underlying infrastructures. It contains information about the available services, the procedures for integrating existing services with the VI-SEEM e-infrastructure and how-to material for establishing new services and integrating them with the existing VRE services. The VI-SEEM wiki consists of the following contents:

1. **Infrastructure resources**
   - HPC resources
   - Cloud resources
   - Grid resources
   - Storage resources

2. **Access to resources**
    - VI-SEEM login
    - VI-SEEM login integration guide
    - Access to VI-SEEM e-Infrastructure resources

3. **VI-SEEM service catalog/portfolio**
4. **Operations and resource management**
    - GOCDB
    - Monitoring
    - Accounting
    - Helpdesk

5. **Source code repository**
6. **Data services**
    - iRODS Documentation
7. **Service and application enabling**
8. **Training**

Related Infrastructure Projects

This section provides links to other Infrastructure projects, which might be of the interest of the user. This section lists projects such as PRACE, EGI, EUDAT, GÉANT and OpenAIRE as well as projects with a DCH orientation such as ARIADNE, CLARIN, DARIAH, E-RHIS and GRAVITATE. A screenshot of this section of the webpage is provided in Figure 16. A brief description for each of these projects is provided in the following subsections.

Figure 16: A screenshot of the "Related Infrastructure Projects" from the VRE.

2.2.1.1 *PRACE*

The mission of PRACE (Partnership for Advanced Computing in Europe) is to enable high-impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. PRACE seeks to realize this mission by offering world class computing and data management resources and services through a peer review process. PRACE is providing computational resources on the seven most powerful European supercomputing machines.

Motivated by the effort to support the largest number of potential users and to train a new generation of scientists and engineers, PRACE has a wide range of training activities the High Performance Computing (HPC). It also promotes collaboration with industry via its SHAPE SME program. Finally yet importantly, PRACE aims to develop the European supercomputing industry. This include prototyping activities and an initiative dealing with the energy efficiency of computing systems and decreasing their impact on the environment.

2.2.1.2 EGI

The mission of EGI is to create and deliver open solutions for science and research infrastructures by federating digital capabilities, resources and expertise between communities and across national boundaries. It achieves this by providing advanced computing services to support scientists, multinational projects and research infrastructures. These services include compute, storage, data and support.

2.2.1.3 EUDAT

EUDAT's vision is Data that is shared and preserved across borders and disciplines. This is achieved by enabling data stewardship within and between European research communities through a Collaborative Data Infrastructure (CDI), a common model and service infrastructure for managing data spanning all European research data centers and community data repositories.

European researchers and practitioners from any research discipline can preserve, find, access, and process data in a trusted environment, as part of the EUDAT Collaborative Data Infrastructure a network of collaborating, cooperating centers, combining the richness of numerous generic and community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centers.

EUDAT offers heterogeneous research data management services and storage resources, supporting multiple research communities as well as individuals, through a geographically distributed, resilient network distributed across 15 European nations and data is stored alongside some of Europe's most powerful supercomputers.

2.2.1.4 GÉANT

GÉANT is a fundamental element of Europe's e-infrastructure, delivering the pan-European GÉANT network for scientific excellence, research, education and innovation. Through its integrated catalogue of connectivity, collaboration and identity services, GÉANT provides users with highly reliable, unconstrained access to computing, analysis, storage, applications and other resources, to ensure that Europe remains at the forefront of research.

Through interconnections with its 38 national research and education network (NREN) partners, the GÉANT network is the largest and most advanced R&E network in the world, connecting over 50 million users at 10,000 institutions across Europe and supporting all scientific disciplines. The backbone network operates at speeds of up to 500Gbps and reaches over 100 national networks worldwide.

2.2.1.5 OpenAIRE

OpenAIRE aims to promote open scholarship and substantially improve the discoverability and reusability of research publications and data. The initiative is a pan-European collaborative endeavor.

2.2.1.6 ARIADNE

ARIADNE, the Advanced Research Infrastructure for Archaeological Dataset Networking in Europe, brings together and integrates existing archaeological research data infrastructures so that researchers can use the various distributed datasets and new and powerful technologies as an integral component of the archaeological research methodology. There is now a large availability of archaeological digital datasets that, together, span different periods, domains and regions; more are continuously created because of the increasing use of IT. These are the accumulated outcome of the research of individuals, teams and institutions, but form a vast and fragmented corpus and their potential has been constrained by difficult access and non-homogenous perspectives. ARIADNE enables trans-national access of researchers to data centers, tools and guidance, and the creation of new web-based services based on common interfaces to data repositories, availability of reference datasets and usage of innovative technologies. It stimulates new research avenues in the field of archaeology, relying on the comparison, re-use and integration into current research of the outcomes of past and ongoing field and laboratory activity. Such data are scattered amongst diverse collections, datasets, unpublished fieldwork reports (grey literature), and in publications. The latter still being the main source of knowledge sharing. ARIADNE will contribute to the creation of a new community of researchers ready to exploit the contribution of Information Technology and to incorporate it in the body of established archaeological research methodology. To achieve this result ARIADNE is making use of a number of integrating technologies that build on common features of the currently available datasets, and on integrating actions that will build a vibrant community of use.

2.2.1.7 CLARIN

CLARIN (Common Language Resources and Technology Infrastructure) is a research infrastructure that was initiated from the vision that all digital language resources and tools from all over Europe and beyond are accessible through a single sign-on online environment for the support of researchers in the humanities and social sciences.

In 2012, CLARIN ERIC (European Research Infrastructure Consortium) was established and took up the mission to create and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences. Currently CLARIN provides easy and sustainable access to digital language data (in written, spoken, or multimodal form) for scholars in the social sciences and humanities, and beyond. CLARIN also offers advanced tools to discover, explore, exploit, annotate, analyze or combine such data sets, wherever they are located. This is enabled through a networked federation of centers: language data repositories, service centers and knowledge centers, with single sign-on access for all members of the academic community in all participating countries. Tools and data from different centers are interoperable, so that data collections can be combined and tools from different sources can be chained to perform complex operations to support researchers in their work.

The CLARIN infrastructure is fully operational in many countries, and a large number of participating centers are offering access services to data, tools and expertise. At the same time, CLARIN continues to be constructed in some countries that joined more recently, and CLARIN's datasets and services are constantly updated and improved.

2.2.1.8 DARIAH

The Digital Research Infrastructure for the Arts and Humanities (DARIAH) aims to enhance and support digitally enabled research and teaching across the arts and humanities. DARIAH is a network of people, expertise, information, knowledge, content, methods, tools and technologies from its member countries. It develops, maintains and operates an infrastructure in support of ICT-based research practices and sustains researchers in using them to build, analyze and interpret digital resources. By working with communities of practice, DARIAH brings together individual state-of-the-art digital arts and humanities activities and scales their results to a European level. It preserves, provides access to and disseminates research that stems from these collaborations and ensures that best practices, methodological and technical standards are adopted. DARIAH was established as a European Research Infrastructure Consortium (ERIC) in August 2014. Currently, DARIAH has 17 Members and several cooperating partners in 11 non-member countries. DARIAH integrates digital arts and humanities research and activities from across Europe, enabling transnational and transdisciplinary approaches. In particular, it provides value to its members and stakeholders through the validation and sharing of data, services and tools; by providing training and education opportunities; by enabling 'bottom-up' organization around emerging research needs; and through the exercise of foresight and policy engagement. Through these activities, DARIAH promotes the further development of research methods in the arts and humanities, documenting the state-of-the-art, supporting the preservation and curation of research data with a focus on particular challenges including diversity, provenance, multimedia collections and granularity, and acting as a coordinator and integrator for a diverse community of practice.

2.2.1.9 E-RIHS

E-RIHS, the European Research Infrastructure for Heritage Science, supports research on heritage interpretation, preservation, documentation and management. Since heritage is a key component of the European identity, the study and preservation of cultural and natural heritage is a global challenge for science and the European society at large. Within, E-RIHS state-of-the-art tools and services will be provided by cross-disciplinary groups of researchers to cross-disciplinary users and scientific communities working to advance knowledge about heritage and to devise innovative strategies for its preservation. E-RIHS Infrastructure connects researchers in the humanities and natural sciences and fosters a trans-disciplinary culture of exchange and cooperation. E-RIHS pursues the integration of European world-class facilities to create a cohesive entity playing a connecting role in the global community of heritage science. Fragmentation, duplication of efforts, isolation of small research groups put at risk the competitive advantage of European heritage science research, spearheaded so well in the past by its unique cultural heritage. The long-term tradition of this field of research, the ability to combine science with innovation, and the support provided by EU-funded projects and integrating activities such as EU-ARTECH, CHARISMA and IPERION CH in conservation science, and ARIADNE in archaeology, represent the background of E-RIHS. E-RIHS exploits the synergy of the cooperation among the academy, research centers, museums and cultural institutions. Both the scientific and the socio-economic importance connected with heritage science are nowadays evident. The research community has achieved the maturity necessary to make the leap towards a permanent European research infrastructure that will affect broadly on society and economy.

2.2.1.10 GRAVITATE

GRAVITATE (Geometric reconstruction and novel semantic reunification of cultural heritage objects) project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 665155. GRAVITATE started the 1st June 2015, with an original time frame of 36 months that was recently extended to 42 months. The project involves 6 partners: IT Innovation Centre (UK), British Museum (UK), The Cyprus Institute (Cyprus), Consiglio Nazionale delle Ricerche - Institute of Applied Mathematics and Information Technologies (Italy), University of Amsterdam (Netherland), Technion – Israel Institute of Technology (Israel) and University of Haifa (Israel). The overall objectives of the GRAVITATE project are to create a set of software tools that will allow archaeologists, curators, conservators and illustrators to identify and Re-Unify heritage artefacts that have been separated across collections, Re-Associate objects of Material Culture that have some relationship (eg. same school, age, pattern…) and eventually Re-Assemble fragments belonging to the same fragmented artefact. These aims will be achieved through the integration of geometric and semantic analysis and matching. Combining those approaches into a single decision support platform, with a full suite of visualization tools, will provide a unique resource for the cultural heritage research community. The project is driven by the needs of the archaeological community and is exemplified by several pertinent use cases. We anticipate that the insights to be gained from the use of these tools will lead to faster and more accurate reconstruction of heritage artefacts for research, conservation and exhibitions activities, to opportunities for reunification of objects between collections and greater insights into relationships between cultural objects.

Download the Guide to the VRE-Portal

Virtual Research Environment also gives the opportunity to the user to make use of the guide to the Virtual Research Environment Portal. This document provides all the needed information in order to get starting using the portal. The VRE-portal guide can be accessed from https://vre.vi-seem.eu/images/VREmanual/VRE-Users_Guide.pdf. A screenshot of the VRE-portal guide is provided in Figure 17.

Give us your Feedback

Finally, the VRE-Portal provides the option to the user to give feedback about the services. Namely, it requires the user to provide the following information:

- **Reasons for using VRE**
- **Whether the user has ever participated in a VI-SEEM integration phase**
- **Whether the user has ever participated in a VI-SEEM open call**
- **Whether VRE was helpful**
- **Whether the user has encounter any problems while using the VRE**
- **To provide any comments about the VRE**
- **Whether another scientific field should be added in the VRE**
- **What additional information would be nice to add**

Figure 17: The front page of the VRE Users' manual guide

2.3 Technical details

The VRE is hosted by The Cyprus Institute upon a Docker [15] container which runs in a virtual machine. The VRE website is built using the Joomla content management system [8] enhanced with the Bootstrap front-end framework [9]. The service is deployed using an Apache webserver and a MySQL Database. An SSL certificate is also installed to allow for secure connections between guests of the VRE and the web server.

Google analytics [16] have also been enabled to monitor the visits to the VRE, allowing detailed access analytics for the different services provided by the VI-SEEM VRE in order to optimize and improve. Figure 18 illustrates the geographical access patterns and visitor statistics for the VRE platform since its first launch.

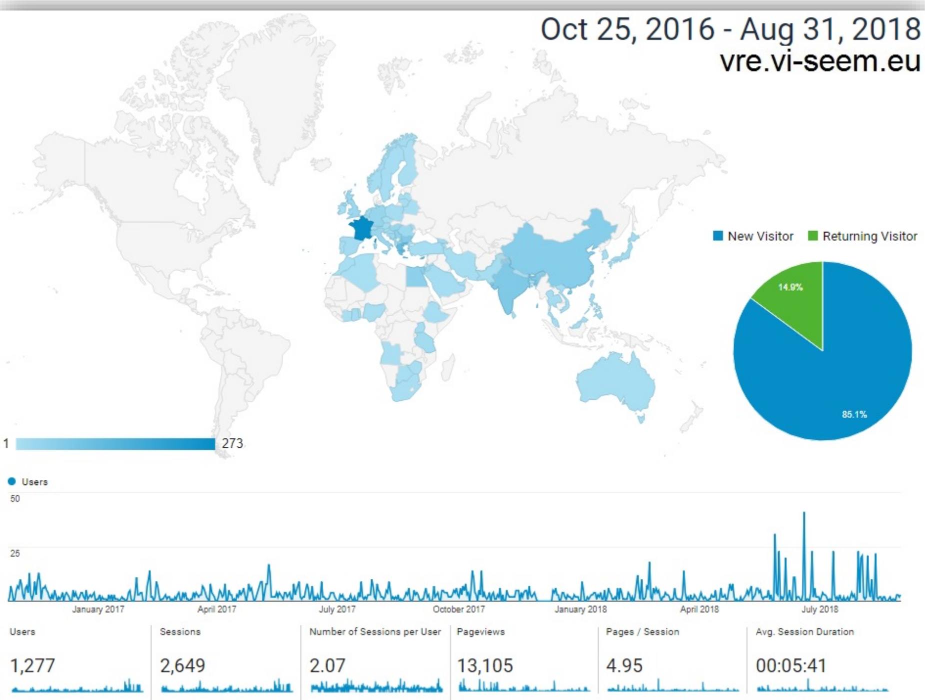


Figure 18: VI-SEEM VRE portal analytics for the period of October 2016 to July 2018

2.4 Users' test of the VRE

After re-shaping VRE according to the reviewers' recommendations, a questionnaire was distributed to randomly chosen users (students, junior researchers, senior researchers) asking for feedback in order to test the overall usability and friendliness of the VRE-portal (as well as of the VI-SEEM webpage). From the 13 feedbacks received, in the question "How easy is for you to find relevant information on the vre.vi-seem.eu" 8 people replied "easy" and 5 people "very easy" while in the question "How useful do you find the vre.vi-seem.eu portal", 5 people replied "useful" and 8 people "very useful" indicating the enhanced usability and user-friendliness of the portal. In addition minor comments were received which are being implemented.

3 VI-SEEM scientific application environment domain specific service

3.1 Scope and service description

The Scientific application environment services provides several modules relevant for the work of the regional scientific communities' interests. Namely, it provides:

- **Optimized applications and libraries**
- **VM images**
- **List of codes**

The applications libraries include short descriptions and access/documentation details for the specific scientific communities' service modules as well as the e-Infrastructure resources supporting them. The VM section provides access and documentation for the available VM images provided by the VI-SEEM scientific application environment and the computational resources supporting them. Finally, the service provides a list of specific developed scientific applications available through the computational facilities of the VI-SEEM scientific community.

Service enablers (one per project partner) are responsible for implementing the smooth integration of services into the VI-SEEM Virtual Research Platform (VRE) and assisting researchers from their country during the process of the service integration. In particular, each service enabler will be responsible for the following:

- Explain the integration procedure to researchers and ensure that the integration follows the agreed timelines as set in the integration plan.
- Assist researcher's access to the VI-SEEM infrastructure and ensure the smooth initiation of service integration.
- Provide technical support and address any problems encountered during the integration possibly with the support of other experts from the project.
- Ensure that researchers receive all required support from the partners that provide the computing resources.

A list with all the services enablers is provided in Table 4.

The scientific application environment domain specific service is available at https://vre.vi-seem.eu/index.php/scientific-application-environment.

3.2 Services integration

The VI-SEEM scientific application environment consists of a collection of scientific service modules provided as a ready-to-use production environment in a number of VI-SEEM partner sites. These scientific modules are provided through the VI-SEEM integration phases described in D5.1 [5] and are requested according to use cases participating in the integration phases. Very briefly, there were three integration phases started on March 2016 and ending on November 2017. Each integration phase lasted for six months and during this period the developers were provided resources to develop their applications. A schematic representation of the integration phases is provided in Figure 19. Due to its capacity the second integration phase was broken down to two stages as the diagram suggests. As a part of WP3, VI-SEEM infrastructure was being continuously updated to make the software modules available in as many infrastructure sites as possible. Tables 1 - 3 list the optimized applications and libraries that are currently available in the VI-SEEM scientific application environment.

Virtual Machine images have also been provided by the scientific communities. Two virtual machine images are available:

1. The Live Access Server VM: a ready to use VM instance of the Live Access Server Climate Scientific Community Application-level service
2. Clowder: pre-configured and ready to use Docker [15] components for installing and using the Clowder digital cultural heritage research data system.

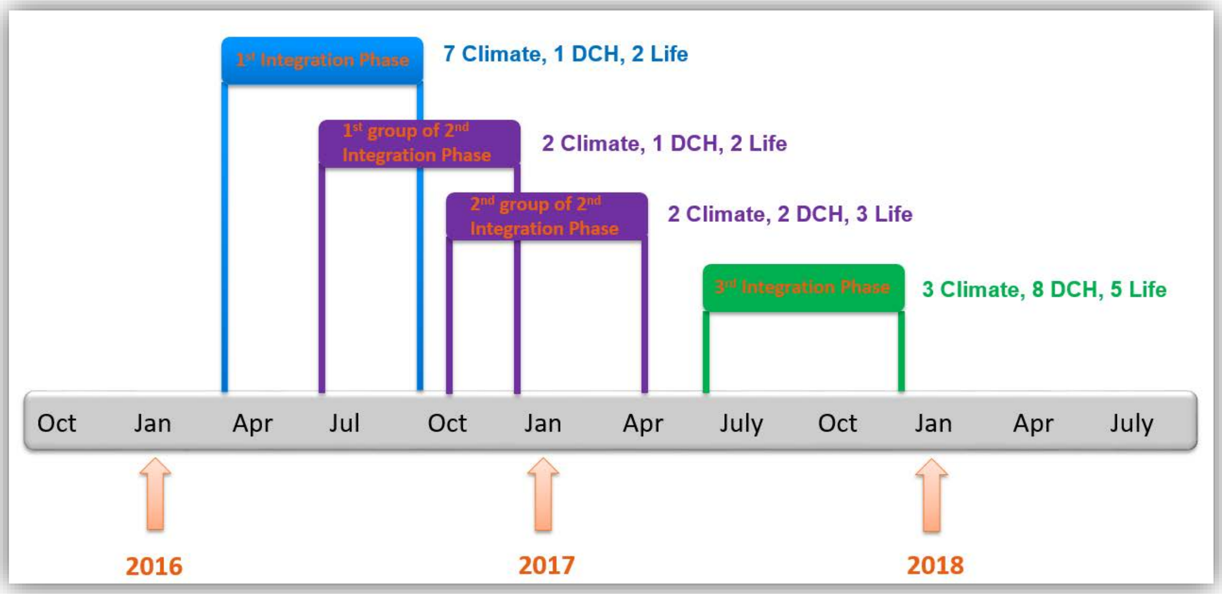


Figure 19: A graphic description of the three integration phases

Climate	CCLM	DREAM	EMAC	FERRET	GIS
	GRADS	IDL	MM5	NCL	OPENFOAM
	Paraview	R	RegCM	WRF	WRF-CHEM

Table 1: Available optimized applications and libraries for the climate community

Digital Cultural Heritage	3DINV	AutoGR	CH-CBIR	CLOWDER	Soft Ontology Layer (SOL)

Table 2: Available optimized applications and libraries for the DCH community

Life Sciences	acml	AMBER	apr	Avogadro	BioCORE
	blacs	blender	boost	CellProfiller	cp2k
	CRYSTAL	cuda	Desmond	espresso	FFTW
	FIREFLY	GAMESS	Gaussian	GROMACS	Hdf5
	ImageJ	ImageMagic	Image-Pro	JVM	LabView
	Lammps	Maestro	maple	Molekel	NAMD
	netcdf	numactl	numpy	OpenCV	Parallel
	paraview	PDB	Phdf5	Pymol	R bbb
	Rasmol	scalapack	scalasca	scipy	siesta
	TCL	teseract	Vaa3D	VMD	WinCoot

Table 3: Available optimized applications and libraries for the life science community

3.3 Access through the VRE platform

Accessing the scientific application environment

The main page of the VI-SEEM scientific application environment, shown in Figure 20, provides to the users general information about the scientific environment. It also provides easy access to categorized software modules and tools hosted. Each section contains information for the corresponding scientific communities; climate, digital cultural heritage and life sciences.

In the optimized applications and libraries category, the service provides a list of available software for the three scientific communities. The climate and weather forecast models are large community codes developed over long time periods by large international user communities. All such codes are presented in Table 1 as well as in the screenshot given in Figure 21. Digital cultural heritage user community employs codes and models, as summarized in Table 2 as well as in the screenshot given in Figure 21. The life sciences research community is providing access to a variety of codes and tools to simulate the processes under investigation; all the associated modules are summarized in Table 3 as well as in the screenshot given in Figure 21. Each module has an additional informal page including general description of the module, linking to documentation and training material and listing the VI-SEEM infrastructures that are currently supporting the module. An example page for the Weather Research and Forecast (WRF) model is illustrated in Figure 22.

Figure 20: VI-SEEM scientific application environment service

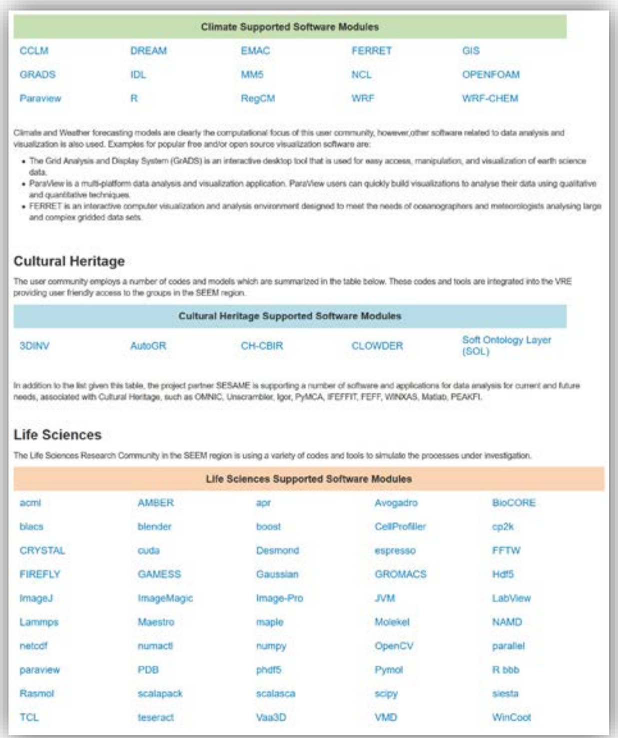


Figure 21: A screenshot with the list of all scientific modules and tools available in the VI-SEEM infrastructure listed by scientific community

VI-SEEM Virtual Research Environment Portal

Vi-SEEM

Home | Scientific Application Environment | Workflow, Software Tools | Regional Community Datasets | Application-Level Services

WRF

The **Weather Research and Forecasting** (**WRF**) Model is a numerical weather prediction (NWP) system designed to serve both atmospheric research and operational forecasting needs. NWP refers to the simulation and prediction of the atmosphere with a computer model, and WRF is a set of software for this. WRF features two dynamical (computational) cores (or *solvers*), a data assimilation system, and a software architecture allowing for parallel computation and system extensibility. The model serves a wide range of meteorological applications across scales ranging from meters to thousands of kilometers.

The effort to develop WRF began in the latter part of the 1990s and was a collaborative partnership principally among the US National Center for Atmospheric Research (NCAR), the US National Oceanic and Atmospheric Administration (represented by the National Centers for Environmental Prediction (NCEP) and the (then) Forecast Systems Laboratory (FSL)), the Air Force Weather Agency (AFWA), the Naval Research Laboratory (NRL), the University of Oklahoma (OU), and the Federal Aviation Administration (FAA). The bulk of the work on the model has been performed or supported by NCAR, NOAA, and AFWA.

WRF allows researchers to produce simulations reflecting either real data (observations, analyses) or idealized atmospheric conditions. WRF provides operational forecasting a flexible and robust platform, while offering advances in physics, numerics, and data assimilation contributed by the many research community developers. WRF is currently in operational use at NCEP and other forecasting centers internationally. WRF has grown to have a large worldwide community of users (over 23,000 registered users in over 150 countries), and workshops and tutorials are held each year at NCAR. WRF is used extensively for research and real-time forecasting throughout the world.

WRF offers two dynamical solvers for its computation of the atmospheric governing equations, and the variants of the model are known as WRF-ARW (Advanced Research WRF) and WRF-NMM (Nonhydrostatic Mesoscale Model). The Advanced Research WRF (ARW) is supported to the community by the NCAR Mesoscale and Microscale Meteorology Division. The WRF-NMM solver variant was based on the Eta Model, and later Nonhydrostatic Mesoscale Model, developed at NCEP. The WRF-NMM (NMM) is supported to the community by the Developmental Testbed Center (DTC).

HPC Resources supporting module				
ARIS	Armcluster	BA-HPC	Cy-Tera	ICAM BlueGene/P
InfraGRID	MK-03-FINKI	PARADOX		

Figure 22: Example of software module specific information for the WRF model

Regarding virtual machines (VMs), the VRE platform provides description and access details for the available VMs for the scientific communities. The VM information includes the operating system version, software packages installed, running services, and access instructions. Figure 23 provides an example of how the VRE platform presents and provide access to the available virtual machines. The example illustrates how the user can get information and the required Docker components for setting up and running a version of the Clowder research data system.

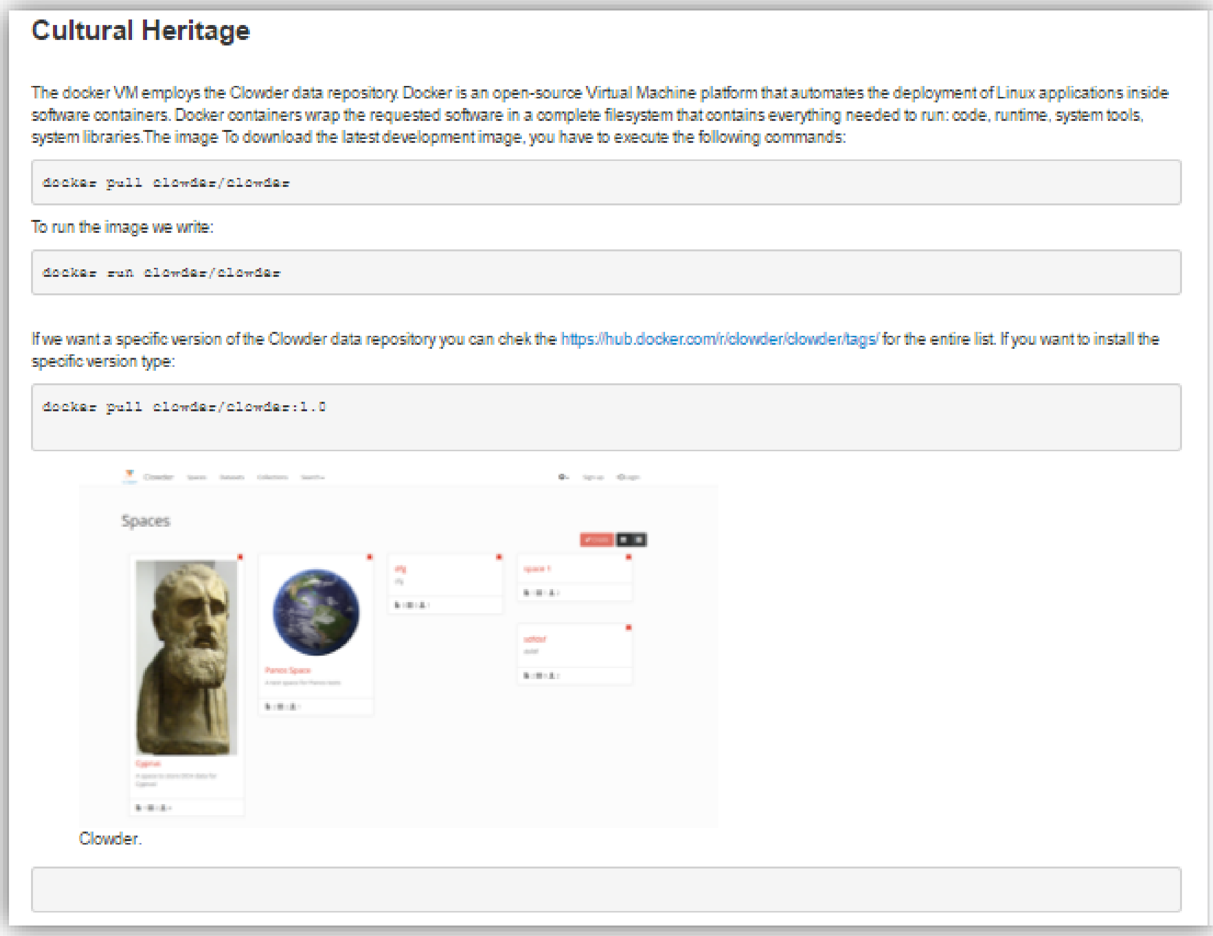


Figure 23: An example of a Digital Cultural Heritage available Virtual Machine Image through the VI-SEEM scientific application environment.

Finally, the VI-SEEM scientific application environment provides a list of use cases developed through the VI-SEEM integration phases by scientific partners. These use cases include development or usage of software modules among scientific partners. In each phase of the project, scientific partners verified the integration, applicability and readiness for use of a number of optimized applications. A table listing the developed cases, the integration phase, completion date, country, institute, and the scientific contact is provided for each of the three scientific communities. An example for the case of the Climate scientific community is illustrated in Figure 24. Interested users can follow the link to the use case of interest for more information, which includes a small description regarding the scope of the module, developers, and contact people (e.g. in Figure 25). Thus, if any user interested in using of any of these developed cases, he/she can find corresponding contacts via the VRE portal.

Climate Scientific Community

	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, 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, and Geography, Bulgarian Academy of Science	Kostadin Ganev
	OPENFOAM	3rd	M28	Albania	UPT	Polytechnic University of Tirana	Neki Frasheri
	Continuous_LST	3rd	M28	Israel	IUCC	Bar-Ilan University, The Department of Geography and Environment	Itamar Lensky

Figure 24: List of the climate specific use cases developed over the VI-SEEM scientific application environment

Figure 25: A detailed view of the information provided for the climate use cases developed over the VI-SEEM scientific application

3.3.2 Contributing to the scientific application environment

To streamline the procedure of contributing to the VI-SEEM scientific environment, each project partner assigned a service enabler who is responsible to coordinate and assist researchers from their own country during the process of the service integration. A list of all the current service enablers is provided in Table 4. In particular, each service enabler is responsible to:

- Explain the integration procedure to researchers and ensure that the integration follows the agreed timelines as set in the integration plan.
- Assist researcher's access to the VI-SEEM infrastructure and ensure the smooth initiation of service integration.
- Provide technical support and address any problems encountered during the integration possibly with the support of other experts from the project.
- Ensure that researchers receive all required support from the partners that provide the computing resources.

The researcher(s) should provide a description of each service module and its use cases on the VRE portal, with links to accompanying documentation, training material, and contact details of the responsible scientist.

Technical documentation is available in the VI-SEEM wiki [19], and entails instructions on how to access, compile/deploy and/or use each service module. Furthermore, the training material covers scientific aspects of the application, and must is hosted on the VI-SEEM training portal. Finally, the researcher(s) should provide the contact details of the person(s) responsible for providing scientific support to future users of the VRE. The responsible scientist for each service module will then need to make sure that each service module, as defined in the VRE integration plan, is set up correctly, and that it works and is accessible to users as expected.

The integration process and specific checkpoints to follow when contributing to the scientific environment are available at the VRE platform at https://vre.vi-seem.eu/index.php/scientific-application-environment/how-to-contribute and illustrated in Figure 26.

Figure 26: Guidelines for contributing to VI-SEEM scientific application environment

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

Table 4: The list of service enablers

4 VI-SEEM workflows and software tools domain specific service

4.1 Scope and service description

This service provides access and information to 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. The service describes and documents scientific workflows provided by the VI-SEEM platform. Similar details are also provided for the list of codes that are available and can be downloaded from the VI-SEEM code repository. Details of how researchers can contribute code are given. Details for contributing workflows – which can take the form of scripts, which automate data generation and processing, interfaces that ease the use of tools and software applications and documents that describe scientific processes, are also provided.

Scientific workflows are made available either as ready-to-use tools that automate the process or as documents describing a specific process. The former category is made available through the VI-SEEM code repository [10], while the latter is uploaded in the VI-SEEM data repository [17]. Codes produce by the VI-SEEM scientific communities are made available through the VI-SEEM code repository. In addition, some VI-SEEM codes created in the framework of Digital Cultural Heritage can be found in the VI-SEEM Clowder. Both the VI-SEEM code repository and the VI-SEEM data repository are integrated with the VI-SEEM Login service, the AAI infrastructure of the VRE that provides the authorization and authentication framework for all users of the region. This service is available at https://vre.vi-seem.eu/index.php/workflow-pipeline-and-software-tools-repository.

4.2 Services integration

As already explained in the previous sections, workflows and software have been developed and subsequently integrated throughout the three integration phases as these are visually explained in Figure 19. As the three open calls for production progress more workflows and codes are expected to be made available through the Virtual Research Environment platform. Table 2 lists the currently available workflows and codes providing the specific links to the VRE platform. Workflows and codes are listed along the specific use case developed through the integration phase as well as the project participating in the open calls for production.

	Application acronym	Codes	Workflows
Life Sciences	MD-Sim	ChemBioServer, Subtract, AFMM, NanoCrystal are stand-alone application level services and their codes are running on VMs.	Workflows for simulating a protein with NAMD and GROMACS
	PSOMI		PSOMI Workflow (NAMD tutorial)
	CCC	CCC code	Usage of the CCC code
	DICOM	DICOM Network is a stand-alone application level service running on a VM	DICOM Workflow
	THERMOGENOME	THERMOGENOME Perl script	THERMOGENOME Workflow
	CNCADD	CNCADD Fortran Code	CNCADD Workflow
	SQP-IRS	SQP-IRS code (PSIPRED) is available in the DSpace at link	
	BioMoFS	BioMoFS codes can be accessed in the iRODS repository: link	BioMoFS workflows can be accessed in the iRODS repository: link
	MDSMS		MDSMS workflows can be accessed in the iRODS repository: link
	NGS1		Best Practices for NGS1
	NGS2		Best Practices for NGS2
	MolSurf	Accompanying scripts and input files to be used in GROMACS: prod.mdp, queue_script.pbs, heat.mdp, min.mdp, relax.mdp, surfactants_water.top	MolSurf Workflow
	MULTIDRUG	Accompanying scripts and input files to be used in GROMACS: heat.mdp, membrane.top, min_env.mdp, min_whole.mdp, prod.mdp, relax.mdp, run_mdrun.pbs	MULTIDRUG Workflow MULTIDRUG Workflow Chart

Table 5: Life Sciences' workflows and codes available through the VRE platform

	Application acronym	Codes	Workflows
Climate	EMAC	Medina: A MECCA - KPP Fortran to CUDA source-to-source pre-processor for CUDA enabled accelerators	MECCA – KPP Installation Guide
	WRF-ARW		WRF Installation Workflow WRF Check Output and Troubleshooting workflow
	VINE	Script for visualization of WRF-Chem output	VINE: Setup and run WRF-Chem model over the south Caucasus domain
	TVRegCM		Tuning and Validation of the RegCM Workflow
	ACIQ		ACIQ Workflow
	ClimStudyArmenia		ClimStudyArmenia Workflows: Installation of WRF Usage of WRF Usage of SLURM
	DRS-ACS	X-H-stretching-potentials 2DSchrodingerEq bound-states-of-the-Morse-well	Workflow for the calculation of Anharmonic X-H Stretching Frequencies of Intramolecular Vibrational Chromophores Relevant to Atmospheric Chemistry

Table 6: Climate's workflows and codes available through the VRE platform

	Application acronym	Codes	Workflows
DCH	Clowder	Complete installation of the Clowder CMS platform as modified by the Cyprus Institute. Also contains new and/or updated extractors. Also contains dockerfiles for all extractors and a convenient Docker Compose file to download and start the server with all the extractors integrated.	Clowder Tutorial with workflows
	Manuscript		Manuscript workflow Article explaining how Manuscript works
	3DINV	3DINV code runs in the Clowder platform	3DINV workflow 3DINV Guide manual including workflows
	AutoGR	AutoGR code runs in the Clowder platform	AutoGR workflow AutoGR Guide manual including workflows
	CH-CBIR	CH-CBIR code: Remote sensing image classification and retrieval using features obtained from convolutional neural networks (CNN).	
	VirMuf	VirMuf code runs in the Clowder platform	VirMuf Pipeline, tutorial and 3d model examples.
	Dioptra		The metadata schema developed by the Science and Technology for Archaeology Research Center (STARC) for the semantic structuring of DCH assets and metadata.
	CHERE		CHERE manual with a workflow

Table 7: DCH's workflows and codes available through the VRE platform

4.3 Access through the VRE platform

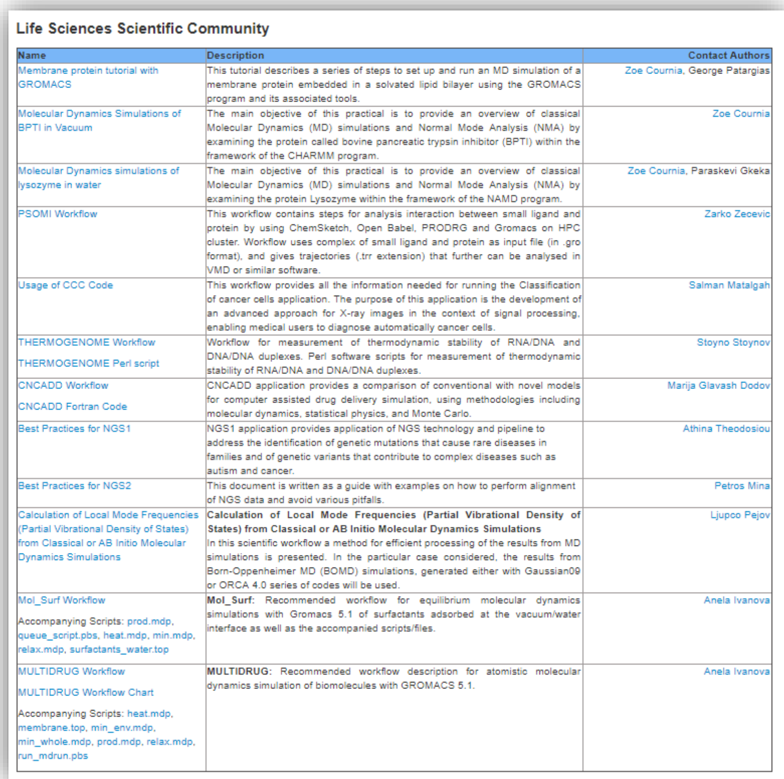
Accessing workflows and software tools

The main page of the “workflows and software tools service” is illustrated in Figure 27. The page provides some general information about the service along with for accessing available scientific workflows (left) and software tools (right). There is also a link for accessing available workflows and codes for the three scientific communities.

The “scientific workflows” page provides a list with the available scientific workflows for each scientific community. Figure 28 shows an example of how the scientific workflows are listed for the life sciences community. The table provides a link to the specific workflows along with a short description and contact information for the responsible scientists for each application.

Similarly, the “code repository” page (Figure 29) provides a list with the available codes for each scientific community along with generic. These codes are software modules built by the VI-SEEM community during the integration of the generic services, and are openly available for use.

Figure 27: Workflow and software tools repository main page



Life Sciences Scientific Community

Name	Description	Contact Authors
Membrane protein tutorial with GROMACS	This tutorial describes a series of steps to set up and run an MD simulation of a membrane protein embedded in a solvated lipid bilayer using the GROMACS program and its associated tools.	Zoe Cournia, George Patargias
Molecular Dynamics Simulations of BPTI in Vacuum	The main objective of this practical is to provide an overview of classical Molecular Dynamics (MD) simulations and Normal Mode Analysis (NMA) by examining the protein called bovine pancreatic trypsin inhibitor (BPTI) within the framework of the CHARMM program.	Zoe Cournia
Molecular Dynamics simulations of lysozyme in water	The main objective of this practical is to provide an overview of classical Molecular Dynamics (MD) simulations and Normal Mode Analysis (NMA) by examining the protein Lysozyme within the framework of the NAMD program.	Zoe Cournia, Paraskevi Gkeka
PSOMI Workflow	This workflow contains steps for analysis interaction between small ligand and protein by using ChemSketch, Open Babel, PRODRG and Gromacs on HPC cluster. Workflow uses complex of small ligand and protein as input file (in .gro format), and gives trajectories (.trr extension) that further can be analysed in VMD or similar software.	Zarko Zecevic
Usage of CCC Code	This workflow provides all the information needed for running the Classification of cancer cells application. The purpose of this application is the development of an advanced approach for X-ray images in the context of signal processing, enabling medical users to diagnose automatically cancer cells.	Salman Matalgah
THERMOGENOME Workflow THERMOGENOME Perl script	Workflow for measurement of thermodynamic stability of RNA/DNA and DNA/DNA duplexes. Perl software scripts for measurement of thermodynamic stability of RNA/DNA and DNA/DNA duplexes.	Stoyno Stoynov
CNCADD Workflow CNCADD Fortran Code	CNCADD application provides a comparison of conventional with novel models for computer assisted drug delivery simulation, using methodologies including molecular dynamics, statistical physics, and Monte Carlo.	Marija Glavash Dodov
Best Practices for NGS1	NGS1 application provides application of 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.	Athina Theodosiou
Best Practices for NGS2	This document is written as a guide with examples on how to perform alignment of NGS data and avoid various pitfalls.	Petros Mina
Calculation of Local Mode Frequencies (Partial Vibrational Density of States) from Classical or AB Initio Molecular Dynamics Simulations	**Calculation of Local Mode Frequencies (Partial Vibrational Density of States) from Classical or AB Initio Molecular Dynamics Simulations** In this scientific workflow a method for efficient processing of the results from MD simulations is presented. In the particular case considered, the results from Born-Oppenheimer MD (BOMD) simulations, generated either with Gaussian09 or ORCA 4.0 series of codes will be used.	Ljupco Pejov
Mol_Surf Workflow Accompanying Scripts: prod.mdp, queue_script.pbs, heat.mdp, min.mdp, relax.mdp, surfactants_water.top	**Mol_Surf**: Recommended workflow for equilibrium molecular dynamics simulations with Gromacs 5.1 of surfactants adsorbed at the vacuum/water interface as well as the accompanied scripts/files.	Anela Ivanova
MULTIDRUG Workflow MULTIDRUG Workflow Chart Accompanying Scripts: heat.mdp, membrane.top, min_env.mdp, min_whole.mdp, prod.mdp, relax.mdp, run_mdrun.pbs	**MULTIDRUG**: Recommended workflow description for atomistic molecular dynamics simulation of biomolecules with GROMACS 5.1.	Anela Ivanova

Figure 28: List of available scientific workflows for the life sciences community as described in the VRE platform

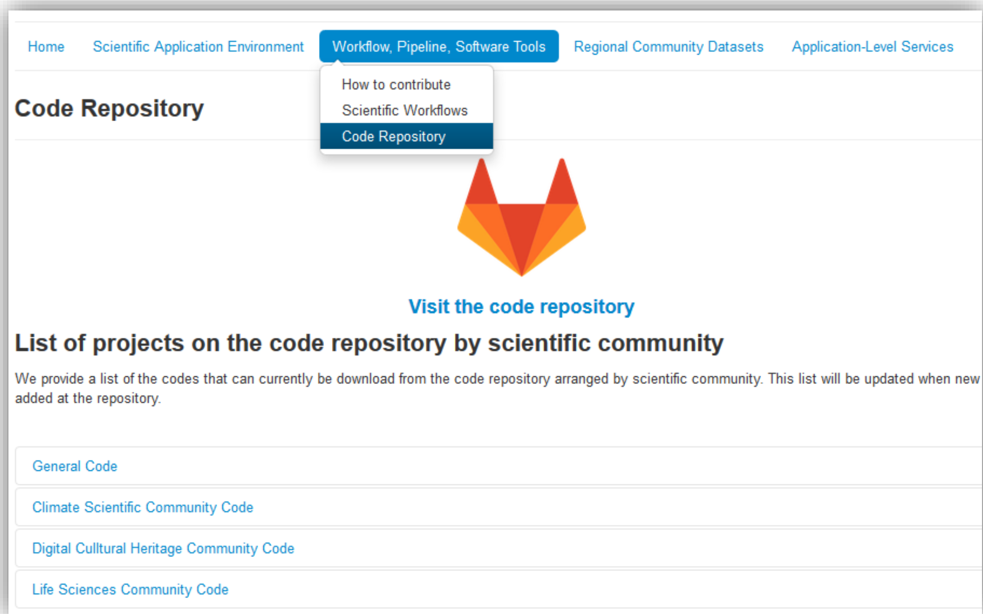


Figure 29: VRE portal code repository page

Making workflows and software tools available to the VRE platform

A specific "how to contribute" page is available to the interested researchers to ease the process of making workflows and codes available to the VI-SEEM communities. The page lists in an easy to follow manner the needed steps and information researchers need to follow and provide in order for the workflows/codes to be made available (Figure 30). Workflow and software tools should be added to the VRE portal accompanied with the following information:

1. Documentation: in case of code, README.md files must be provided with instructions on how to access, compile and/or use them.
2. Training material: the training material will cover scientific aspects of the application, and must be hosted on the VI-SEEM training portal.
3. Contact details of the responsible scientist, who will be providing scientific support to future users of the VRE. The responsible scientist for each contribution will then need to make sure that each contribution, as defined in the VRE integration plan, is set up correctly, and that it works and is accessible to users as expected. Finally, this scientist is responsible for contacting the platform support (support@vi-seem.eu) so that the VRE platform is updated to reflect the contribution.
4. License: Licensing information on the use of the material uploaded in the core repository must be also provided.

The complete guidelines are available at https://vre.vi-seem.eu/index.php/workflow-pipeline-and-software-tools-repository/how-to-contribute

How to Contribute Workflows and Code

Workflow and code repositories usually require small amounts of storage space and it is envisaged that after quality control, researchers should be provided with the appropriate storage to store their datasets for them to be accessible and freely available for other researchers to access without any form of restrictions. Workflow, pipeline and software repository tools should be added to the VRE portal through the following guidelines.

1. **Accompanying documentation:** In case of code, README.md files must be provided with instructions on how to access, compile and/or use it.
2. **Training material**: The training material will cover scientific aspects of the application, and must be hosted on the VI-SEEM training portal.
3. **Contact details of the responsible scientist**, who will be providing scientific support to future users of the VRE. The responsible scientist for each contribution will then need to make sure that each contribution, as defined in the VRE integration plan, is set up correctly, and that it works and is accessible to users as expected. Finally, this scientist is responsible for contacting the platform support (support@vi-seem.eu) so that the VRE platform is updated.

Existing workflows and code can be found at the following pages:

1. Scientific Workflows
2. Code Repository

Contributing Workflows

Scientific workflows/pipelines can come in various forms:

1. scripts that automate data generation and processing,
2. interfaces that ease the use of tools and software applications and
3. documents that describe scientific processes and/or training material in various formats (e.g., pdf, doc, ppt).

Depending on the workflow type (code or documents), files are submitted as follows:

1. Scripts/Code: on the code repository (see also Contributing Code below)
2. Documents: on the VI-SEEM repository following the guidelines for uploading datasets

Contributing Code

These steps should be followed when uploading code to the VRE portal:

1. Upload your software to the VI-SEEM code repository
    1. Create a repository for each project
    2. Create a README.md file that describes the project and how to compile and run
2. Upload training material regarding the code in the "Training portal"
3. Contact support@vi-seem.eu to make sure the application is listed in the "List of codes" under the relevant scientific community providing
    1. A description of the application
    2. Responsible scientist details
    3. Links to documentation material
    4. Links to training material

Figure 30: Step-by-step guidelines for making workflows and software codes available to the VRE platform

5 VI-SEEM regional community datasets domain specific service

5.1 Scope and service description

This service provides access and information regarding datasets of regional importance for the scientific communities of interest. The service integrates the generic data services provided by WP4 and links them to the scientific communities [4].

Datasets can be made available to the scientific communities through three different methods:

1. Uploaded to the VI-SEEM dataset repository service [19]
2. Uploaded to an application-level service as those described in chapter 6
3. Linked through an external source

The first two methods provide to the data owned the storage infrastructure and services for making the dataset available. The latter case covers data that are already hosted in an external infrastructure. In all cases, the datasets should follow the VI-SEEM data management plan as described in D5.2 [6].

5.2 Services integration

As already explained in the previous sections the Virtual Research Environment has been enriched with datasets created through the development of the associated applications, which took part in the three distinct integration phases as these are explained graphically in Figure 19.

In Table 8, Table 9 and Table 10 we are listing the currently available datasets as well as providing the specific links to the datasets for the scientific disciplines of climate, digital cultural heritage and life sciences respectively. Datasets are listed along the specific use case developed through the integration phases. In the VI-SEEM repository the user can view all datasets available, organized by the different scientific communities that the datasets belong to, as shown in Figure 31. Selecting the specific scientific community, the user can then get access to the available datasets together with all relevant metadata that accompany the dataset. The minimum requirements datasets need to obey are described in detail in the VI-SEEM data management plan D5.2 [6].

	Application Acronym	Datasets
Climate	WRF-ARW	Current atmospheric and weather predictions available from the Cyprus Department of Meteorology: • Current Weather Predictions DSM Values • Current Weather Predictions Point Values • Current Atmospheric Predictions (RAOB)
	VINE	GRENA and National Environmental Agency of Georgia • Simulation of dust distribution over Caucasus region by WRF-Chem • Regional Climate Model simulation for Caucasus Region
	WRF–Chem	WRF-Chem Regional Climate Modelling • WRF–Chem: ECMWF-ERAINT dataset
	DREAMCLIMATE	Simulation of aerosol optical thickness and surface dust concentration for North Africa, Southern Europe and Middle East by DREAM model • Datasets
	ACIQLife	Annually averaged output fields of the Air Quality Indices for the territory of of the country of Bulgaria and the city of Sofia: • ACIQLife Dataset for the city of Sofia • ACIQLife Dataset for the county of Bulgaria
	TVRegCM	Tuning and Validation of RegCM: • TVRegCM_r11255 Results • TVRegCM_r11221 Results
	ClimStudyArmenia	Evaluation of WRF Model Output for Severe Weather Forecasting over the territory of Armenia for June 2016. • ClimStudyArmenia datasets
	RCM Mena-CORDEX	Regional Climate Modelling (Middle East, North Africa) – WRF • RCM Mena-CORDEX datasets
	DRS-ACS	The dataset contains the complete molecular dynamics (MD) trajectories of free irinotecan molecule at several temperatures. MD study of irinotecan molecule was carried out with the atom-centered density matrix propagation scheme at AM1 semiempirical level of theory, at series of different temperatures, ranging from 5 K to 300 K. • Access the dataset
	Continuous_LST	Continuous Land Surface Temperature (LST) database at satellite (MODIS) resolution of 1 km using combination of Satellite and Model data (Numerical Weather Prediction data, CFSV2). • Access the datasets

Table 8: Regional community datasets on climate available through the VRE platform

	Application Acronym	Datasets
DCH	BVL	A set of rare books that represent documentation sources for the culture and civilization on the Banat region (Banatica collection). A subset of 200 digitized books have been made available in UVT's GridFTP. • Banatica Dataset
	ELKA	The Electronic Corpus of Karamanlidika (ELKA) offers access to a number of manually digitized Karamandlidika texts available for testing and improving OCR methods. • Karamaldidika texts
	Manuscript	The Historical Arabic Documents Dataset for Recognition Systems: Annotation on sub-word level of five books written by different writers from the years 1088-1451. • Book 1 • Book 2 • Book 3 • Book 4 • Book 5
	3DINV	Three-dimensional (3-D) inversion of surface Electrical Resistivity Tomography (ERT) data in order to automatically determine a 3-D resistivity subsurface model. Real ERT data from survey at • the Sikyon archaeological area located in Peloponnesus regional unit, Greece. • the Istro archaeological area located in Crete regional unit, Greece. • the Ierapetra area located in Crete regional unit, Greece. • the Evropos area located in Kilkis regional unit, Greece.
	AUTOGR	Automatic image geo-referencing script Datasets for testing
	CH-CBIR	CNN Features for Remote Sensing Image Classification – Testing datasets • AID dataset • Deepsat dataset
	CSAD	• RTi datasets • Datasets of Inscriptions Ptolemaic and Roman Egypt
	PETRA	• Painted architectural surfaces including wall paintings and attached decorative elements from Nabataean Petra • Painted marble sculpture from Nabataean Petra
	VirMuf	VirMuF is an application that aims to Popularize the museum experience by adding many ways to simplify and present archaeological and historical data. • Access the datasets
	Dioptra	The metadata schema developed by the Science and Technology for Archaeology Research Center (STARC) has the goal to enable data interoperability and access to the digital resources stored in the local repository. • Access the dataset

Table 9: Regional community datasets on DCH available through the VRE platform

	Application Acronym	Datasets
Life Sciences	MD-Sim	MD trajectories of oncogenic proteins with mutations relevant to the SEEM area: • MD simulation Trajectory for the wild-type PI3Ka protein solvated in water (5 replicas of 100 ns each and 2 replicas of 800 ns each) • MD simulation Trajectory for the oncogenic PI3Ka mutant H1047R protein solvated in water (5 replicas of 100 ns each) • MD simulation Trajectory for the oncogenic PI3Ka mutant E545K protein solvated in water (4 replicas of 800 ns each) • MD simulation Trajectory for the wild-type RXRa protein (1 simulation of 800 ns each) • MD simulation Trajectory for the oncogenic RXRa mutant S427F RXRa (2 replicas of 800 ns each) • MD simulation Trajectory of an Arachidonic acid membrane in cis and trans conformations
	PSOMI	Trajectories for PSOMI application: Link to datasets
	CCC	Datasets used as examples for running CCC code. Some of these datasets have been taken from DICOM Network Link to datasets
	THERMOGENOME	Datasets on the thermodynamic pattern of the genome Link to datasets
	CNCADD	Datasets from calculation of local mode frequencies (partial vibrational density of states) from classical or ab initio Molecular Dynamics Simulations Link to datasets
	SQP-IRS	The datasets contain the results of the SQP-IRS analysis: Link to datasets
	BioMoFS	BioMoFS produced datasets, containing results of simulations concerning disordered proteins and single - stranded nucleic acids can be found here: Link to datasets
	MDSMS	In the framework MDSMS, three-dimensional structures of various complexes are available in: http://irods.asnet.am/Jmol A visualisation tool for the three-dimensional structures can be accessed here: http://irods.asnet.am/Jmol/jsmol.htm

Table 10: Regional community datasets on life sciences available through the VRE platform

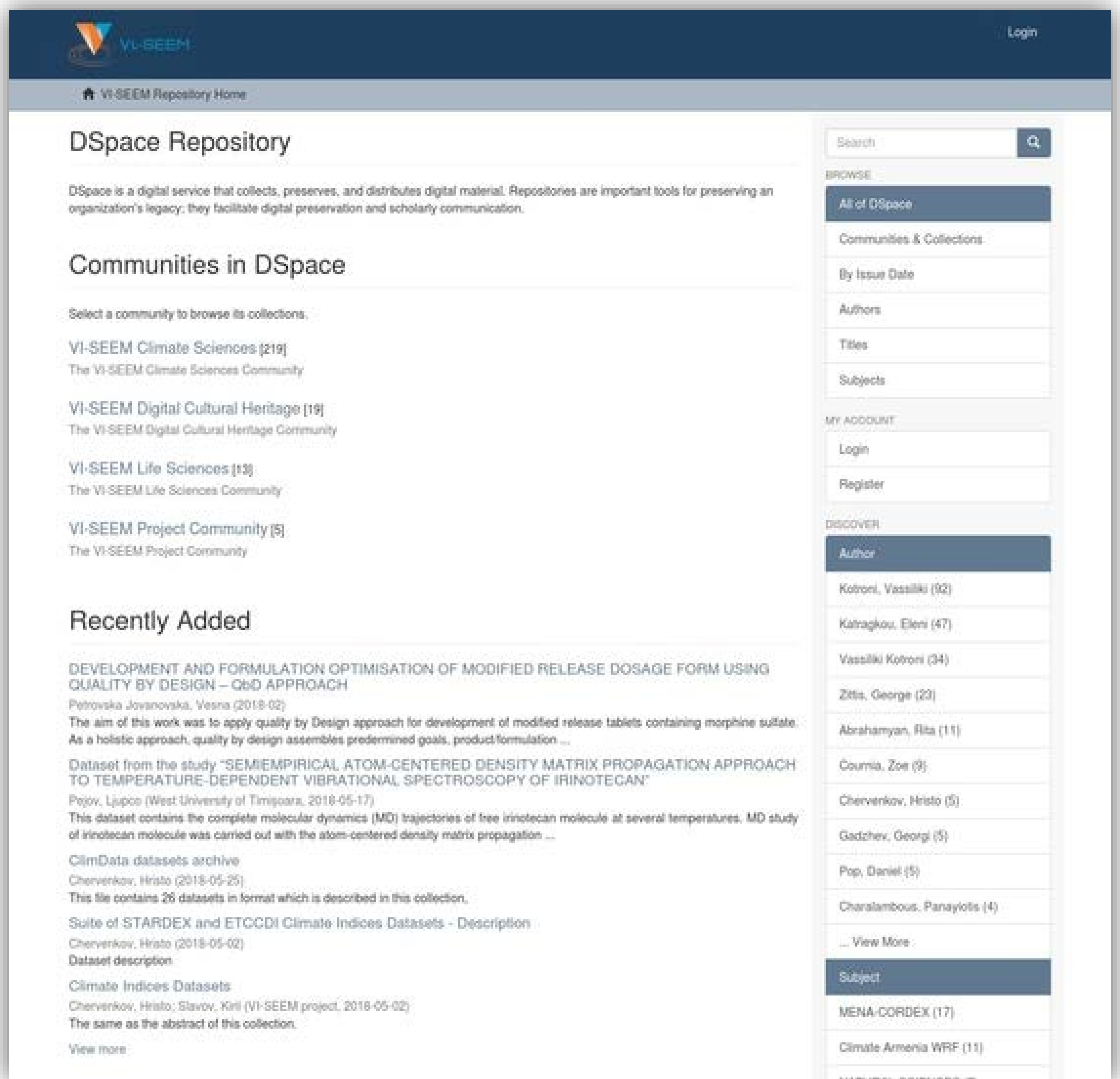


Figure 31: The home page of the VI-SEEM repository

5.3 Access through the VRE platform

Accessing datasets

The main page of the "regional community datasets" service is shown in Figure 32. The service initially provides some general information about the types of data available through VI-SEEM and the corresponding target user categories. For data producers interested in making their dataset available to the VI-SEEM communities, a link to the dataset uploading instructions is provided. Finally, links to the available datasets for the three scientific communities are provided.

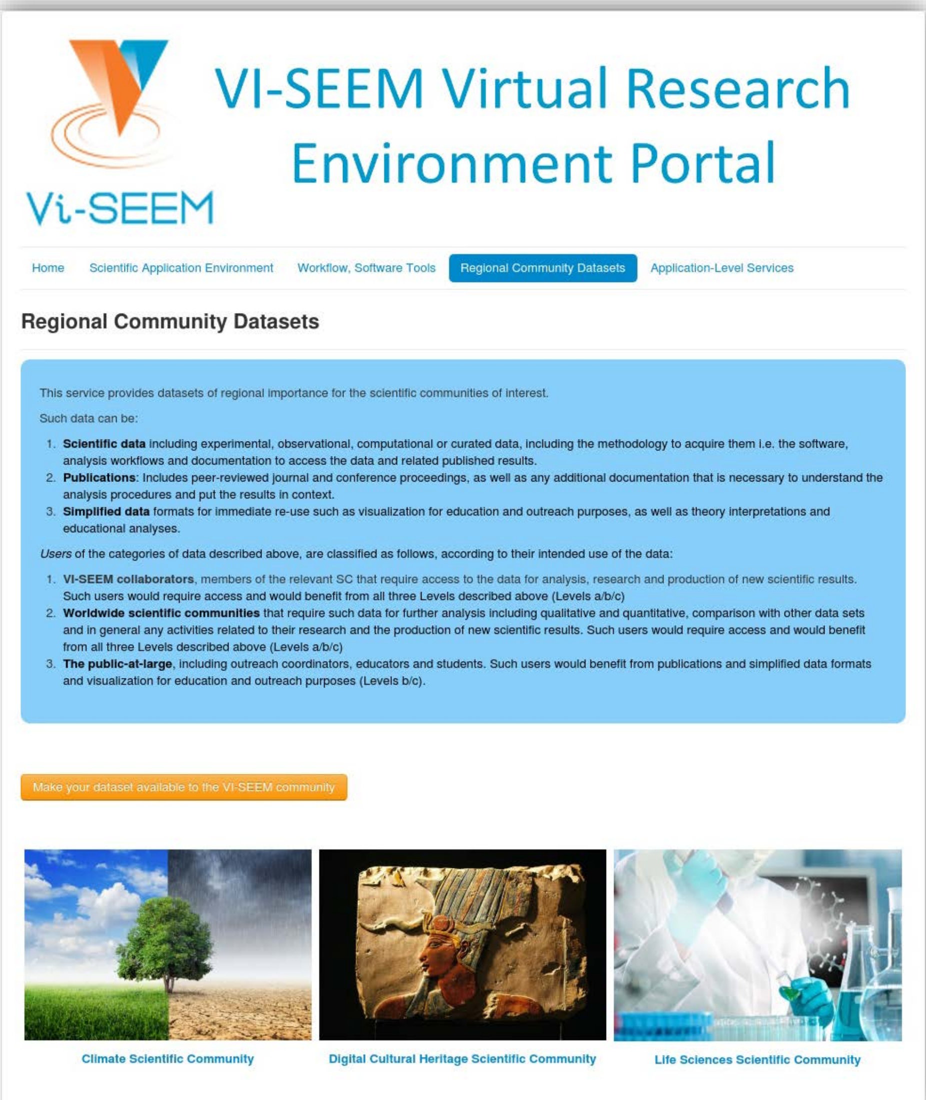


Figure 32: Regional community datasets service main page

For each scientific community the available datasets are listed in a different page in the VRE platform. Figure 33 shows an example of the page using the climate scientific community. The page explains the types of datasets provided by the scientific community and the structure of the metadata accompanying each dataset. A list of the available datasets is given for each dataset type.

Continuing the example for the climate scientific community Figure 34 shows the description for the climate simulation data; figure also depicts how available datasets are shown to the user. A direct link to the list of provided datasets is given through the orange button. For each dataset, a small description is provided to help user determine whether the dataset is of use to him. Additionally, metadata regarding the openness, preservation, pre-processing and responsible scientists are given for the dataset.

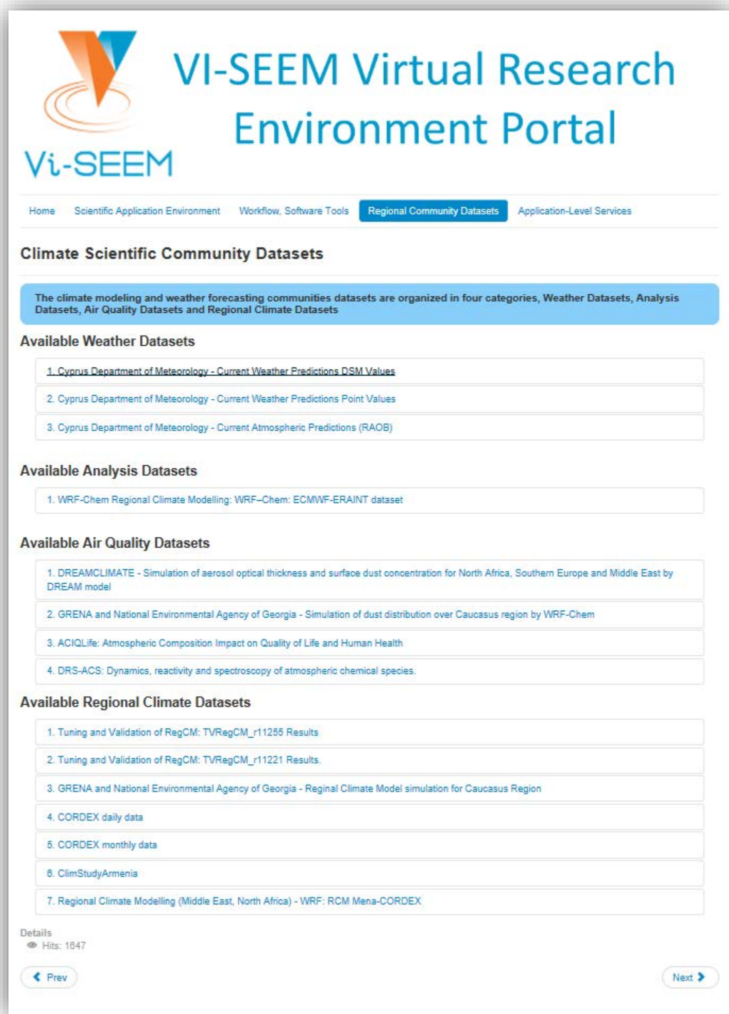


Figure 33: Climate scientific community datasets main page

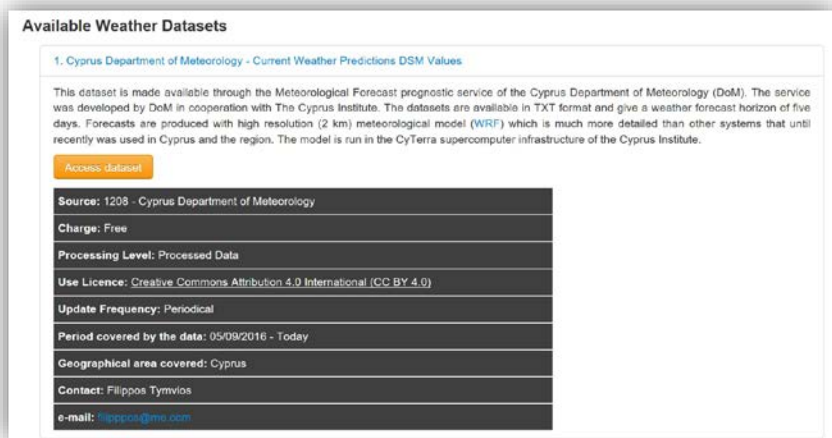


Figure 34: Climate scientific community dataset page showing a description of the simulation data for climate research

Making datasets available to the VRE platform

A "how to contribute" page is provided to the data producers to ease the process of making datasets available to the VI-SEEM communities. The page has detailed instructions data producers need to follow and the information they need to provide in order for the dataset to be made available (Figure 35https://vre.vi-seem.eu/index.php/regional-community-datasets/how-to-contribute"https://vre.vi-seem.eu/index.php/regional-community-datasets/how-to-contribute. Instruction on how to use the different services to upload datasets are given. For example, Figure 19 shows the guidelines for adding a dataset in the VI-SEEM repository service. The guidelines are available at https://vre.vi-seem.eu/index.php/regional-community-datasets/how-to-contribute?id=96 shows the guidelines for adding a dataset in the VI-SEEM repository service. The guidelines are available at https://vre.vi-seem.eu/index.php/regional-community-datasets/how-to-contribute?id=96

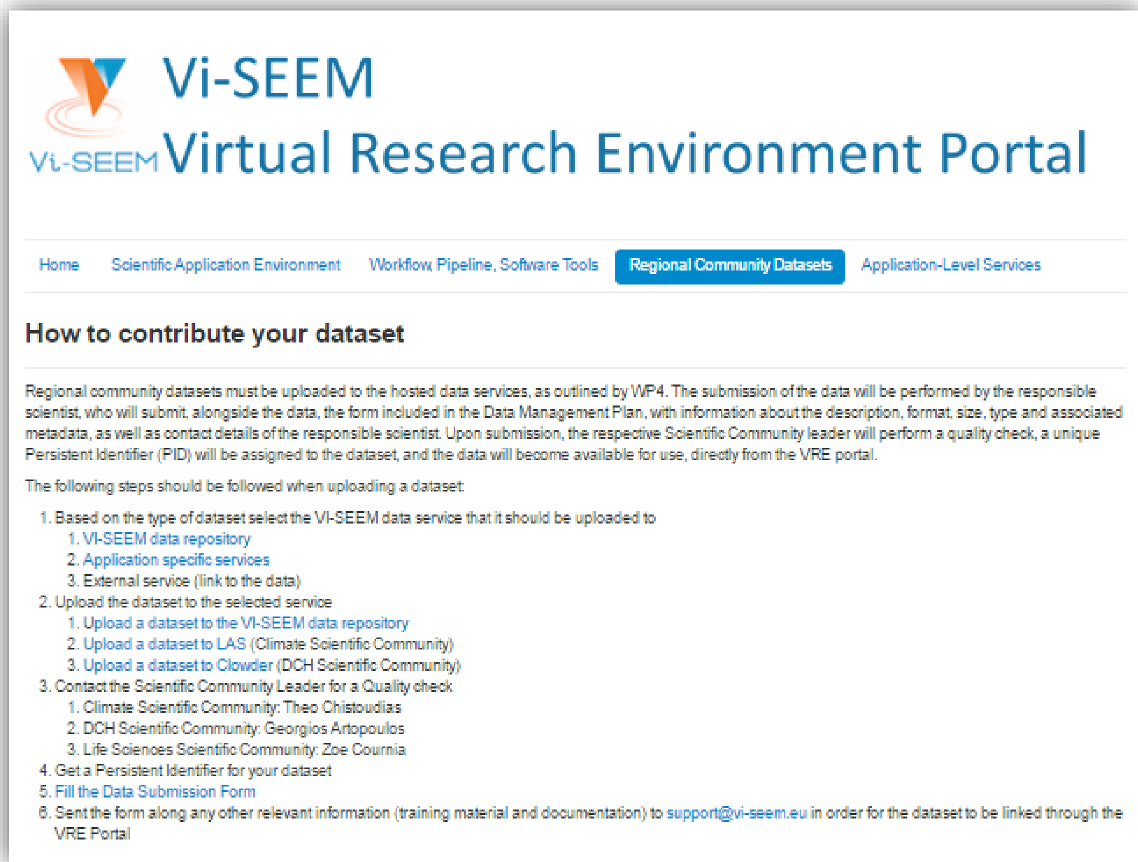


Figure 35: Guidelines for listing datasets through the VRE platform

6 VI-SEEM application-level services

6.1 Scope and service description

This category of services contains web-based or visualization services providing easy access to underlying workflows, applications, and resources. Currently there are eleven available application-level services. In life sciences five such applications-level services exist, in climate just one and in digital cultural heritage five more. These are mostly standalone services, which are providing the necessary hosting and storage infrastructure for their operation through the VI-SEEM generic services. The service allows for uploading and pre-processing datasets that are then made available to the scientific communities. The listing of the application-level services is available at https://vre.vi-seem.eu/index.php/application-level-services-for-the-regional-communities

6.2 Service integration

Currently eleven application-level services are available through the VRE platform. These are the following:

Application Level Services in Life Sciences

6.2.1.1 ChemBioServer

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

It provides researchers with the ability to:

- Browse and visualize compounds along with their properties.
- Filter chemical compounds for a variety of properties such as steric clashes and toxicity.
- Apply perfect match substructure search.
- Cluster compounds according to their physicochemical properties providing representative compounds for each cluster.
- Build custom compound mining pipelines.
- Quantify through property graphs the top-ranking compounds in drug discovery procedures.

The layout of the front page of the ChemBioServer is provided in Figure 36. The ChemBioServer can be accessed at http://chembioserver.vi-seem.eu/.

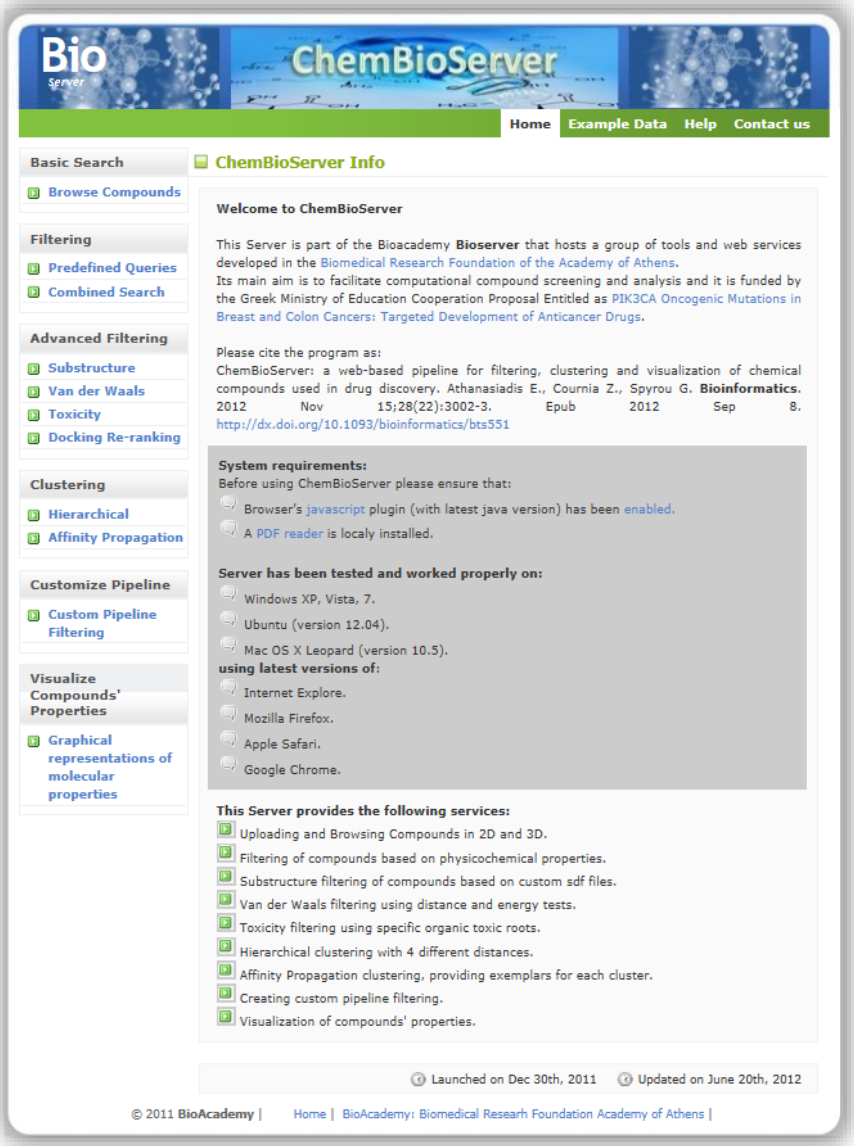


Figure 36: The ChemBioServer layout

6.2.1.2 AFMM

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

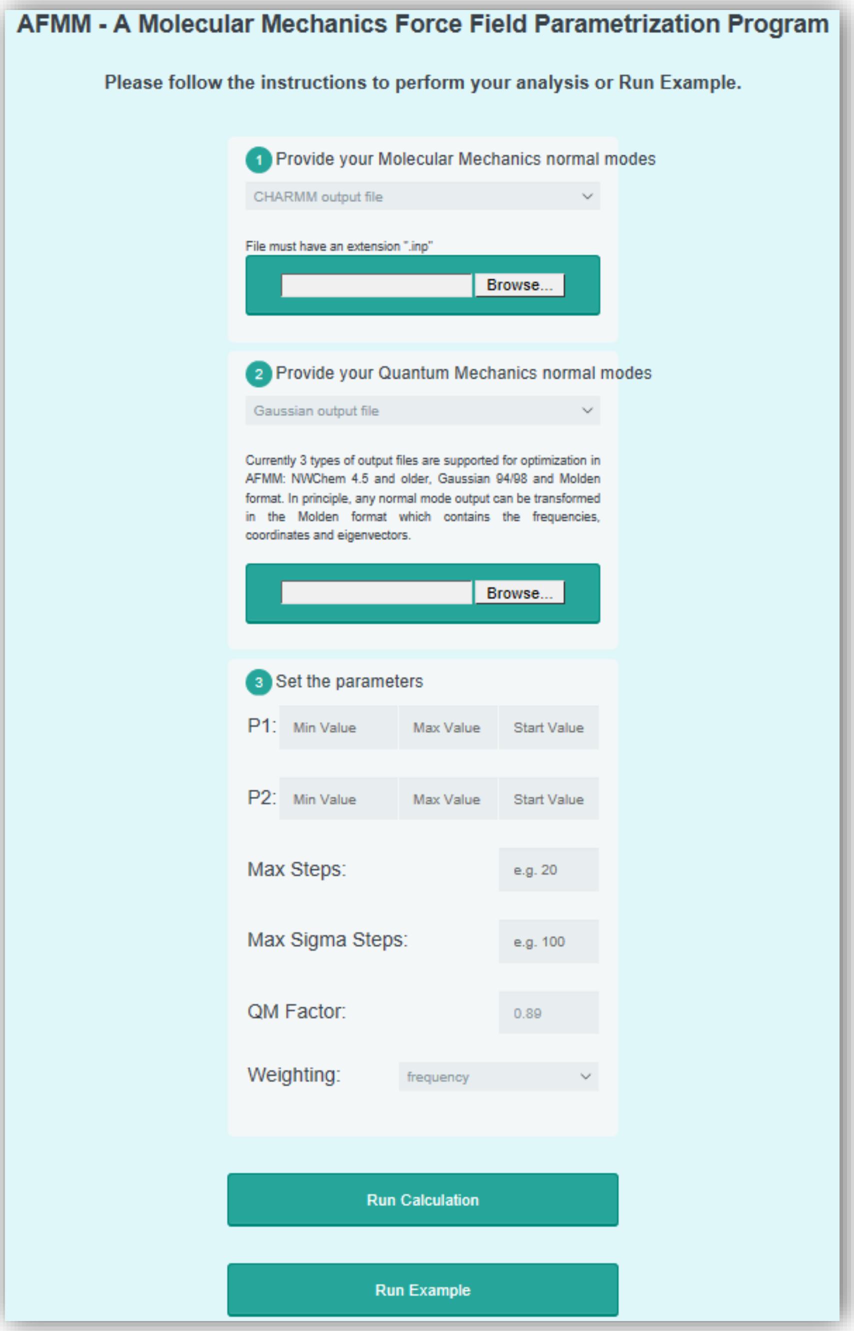


Figure 37: The layout of AFMM

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

The layout of the front page of the AFMM is presented in Figure 37. AFMM can be accessed through http://afmm.vi-seem.eu/.

6.2.1.3 Nano-Crystal

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

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

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

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

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

The layout of the front page of the Nano-Crystal is presented in Figure 38.

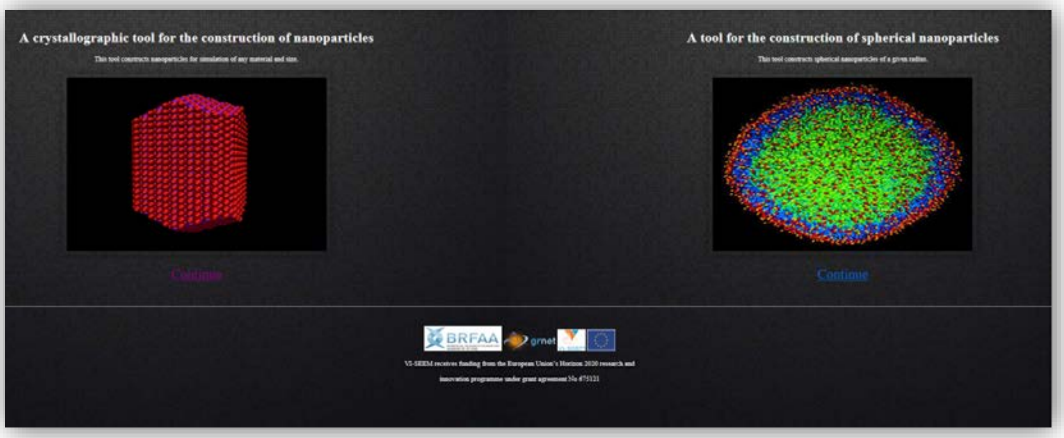


Figure 38: The layout of the Nano-Crystal

6.2.1.4 Subtract

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

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

The layout of the front page of the Subtract application-level service is presented in Figure 39. Subtract can be accessed through http://subtract.vi-seem.eu/.

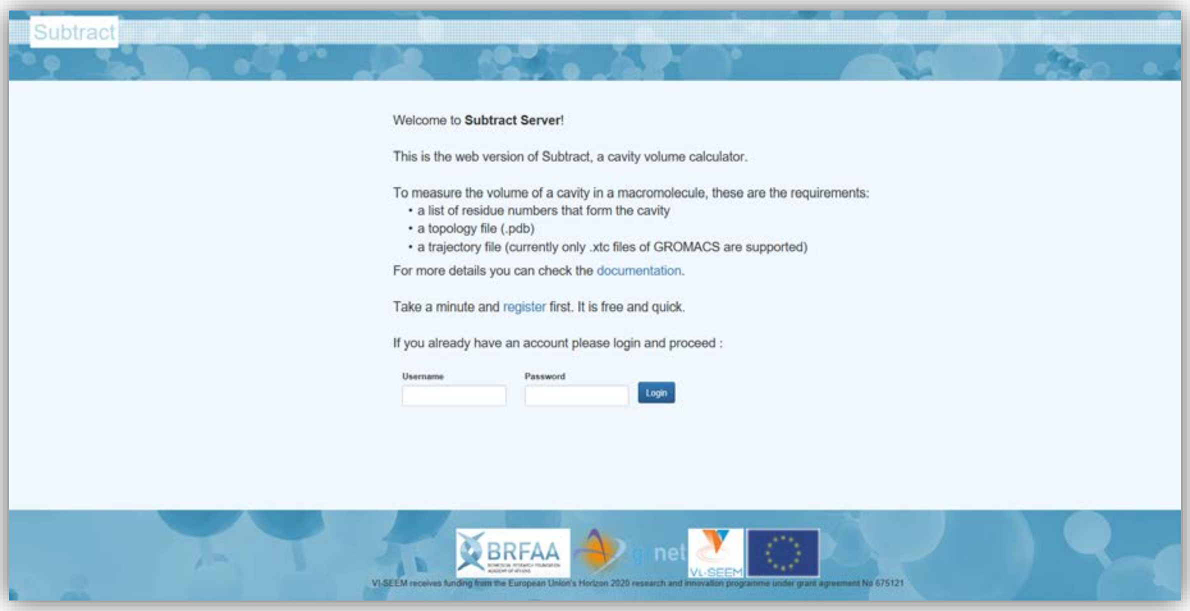


Figure 39: The layout of the Subtract

6.2.1.5 DICOM Network

DICOM stands for Distributed Informational System for Medical and Diagnostically Institutions for collections, processing and visualization of medical images and can be accessed through the link http://viseem.dicom.md/.

"DICOM Network" offers necessary functionality for quality medical management and secure investigation access in ONLINE mode. This helps doctors, specialists and penitents to access, save and document Medical Imaging Investigations and help institutions to reduce investigation costs, enchasing quality of service.

"DICOM Network" provides access to investigations for medical staff with the appropriate access rights and as well as patients to the personal radiography investigations. Nowadays the system collects and processes more than 500 gigabytes of data per month. The system is based on Data Storage and Data Processing components distributed between different processing units and storages, which could be customized using specific interfaces. VI-SEEM integration connects national DICOM Network application, that includes National DICOM Portal http://dicom.md/, with the DICOM Portal installed on VI-SEEM platform resources. DICOM DATA interface grants the interconnectivity for different users of the both portals and allows displaying DICOM investigations using both portals interfaces. Public DICOM Server grants possibility for any VI-SEEM platform member to pull and retrieve the investigations from DICOM Network application and use the developed facilities based on configured access rules. VI-SEEM platform will offer possibility to install and configure publicly available DOCOM Portal that can be used by any interested institutions to store, access and share medical images. Setting up public DICOM Portal for instance will increase the level of access to DICOM investigations and will help to promote DICOM Network services to regional medical research and practicing community. The DICOM layout is provided in Figure 40

The main features of DICOM are the following:

- DICOM system covers all the workflow for processing and documentation of medical investigations from collecting images from equipment to archiving investigation in the patient medical record.
- All the investigations (DICOM Images) are archiving on DICOM Servers, but the information about investigation is stored in DICOM Portal (like www.dicom.md) database. Many DICOM Services can be connected to one DICOM Portal.
- DICOM Portal stores all data like user's info, access info, system settings, DICOM Server settings and some other, but not DICOM images it salves. Each Institution can deploy DICOM Portal internally on own server.
- DICOM DATA Interface collects information about users and investigations from all DICOM Portals and provides functionality for data exchange and unification.
- "DICOM Server" modules can be installed in the same location with the used medical equipment or can be distributed through other institutions and even countries.

Figure 40: The layout of DICOM Network

Application Level Services in Climate

6.2.2.1 Live Access Server

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

LAS enables the web user to:

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

The frontpage of the Live Access Server is presented in Figure 41.

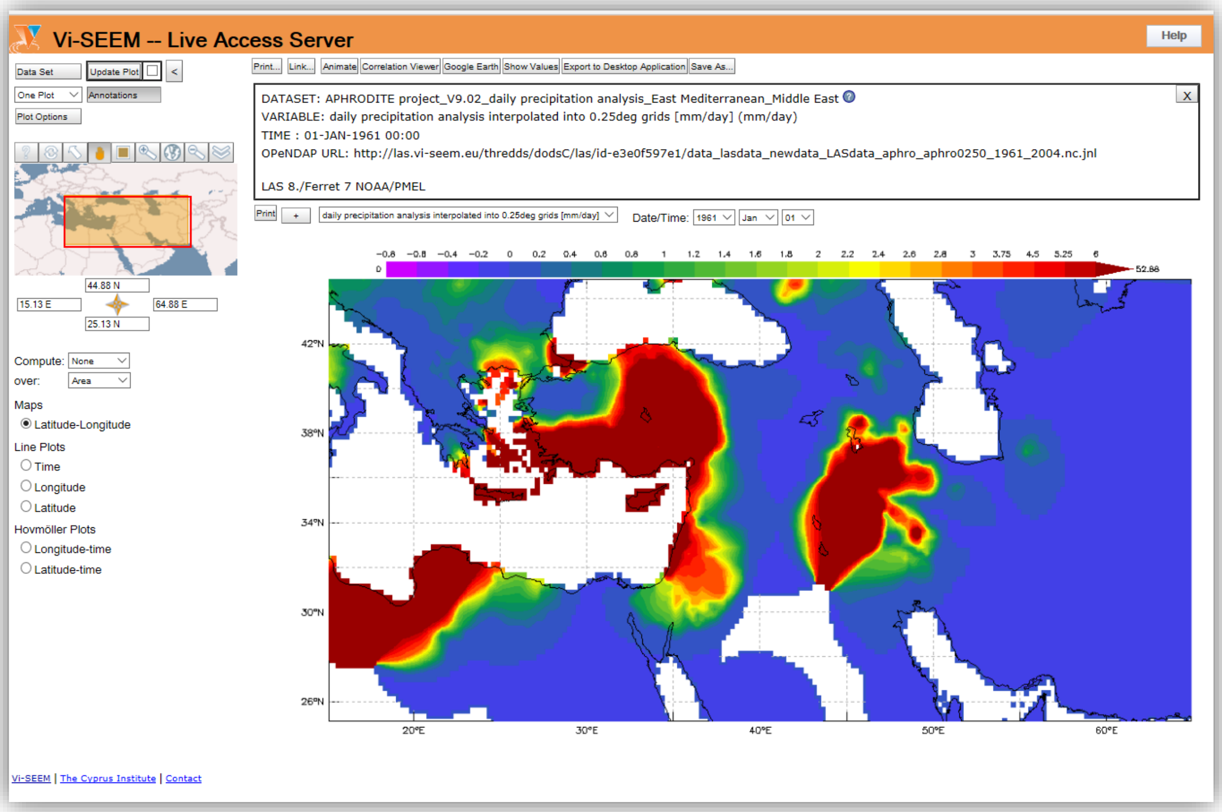


Figure 41: The layout of the Live Access Server

Application Level Services in Digital Cultural Heritage

6.2.3.1 VI-SEEM Clowder

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

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

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

- 3D Models: where extractors clean up and prepare for visualization on the platform itself.
- Scanned books and their metadata: OCR algorithms will be used to extract the text in the documents so that users can find books using both metadata information and the book's contents.
- Image, text, video and sound files and their metadata, organized in collections.

- Advanced documentation data, such as computational imaging (Reflectance Transformation imaging), and analysis of material properties of structures, works of art and artefacts.

The frontpage of the VI-SEEM Clowder is presented in Figure 42.

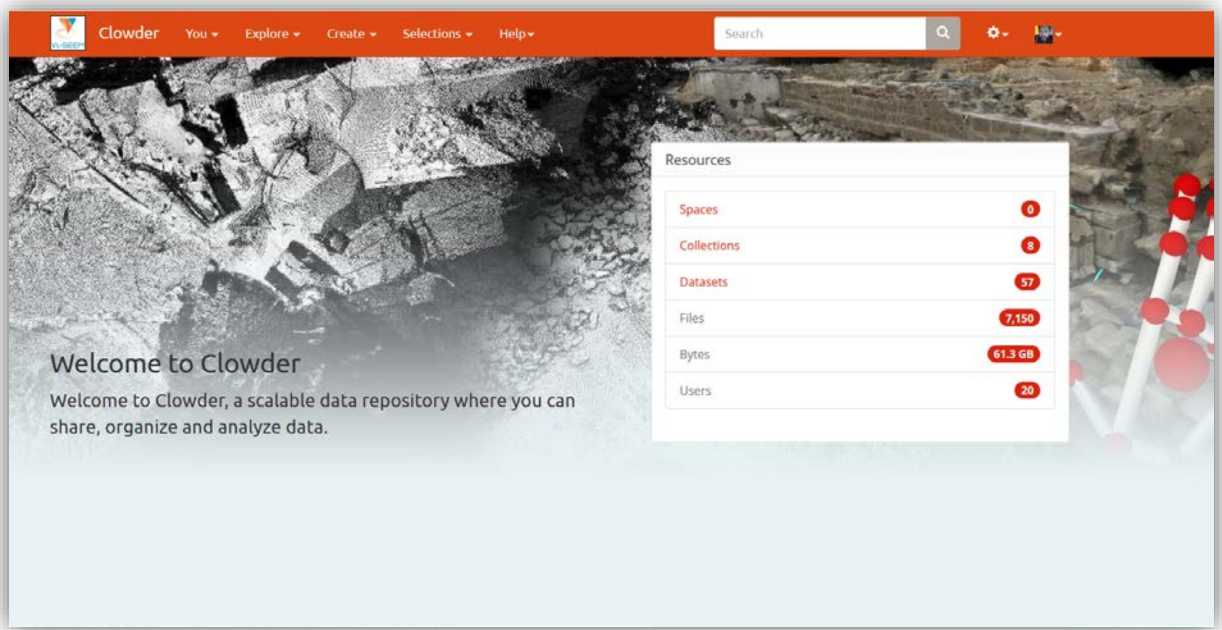


Figure 42: The layout of the VI-SEEM Clowder

6.2.1.2 VI-SEEM CHERE

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

The main features of CHERE are the following:

- Structure-from-Motion reconstruction of 3D objects (online service and docker container).
- Online 3D mesh editing and measurements.
- Generate progressive 360 panoramas from equirectangular 360x180 photos.
- Management of digital cultural items and item sets.

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

The Login page of CHERE is shown in Figure 43.

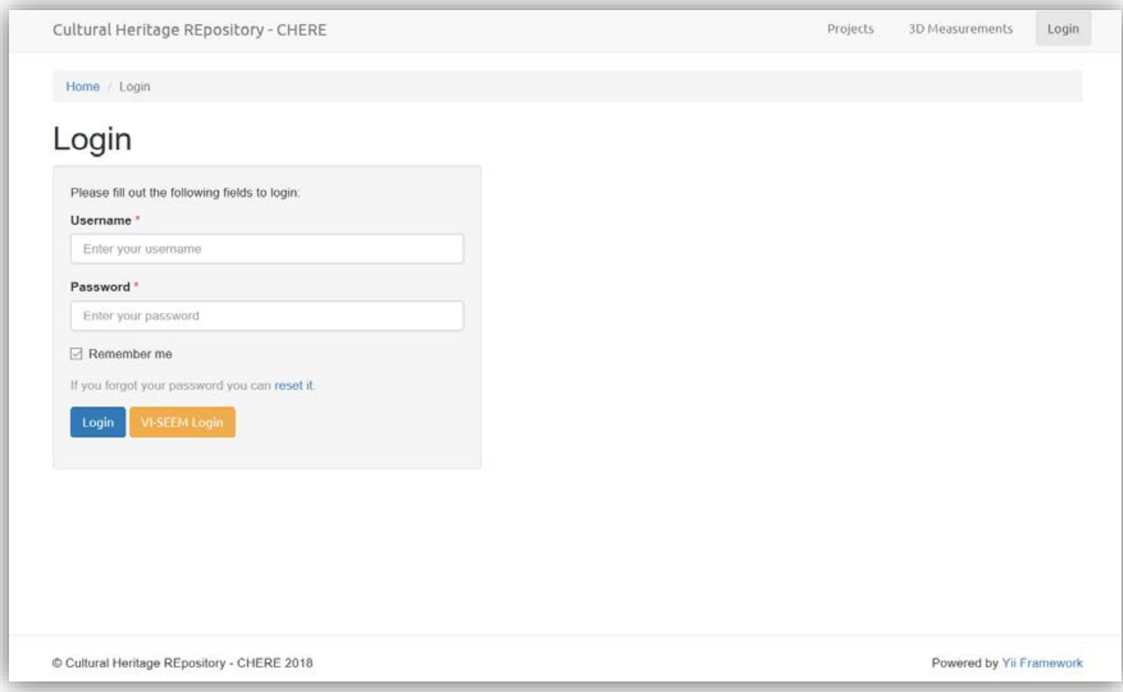


Figure 43: The layout of CHERE

6.2.1.3 VI-SEEM ARCHES – MC4CH

MC4CH (Management Cloud for Cultural Heritage) is a service which significantly lowers the required technical knowledge required to start using digital repositories and conduct structure from motion reconstructions without installation of additional software. It is aimed at museum staff with limited technical know-how but can also be used in standalone version via Docker containers for more technically inclined users. It can generate tiled models of panoramas suitable for web consumption without the restriction of maximum resolution supported by the browser.

Figure 44: The layout of Management Cloud for Cultural Heritage

6.2.1.4 3DINV

3DINV standing for three-dimensional (3-D) inversion of surface Electrical Resistivity Tomography (ERT) is an executable code embedded into the Clowder platform.

3DINV is a program for the three-dimensional (3-D) inversion of surface Electrical Resistivity Tomography (ERT) data in order to automatically determine a 3-D resistivity subsurface model. The program performs smoothness constrained (Occam's) inversion in order to address the non-uniqueness of the inverse problem and stabilize the procedure. The subsurface is divided in homogeneous and isotropic hexahedral elements and a 3-D Finite Element Method (FEM) routine is employed to calculate the resistivity response of 3-D bodies. The adjoint equation technique has been incorporated into the FEM scheme to calculate the Jacobian matrix. An iterative method (LSMR) is used to update the resistivity model though the inversion iterations.

The program can cope with 3-D data sets measured using any kind of conventional (Dipole-Dipole, Pole-Dipole, Pole-Pole, Gradient, Schlumberger, Wenner), non-conventional and mixed surface arrays (e.g. Dipole-Dipole with Pole-Dipole). The ideal 3-D survey employs the arrangement of the electrodes in a rectangular grid and gathering the measurements along all possible directions. An alternative and nowadays most common strategy is to gather the measurements along two perpendicular directions (X and Y survey) or along a single direction (e.g. X-survey). In 3-D surveys composed of parallel two-dimensional (2-D) lines the basic inter-electrode spacing should be equal to the inter-line distance to ensure the true "3-D" coverage of the subsurface resistivity properties.

Although, 3DINV is embedded in the VI-SEEM Clowder it can also be used as a software in private computers.

6.2.1.5 AutoGR

AUTOGR is an executable code embedded into the Clowder platform which makes use of the AutoGR-SIFT.

AutoGR-SIFT is a module of the larger software AutoGR-Toolkit (now at version 3.5 - http://www.ims.forth.gr/AutoGR), which is freely available online for download and it is consistently used by Universities, Research Centres, Commercial companies and private users all over the world (at the moment, more than 1000 users are counted).

The toolkit was intended for researchers dealing with large collection of images to be used in a GIS environment, with velocity and accuracy as keywords.

Most of the geo-referencing procedures are normally based on subjective feedback and visual evaluation of algorithm approximation. Typical is the case of an aerial image georeferenced just by clicking on few identified features (crossroads, field boundaries or even ground-measured targets) on the given image and the corresponding points in real world coordinates on an orthophoto or a map. In general terms, if higher accuracy is required, more common points should be found between the given image and the base map/orthophoto used for geo-referencing. If we multiply these points for all images in a dataset (typically made of hundreds images, especially when dealing with photographs captured at regular intervals with a flying platform such as a drone), we immediately get to a load of work that can hardly be handled in minutes.

AutoGR-SIFT takes the move from the branch of Computer Vision dealing with Image Matching. According to the most recent developments in IT, if two images can be "described" on a pixel-by pixel base, they can also be compared and (ideally) matched with some level of tolerance with regards to the reciprocal scale and orientation.

6.3 Access through the VRE platform

Accessing available services

The main page of the "application-level services" category, as depicted in Figure 45, provides a short introduction on the services and easy one-click access to the available services. The page also provides a link to specific information for making new application-level services available through the VRE platform. Following the links and through the menu, the user can get more information and access to the available application-level services. The available services in each community are listed along a short description. Direct access is provided to the service. Each service is also registered in the VI-SEEM service catalogue [18] offering more information about the service current status, including version and support information. As an example, in Figure 46, we show the information provided to the VI-SEEM platform user for the available application-level services for the scientific community of Life Sciences.

Figure 45: VRE platform main page for application-level services

Figure 46: Available life sciences application-level services

Contributing new application-level services

A specific "how to contribute" page is available to the interested application-level service owner to ease the process of listing a service to the VI-SEEM communities. The page lists in an easy to follow manner the needed steps and information the service owner needs to follow and provide in order for the application-level service to be made available (Figure 47). The detailed guidelines can be found at

https://vre.vi-seem.eu/index.php/application-level-services-for-the-regional-communities/how-to-contribute

Figure 47: Guidelines for listing an application-level service to the VRE platform

7 Update of existing and integration of new services

7.1 Scope of this section

As we have explained in the previous chapters, the VI-SEEM thematic services have been developed and integrated throughout the three integration phases. A pictorial representation of these three phases in provided in Figure 19. These services included software tools, workflows, codes and datasets. Although the integration phases have ended, more services are expected to enrich further the list of new software tools, workflows, codes and datasets and update the existing services. These new services and updates are expected to be provided throughout the three open calls for projects accessing the VI-SEEM services and associated infrastructure. These open calls started the production of results on January 2017 and are expected to finish in May 2020. A diagram revealing the plan for the three open calls is provided in Figure 48. The initial stage of each open call denotes the allocation start date of awarded proposals and the final stage the submission of the final report for each project.

A number of projects participating in these three open calls are expected to contribute in the workflows and software tools service as well as in the regional community dataset service. However, the principal investigators have not explicitly defined what they will provide in the Virtual Research Environment. Hence, it is clear that throughout the open calls more applications will be integrated in the Virtual Research Environment, albeit without knowing what these services will provide.

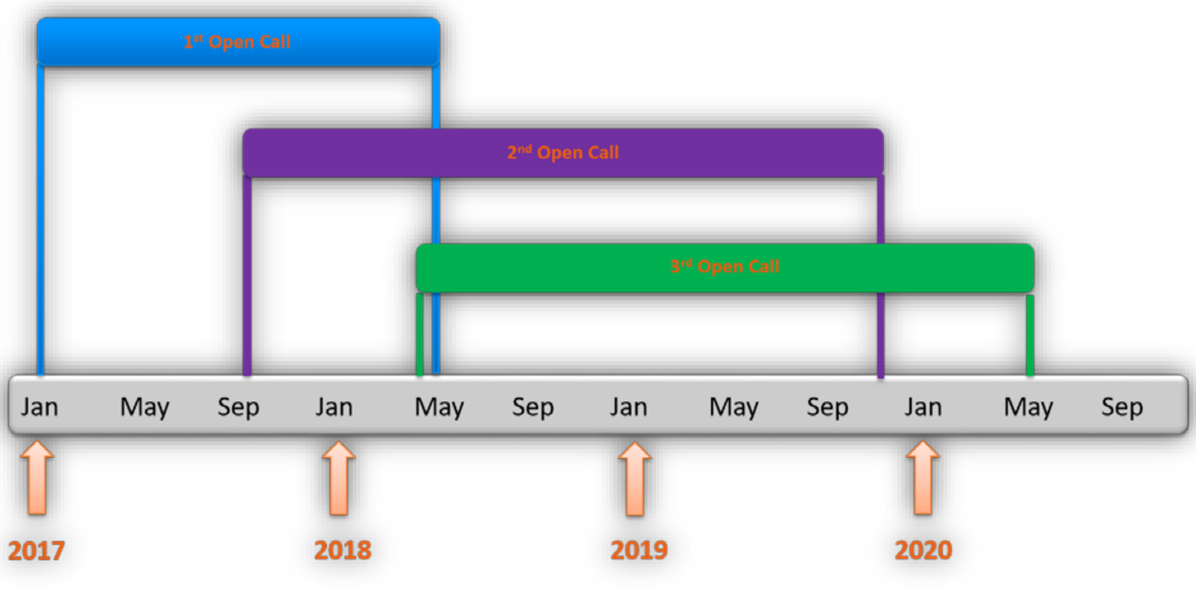


Figure 48: A figure representing the progress of the 3 open calls.

7.2 Integrated applications from the 1st open call projects

The 1st open call ended on May 2018. Throughout the 1st open call a number of applications have been integrated. Namely, five applications have been integrated in the Virtual Research Environment. Climate community provided two new applications, namely AUTH_WRF371M and Continuous_LST. In Digital Cultural Heritage, there were no new contributions, while in Life Sciences three more applications have been added in the VRE applications, namely MULTIDRUG, D3R and MolSurf. More details regarding the projects participating in the open call and contribute in the Virtual Research Environment are provided in Table 11. The descriptions of all the projects participated in the 1st open call can be found in the deliverable D6.2.

	Application Acronym	Contribution
Climate	AUTH_WRF371M	AUTH_WRF371M has contributed with datasets
	Continuous_LST	Continuous_LST has contributed with more datasets
Life Sciences	MULTIDRUG	MULTIDRUG has provided workflows, and codes
	D3R	D3R have provided workflows, datasets as well as training material into the Virtual Research Environment
	MolSurf	MolSurf has provided workflows, and codes

Table 11: The projects participating in the 1st open call and providing new services in the Virtual Research Environment

7.3 Potential new services from the 2nd open call

The 2nd open call is ongoing and is expected to end in November 2019 as the diagram of Figure 48 suggests. In total 18 projects are participating in the 2nd open call. A number of these projects are based on applications that have been embedded into the VRE via the integration phases. These applications are expected to provide only updates on existing datasets, codes as well as workflows. However, a number of projects are completely “fresh” and are expected to enrich further the Virtual Research Environment. There are 12 such projects, the acronyms of which can be found in the following table (Table 12). The descriptions of all the projects participated in the 2nd open call can be found in the Deliverable D6.2.

	Application Acronym	Expected Contribution
Climate	WheAirCYEM	WheAirCYEM can contribute with datasets
	WRF-Solar	WRF-Solar can contribute with datasets
	3DVAR_WRF	3DVAR_WRF is expected to provide datasets

	NBBM4RHMS	NBBM4RHMS is expected to provide datasets
DCH	DCH	DCH has committed to provide datasets. Part of the dataset DCH will proved can already be found in Clowder. More information on this project can be found on section 8.6.2
	DataCrowds	DataCrowds, a cross-disciplinary service, has committed to provide datasets and workflows. Part of the datasets DataCrowds will provide can already be found in Clowder and DSpace. More information about DataCrowds can be found on section 8.3.1
	HaPPEn	HaPPEn can provide codes, workflows and datasets
Life Sciences	BRING-MD	BRING-MD can provide workflows and datasets
	OP4D	OP4D is expected to provide workflows
	SIPD	SIPD can provide datasets
	Surf_Prop	Surf_Prop is expected to provide workflows and codes

Table 12: The projects participating in the 2nd open call and could provide new applications in the Virtual Research Environment.

7.4 Potential new services from the 3rd open call

The starting of production of the 3rd open call was April 2018 and it is expected to end on May 2020. In total 22 projects are participating in this call from which 5 from the climate community, 8 from the life sciences, 5 from the Digital Cultural Heritage community and in addition 4 from cross-disciplinary projects. Hence, in addition to the three "traditional" communities VI-SEEM consortium also supported cross-disciplinary projects in the 3rd open call. A number of cross-disciplinary services which already can be found on the VRE portal are described in the next chapter. From these 22 projects, 17 are completely new, meaning that they did not participated in the integration phases. The acronyms of the new projects expected to provide some input into the Virtual Research Environment are presented in Table 13. The descriptions of all the projects participated in the 3rd open call can be found in the deliverable D6.3.

	Application Acronym	Expected Contribution
Climate	EXMED	EXMED can provide workflows and datasets
	ArmWRF	ArmWRF can workflows and/or datasets
DCH	DESPHOTMAT	DESPHOTMAT can provide datasets and/or workflows
	CDFOLK	CDFOLK can provide datasets
	GAMMOS	GAMMOS can provide workflows and/or datasets
Life Sciences	RNA_LUPUS	RNA_LUPUS can provide workflows and datasets
	EuGenia	EuGenia can potentially provide workflows and datasets
	BSI	BSI is expected to provide a workflow and/or a code

	GQL	GQL can provide workflows and datasets
	AT1R	AT1R could provide datasets and/or workflows
	RXRa	RXRa project already provided datasets.
	CMMC	CMMC can provide datasets
Cross-Disciplinary	ClimData	ClimData project already provided cross-disciplinary datasets. For more information read section 8.4.1.
	SARISI	SARISI has agreed to provide cross-disciplinary datasets
	CAT-ICE	CAT-ICE is expected to provide codes and/or workflows
	CDPrSc	CDPrSc can provide codes/workflows/datasets

Table 13: The projects participating in the 3rd open call and could provide new applications in the Virtual Research Environment

8 Cross-Disciplinary Applications

8.1 Scope of the cross-disciplinary applications chapter

VI-SEEM has a strong cross- as well as inter-disciplinary character. Serious efforts have been invested in bridging different disciplines by simultaneously integrating applications that provide connections between the three communities as well as developing an ontological solution capable of addressing the interoperability between the communities. Hence, a new section dedicated to cross-disciplinary applications has been created in the VRE-portal and since then it has continuously been enriched with new services. Therefore, the purpose of this chapter is to describe all the available cross-disciplinary applications which exist on VI-SEEM.

8.2 Data Visualisation Applications

Nicosia Climate Change Hot Spot: Future's Extremes

Disciplines: Climate + Digital Cultural Heritage

Link: https://vre.vi-seem.eu/index.php/data-visualization#DCHAndClimatology

The first successful cross-disciplinary research activity between the digital cultural heritage and Climate communities that capitalized on recourses offered by the VI-SEEM project was recently exhibited at a popular international venue. The Seoul International Biennale on Architecture & Urbanism is a large-scale public event that addresses themes of particular relevance to urban settings on a global scale. This exhibition is organized by the Seoul Metropolitan Government and Seoul Design Foundation and received 4M visitors over the course of its duration. Titled "Imminent Commons", the exhibition was co-directed by Hyungmin Pai, acclaimed architectural historian, critic, alumnus of MIT and Alejandro Zaera - Polo, the award-winning architect and tenured Professor at Princeton University, U.S.A. From September 1 to November 5, 2017, the Biennale provided a forum for debate to policy makers, experts and citizens at large.

Following the UN's World Urbanization Prospect Report of 2014, 54% of the world's population now live in metropolitan areas. By 2050, this percentage will increase to 86% in advanced countries, and 64% in developing nations. Already now, the MENA region, renowned for its wealth of cultural heritage, ancient civilizations' monuments and major sociocultural developments during medieval and early modern times, is experiencing a high degree of urbanization. The region saw a 400% growth in urban structures during 1970 to 2010. It is expected that over the next 40 years there will be an additional 200% growth of larger cites. Thus, the aforementioned sustainability challenges are of particular relevance to the MENA regions and its larger urban structures, including the rich in heritage cities of eastern Mediterranean region, such as Nicosia. Climate change will have particularly strong manifestations in the lived experience of urban settings (e.g., Lelieveld et al., 20141), and will pose great challenges to the material integrity as well as use of built heritage in these environments.

Nicosia, the capital of the Republic of Cyprus and the only major inland city of this Eastern Mediterranean Island has been continuously inhabited for over 4500 years. Estimated to become a climate change 'hotspot' in the foreseeable future, the people of this city already face the effects of the region's changing weather patterns and climate trends. Recognizing the need to prepare for a range of extreme conditions has become crucial, as projected changes will bring a higher number of extreme weather events, which are expected to threaten the rich heritage of the city. This includes prolonged heat waves with enhance urban warming, urban flooding from high-intensity rainstorms, and deteriorating air quality due to atmospherically transported desert dust from North Africa or the Arabian Peninsula and pollutant influx from neighboring Mediterranean countries in the north and west.

The VI-SEEM presented collaborative research activity in the form of an interactive audiovisual exhibit of immersive simulations that illustrated possible futures of this city, visualizing forthcoming conditions of heat, dust and floods using scientific data of climate observations and (computationally) simulated projections.

The following researchers worked to successfully implement this project: Georgios Artopoulos, Theodoros Christoudias, Panayiotis Charalambous, Colter Wehmeier, Charalambos Ioannou, Charis Iacovou, Harry Varnava, Adriana Bruggeman, Panos Hadjinicolaou, Katerina Charalambous, Jonilda Kushta.

Figure 49 - Visualization of the extreme dust event that took place on September 8, 2015 as seen in virtual reality using Nicosia simulation model.

8.3 Simulation Data

DataCrowds

Disciplines: Climate + Life Sciences + Digital Cultural Heritage

Link: https://vre.vi-seem.eu/index.php/simulation-data#DCHAndLifeSciencesAndClimatology

Today more people are living in urban environments than in rural areas. It is forecasted that 70% of the global population will be living in cities by 2050. This intense urbanization poses huge challenges in overcrowding, segregation, demographics and use of resources. The main goal of this project is to innovate in the unified area of research that is occupied with the transdisciplinary study of crowds in built environment. We envision a web-accessible, social platform that allows researchers from very diverse fields, such as Crowd Simulation, Urban Modeling and Simulation, Pedestrian Dynamics, Computer Graphics, Social Dynamics and Architecture to collaborate, share data and take advantage of each field's breakthroughs in order to contribute more accurate crowd simulations for the future sustainability of urban environments. An example of visualization of such data crowds is presented in Figure 50.

By getting access to share crowd data in a platform where researchers can download, comment and enhance the content, new data-driven methodologies to simulate and understand crowd behavior will be created and existing methods will be enhanced. Better models will help in applying these findings in different scenarios such as evacuations, urban planning, virtual reality, entertainment, robotics, etc. It enables users to form an online collaboration environment to support research communities and activities, and disseminate results.

Figure 50: Visualization of the crowd data at the old town of Nicosia.

8.4 Data Analytics and Processing

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

Disciplines: Climate + Life Sciences

Link: https://vre.vi-seem.eu/index.php/data-analytics-and-processing#LifeSciencesAndClimatology

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

8.5 Geographic Datasets

DroneMapper

Disciplines: Climate + Life Sciences + Digital Cultural Heritage

Link: https://vre.vi-seem.eu/index.php/geographic-datasets-description#Montenegroviseem

DroneMapper is photogrammetry cloud application for automatic image mosaicking and geo-referencing. It supports several users, whereby every user is able to upload a set of captured images via a web interface (Figure 51), begin their processing and make an overview of already created maps. The user can download maps in standard GeoTiff format and use them in one of the existing GIS tools. DroneMapper uses numerous open source image processing tools, while the most computationally demanding among them are able to perform multi-core parallel processing, which provides a better usage of the cloud resources.

DroneMapper has an impact in the general areas of Several Communities, Service, Data, Computing, Operations and marketing. Regarding communities, it is expected to impact the disciplines of digital cultural heritage, weather, life sciences as well as other communities such as agriculture, environment monitoring and protection.

Users can upload images, process them and preview results without big knowledge from image mosaicking and geo-referencing. The processed images can be downloaded in standard formats and used with existing GIS tools. Aggregated and processed data can be easily presented to decision makers and stakeholders.

Figure 51: The front-page of the DroneMapper application

8.6 Analytical Studies and Portfolio

Hamaarag Israel National Natural History Collection

Discipline: Climate + Life Sciences + Digital Cultural Heritage

Link: https://vre.vi-seem.eu/index.php/analytical-studies-portfolio#DCHAndLifeSciencesAndClimatology

A collection, dated back to the beginning of the 20th century representing the flora of the Southern Levant, a culture heritage. It is very important to catalogue the collection and photo the specimens in 3D. The morphological and molecular studies of the specimens will use state of the art technologies. Their innovation is to associate species with their parasite and compare it to modern days in order to determine the changes over time.

DCH: The "Aharoni" Online Digitized Collection

Discipline: Climate + Life Sciences + Digital Cultural Heritage

Link: https://vre.vi-seem.eu/index.php/analytical-studies-portfolio#DCHAndLifeSciencesAndClimatology

The "Aharoni" Online Digitized Collection is an innovative project aimed at creating a suitable platform for presenting and preserving the greatest Levantine faunal collection from the beginning of the 20^{th} century. Natural History collections are rich repositories that document our planet's past and present ecosystems and represent a monumental societal investment in research and applied environmental science. The National Natural History collections at the Hebrew University curate sole collections, The "Aharoni" collections, that are comprised of unique fauna (avian, amphibian, reptiles and mammalian) collections and archive material, are the sole direct evidence of the species richness and biodiversity of the Levant region at the beginning of the 20^{th} century. The archival material includes historical documents describing Aharoni' s numerous zoological expeditions through the Levant, his comments on specimens, ecological description of habitat that are gone, buried under cities and roads, and other valuable information associated with his studies.

8.7 Source Code

X-H-stretching-potential

Discipline: Climate + Life Sciences

Link: https://vre.vi-seem.eu/index.php/source-code#LifeSciencesAndClimatology

X-H-stretching-potentials is a program which generates geometries/configurations to compute X-H stretching potentials in the case of X-H stretching oscillators for free and interacting molecular species relevant to atmospheric chemistry and climate science. Geometries are generated such as to mimic as closely as possible the realistic X-H motion within the local mode approximation, keeping the center-of-mass fixed. First, a one should define the number of configurations that need to be generated, the atomic number of atom X (oxygen in the present case), of the atomic number of atom H, the atomic mass of atom X and the atomic mass of atom H. The input of this program is a file that contains the cartesian coordinates of the X and H atom (e.g. from an ab initio optimized molecular geometry). After the file is read, a one should define the increment. Next, the program calculates the following: position of the center of mass, the parameters of the unit vector along the X-H direction with respect to the motion with the center-of-mass fixed (unit vector parameter for X atom and for H atom). At the end, the program calculates the novel coordinates for H and X atoms and it writes an output. The program is written in Fortran and it requires GNU Fortran compiler. Examples of input (test-input-OH.txt) and output file (test-output-OH.out) are provided. To run the program the following command should be executed: ifort xh-stretch-1.f -o out

DSchrodingerEq

Description: Climate + Life Science

Link: https://vre.vi-seem.eu/index.php/source-code#LifeSciencesAndClimatology

2DSchrodingerEq is a program for solving 2D Schrödinger equation. It provides a solution for the time-independent Schrödinger equation for certain values of energy by finding the eigenvalues of energy. The DVR method for solving one-dimensional and two-dimensional Schrödinger equations uses the matrix representation and is based on the truncated standard orthonormal polynomial bases and the corresponding Gaussian quadratures. The program contains the following modules:

1. Addition of two matrices - (input: 2D array A and 2D array B, output: 2D array A+B)
2. Multiplication of two matrices - (input: 2D array A and 2D array B, output: 2D array A × B)
3. Multiplication of scalar and matrix - (input: 2D array A and scalar a, output: 2D array a × A)
4. Matrix diagonalization - (input: 2D array A, output: 2D array - diagonalized A)
5. Matrix transposition - (input: 2D array A, output: 2D array - transposed A)
6. Generating identity matrix of size n, (input 2D array A, number n, output: 2D array-identity matrix of size n)
7. Division of vector by scalar - (input vector v, scalar a, output: the vector v divided by a)
8. Reading file content - (input: string -file path, output: string - file content)
9. Printing matrix elements - (input: 2D array A, output: string - content of 2D array A)
10. File rows comparison - (input: file, output: Boolean)

11. File rows sorting - (input: file, output: sorted rows)
12. Calculation of eigenvalues (input - 2D array, output - 1D array)
13. Transformation and calculation of array of points specified in finite basis representation (FBR) to discrete value representation (DVR) (input - number of points, min and max values for the points, output - arrays with points in discrete values representation (two dimensional arrays for solving two dimensional Schrödinger equation).

The requirements for this program are gcc compiler and GSL (GNU Scientific Library) library. An input file example (q1q2.kcal.mol.dat-0GPa) with x and y coordinates and energies is provided. To run the program, a one should first compile the program: gcc 2DSchrodinger.c -lgsl -lblas -o out and then execute the binary file ./out

bound-states-of-the-Morse-well

Disciplines: Climate + Life Sciences

Link: https://vre.vi-seem.eu/index.php/source-code#LifeSciencesAndClimatology

bound-states-of-the-Morse-well is a program that *solves the bound states of the* Morse oscillator well. In order to solve the bound states of the Morse oscillator well, a 100-point basis ranging from [-3,32] is used. First, the Hamiltonian matrix in the DVR basis is constructed and then it is diagonalized to determine the eigenvalues and eigenvectors. The eigenvalues (energies) below 0 are bound states. The eigenvectors and eigenvalues are used to plot the wave functions. The part for checking the correctness of wavefunctions by checking the values of the first two eigenvectors is automated. This code is written in C language and it is based on the code written in Mathematica (Bittner, Eric R. Quantum dynamics: applications in biological and materials systems. CRC Press, 2009.). DVR (Discrete Value Representation) Hamiltonian is constructed by constructing a diagonal matrix of the potential evaluated over the DVR points. Then, the kinetic energy contribution is added by transforming the diagonal matrix (fbrke) from the basis representation to the DVR. The requirements for this program are gcc compiler and GSL (GNU Scientific Library) library. The command to compile the program is gcc morse.c -lgsl -lblas -o out and to execute the binary file ./out.

pACF

Disciplines: Climate + Life Sciences

Link: https://vre.vi-seem.eu/index.php/source-code#LifeSciencesAndClimatology

pACF is a program to compute the partial autocorrelation function pACF of the time-derivative of the X-H bond length in order to further calculate the partial vibrational density of states corresponding to the local X-H stretching mode, i.e. the contribution of this motion to the overall vibrational density of states – vDOS. First, one should define the number of steps for the computation of the pACF. Next, the following input data should be read: xt(i) - time; xr(i) - X-H distance; xdr(i) - time derivative of the X-H distance. Then the computation of the pACF is performed. Next, a one should provide a definition of the maximum number of lags that are physically/numerically acceptable. After the calculations, the following output is written: xt(i) - time; xc(i) - autocorrelation function; xcnor(i) - normalized autocorrelation function. The program is written in Fortran and it requires GNU Fortran compiler. To run the program a one should compile the program: gfortran pacf.f -o program.out and then execute the binary file ./program.out.

8.8 Towards an ontological solution for VI-SEEM services

Different scientific communities produce different kinds of datasets that rely on different data descriptions, approaches, and logical organizations. Hence, in such an environment it is important to establish a knowledge communication framework that is able to guarantee some fundamentals, such as an inclusive description and documentation of the inter-disciplinary digital resources, their long-term preservation, access, use and reuse.

In this vein, VI-SEEM aims at developing a solution in order to harmonize that variety of approaches and descriptions in the three different research communities, as well as providing a semantic tool for retrieval and discovery of relations aimed at an inter-disciplinary use of their content. In this way, data can be shared by different disciplines and used in different ways.

For what concerns the semantic interoperability of databases and datasets integrated into the VI-SEEM Virtual Research Environment, a methodology and a process towards the development of a cross-disciplinary ontological solution has been set up.

The first step of the research consists of the development of a methodology and a process towards the development of such a cross-disciplinary ontological solution.

In the Deliverable D5.2 “Data Management Plans”, the first step of this research was reported. This step required to adopt a strategy to understand the diversity of the provided information. To address the questions about the data available WP5 took the initiative and performed a Data Management Plan survey. After collecting the contributions from the partners we analysed the responses and provided directions regarding the progress of the development of the framework to address the cross-disciplinary structure.

A number of issues came out from the first survey. Specifically for what concerns the metadata structures of the data in VI-SEEM, it appeared that:

- Some datasets come without metadata nor a particular metadata schema. This provides inconsistencies in the way one handles different metadata formats.
- There are similarities between the Climate and Life Sciences metadata formats in the sense that most of the metadata schemas are produced by software, while in the case of Digital Cultural Heritage metadata structure requires human input (html, Text formats).

For more detailed report on this survey, we refer to D5.2 “Data Management Plans”.

The following section presents the methodology and the process towards the development of such a cross-disciplinary ontological solution within VI-SEEM. The steps of the research consist of:

- Development of the methodology itself.
- The state of the art about the Vi-SEEM registries and datasets metadata descriptions.
- Identification of cross- /inter-disciplinary datasets within the three scientific communities that can be used as case studies for the development of the ontological solution.
- Analysis of the DSpace platform (the VI-SEEM repository) to investigate the possibility of developing the ontology on the base of its entities (the analysis involved also the assessment of the users' needs related to the descriptions and retrieval of their data).
- Finally the development of the VI-SEEM ontological solution, given an assessment of existent ontologies.

The efforts towards the development of an ontological solution gained from the organization of an event dedicated to data management and semantic structures for cross-disciplinary research. The workshop, namely "Data Management and Semantic Structures for Cross-disciplinary Research in the South East European and Eastern Mediterranean (SEEM) region", was held at the Cyprus Institute on 25-27 June 2018. The Workshop aimed at bringing users and experts from different scientific communities together, by providing a stage for sharing knowledge, good practices and experiences towards the development of the Vi-SEEM ontological solution. The workshop was attended by experts in the field of data management; various examples from other digital research infrastructures and case studies in different fields were presented and discussed. Such workshop highlighted the challenging areas in information integration and provided an overview of some available solutions for cross-domain applications, as well as tools to facilitate the construction of semantic cross-disciplinary databases for data that can be used as the information integration backbone for e-Research Infrastructure.

The ontological solution is conceptualized as an inter- and cross-disciplinary tool in order to:

- Access digital resources
- Enable data interoperability
- Allow retrieving of 2D/3D objects, Decisions, Activities, Events, Subjects
- Digital provenance
- Data interpretation
- Data re-use

The last step of the methodology consists of the alignment/mapping of the metadata fields of the datasets chosen as case studies and as a starting point towards the development of the cross-disciplinary ontological solution for the VI-SEEM infrastructure.

To the final development of the cross-disciplinary ontological solution, an assessment of existing semantic solution has been performed. Moreover, an evaluation of available metadata mapping and ontology creation tools has been accomplished.

This phase of the work consists of the following three steps:

- Choice of a data oriented procedure

- Choice of cross-disciplinary case studies within VI-SEEM
- Use of a stable and adaptable (for possible future development) ontological solution

The procedure followed towards the development of a cross-disciplinary solution is a “data oriented” one. The procedure is in fact based directly on the data available in VI-SEEM. Since it is a procedure that starts from the data, we therefore identified case studies that could cover the scientific research fields participating in the project. These datasets are used as case scenario of the ontological solution within the infrastructure, to test the benefits of cross-domain research.

The ontological solution is tested on two case studies: the “Aharoni” Online Digitized collection and on the Banja Luka museums datasets.

The “Aharoni Online Digitized” collection (Figure 52) consists of an interesting case of split collection into different countries and reunified digitally. It consists of 12 collections that cover various scientific fields (paleontology, life science, ecology, cultural heritage etc.). It presents simple metadata fields (not attributable to a standardized metadata schema), enriched by the presence of an extensive taxonomy.

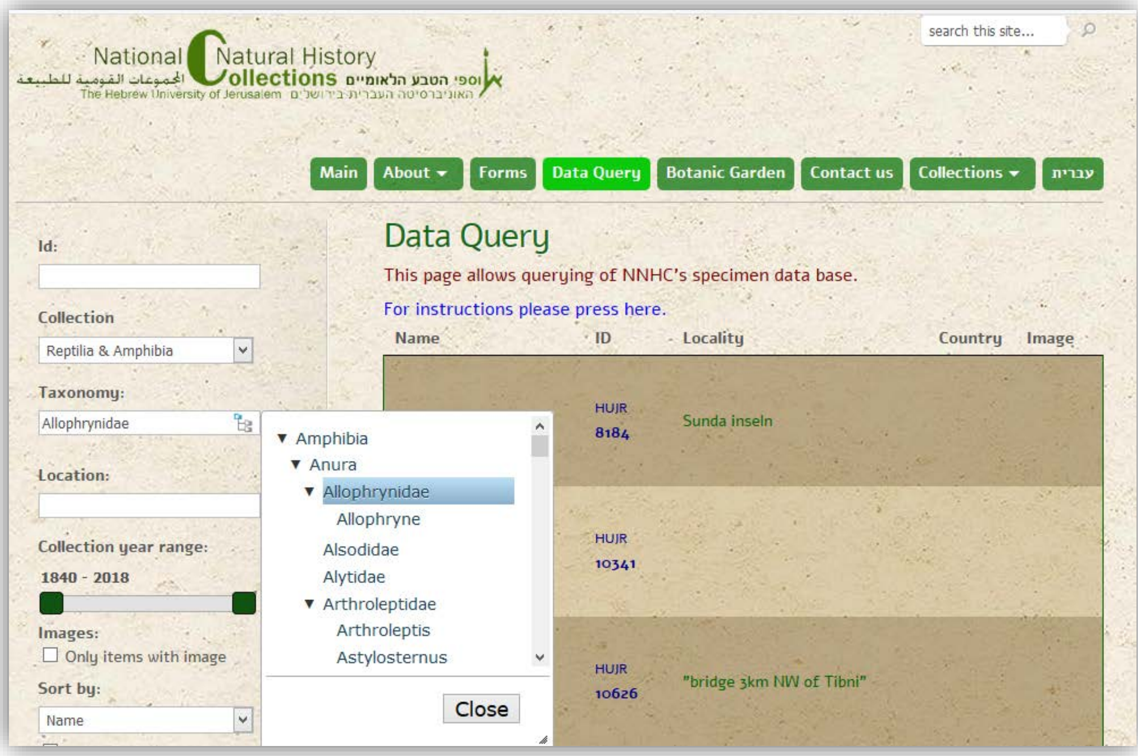


Figure 52: “Aharoni” online digitized collection datasets with their extended taxonomy

The Banja Luka datasets (Figure 53) represent data from 5 collections/museums that cover different fields: archaeology, ethnology, art history, natural history, all described through not standardized metadata fields.

For this specific case study we have taken into consideration the “raw data” of the collections in order to have their original descriptions. In fact the data have been transformed into an extended Dublin Core for the aim of adapting them to the DSpace structure of VI-SEEM. Such transformation flattened the information relative to the datasets.

ARCHAEOLOGY	ETHNOLOGY	HISTORY	NATURE HISTORY	HISTORY OF ART
NAME AND LOCATION OF THE MUSEUM	Inventory number	Inventory number	NAME AND LOCATION OF THE MUSEUM	NAME AND LOCATION OF THE MUSEUM
Number of main documentation	Number of main documentation	Number of documentatio n	International museum code	Inventory number
Collection	Number of pieces	Object	sex, age, development stage: number of specimens or fragments	Number of main documentation
Object	Name of the object	Collection	Object name, code	Name of the artist
Placement	Folk term	Negative number and mark	Collected by: name, date	Negative number and label
Negative mark and number	Collection	Photograph	Identified by: name,	Title

Figure 53: Banja Luka datasets

The assessment of existing semantic solutions that take into consideration the presence of cross-disciplinary data was necessary. The cross-disciplinarity of the data needs of flexible data models and tools for a proper and efficient management, integration and interoperability with other similar information coming from other research fields. Ontologies could provide data with the necessary structure and semantic richness required to fulfill these goals.

On the base of these case studies, we opted for a stable ontological solution already tested in different fields (therefore serving the need of our cross-disciplinarity) and that it is possible to further develop in the future, according to the needs of Vi-SEEM or of any cross-disciplinary e-Infrastructure.

It was decided to choose the CIDOC CRM ontology since it is a stable and also flexible ontological solution already tested in different research fields. CIDOC CRM is a core ontology designed to become a platform-independent and machine-understandable language, suitable for data description in a formal language. CIDOC CRM also comes with a set of extensions intended to satisfy the needs of specific disciplines, such as cultural heritage and scientific data.

This ontology has different extensions that can be used for the description of data coming from different research fields. Moreover, the structure both of the core ontology and of the extensions allows us to further extend or developing the ontology.

The alignment of the case studies structures into CIDOC CRM and its extensions is showed in the following schemas. Figure 54 shows the mapping of the “Aharoni” Online Digitized collection datasets to CIDOC CRM.

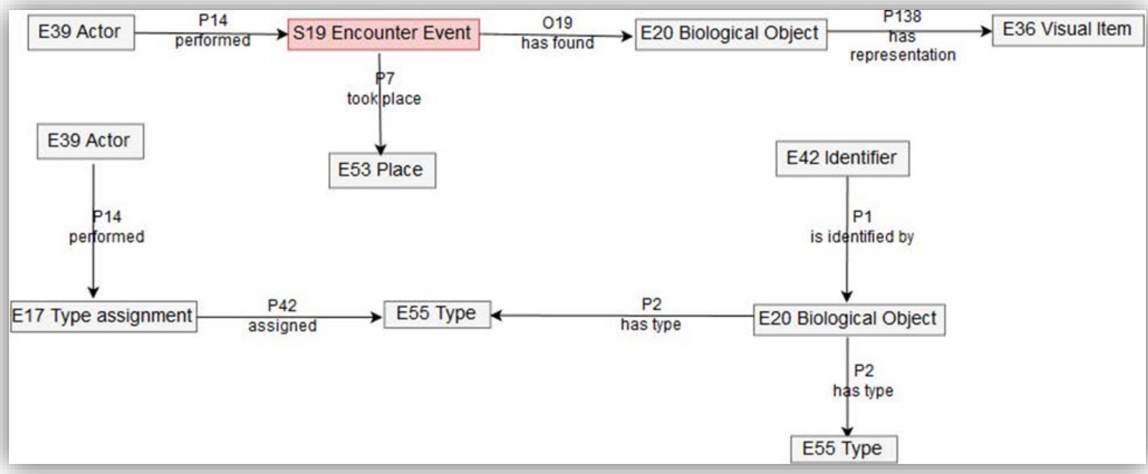


Figure 54: Mapping of the “Aharoni” online digitized collection datasets to CIDOC CRM

Figure 55 shows the mapping of one of the collection from Banja Luka, (specifically the archaeological museum) to CIDOC CRM.

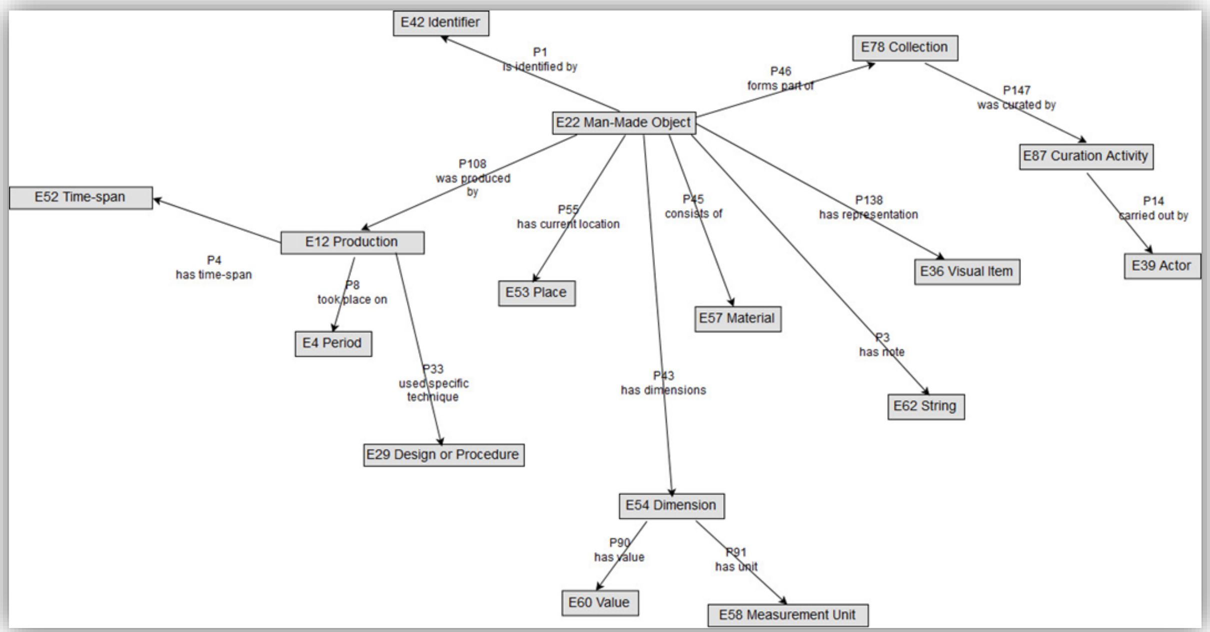


Figure 55: Mapping of the Banja Luka dataset to CIDOC CRM

In this case we can see also the different alternatives that can be used to describe the data and the proof of the flexibility of this solution. Figure 56 shows the alternative in the description of parts of the dataset.

Moreover, this solution allows us to describe and make coherent the description of different data, without flattening them.

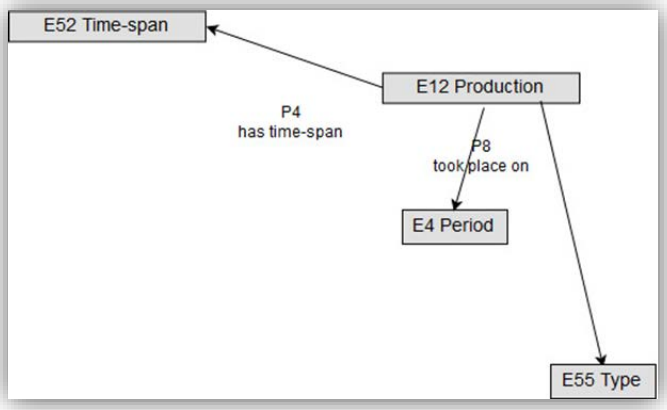


Figure 56: Alternative mapping of the Banja Luka dataset to CIDOC CRM

As a further step, the solution drawn on the case studies chosen can be integrated also with other ontological solution such as the Parthenos ontology. In fact, this solution is devoted to the ontological description of e-infrastructures and it is based on CIDOC CRM. This allows us to be in line with other research in the field in order to collaborate in the future towards a common integrated solution.

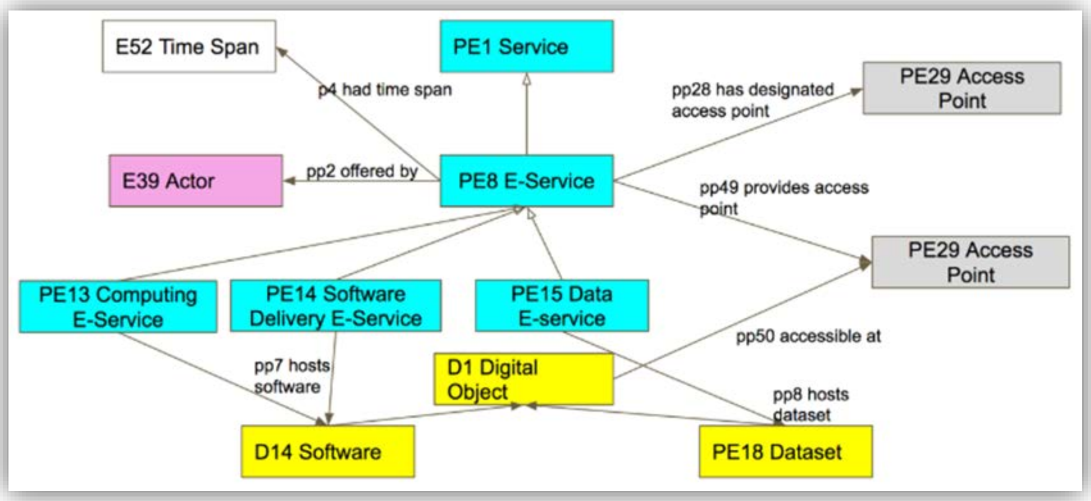


Figure 57: Parthenos ontology (devoted to the ontological description of e-infrastructures).

A scientific publication about the solution adopted towards the development and use of this ontological solution is under submission.

9 Conclusions

The deliverable D5.5 is the main output of the Tasks T5.3, T5.4 and T5.5. It is reporting in detail the final version of Virtual Research Environment, making use of the outcome of the focused effort invested in deploying the VRE platform, the overall integration of services as well as the scientific support. Hence, it describes the design behind the Virtual Research Environment platform, how this reflects the integration of the generic and domain specific VI-SEEM services, and what the content of the portal is.

More specifically, the deliverable describes the different sections of the VRE platform, illustrating how each section manages to integrate the VI-SEEM offered services. Details on how the user can easily get information and use the VI-SEEM provided services are given, together with the current services integration status. Detailed instructions on how the user can integrate hers/his own applications and services to the VI-SEEM Virtual Research Environment are given in each section. A description of possible future additions in the VRE is given and the cross-disciplinary actions of the project are presented in detail.

This deliverable can be considered to be an update of the deliverable 5.4. Considerable changes have been made due to comments and requests received by the reviewers during the midterm review which took place on May 2017 as well as the unofficial check which was held on May 2018. The suggested changes enhanced the usability and user-friendliness of the Virtual Research Environment. This is reflected in the user reports received by randomly chosen VI-SEEM users.

