Sharing value-added services for Research and Education communities

Dr. Ognjen Prnjat
European and Regional e-Infrastructure manager
GRNET

The VI-SEEM project initiative is co-funded by the European Commission under the H2020 Research Infrastructures contract no. **675121**

VRE for regional interdisciplinary communities in SEE and EM

Vi-SEEM

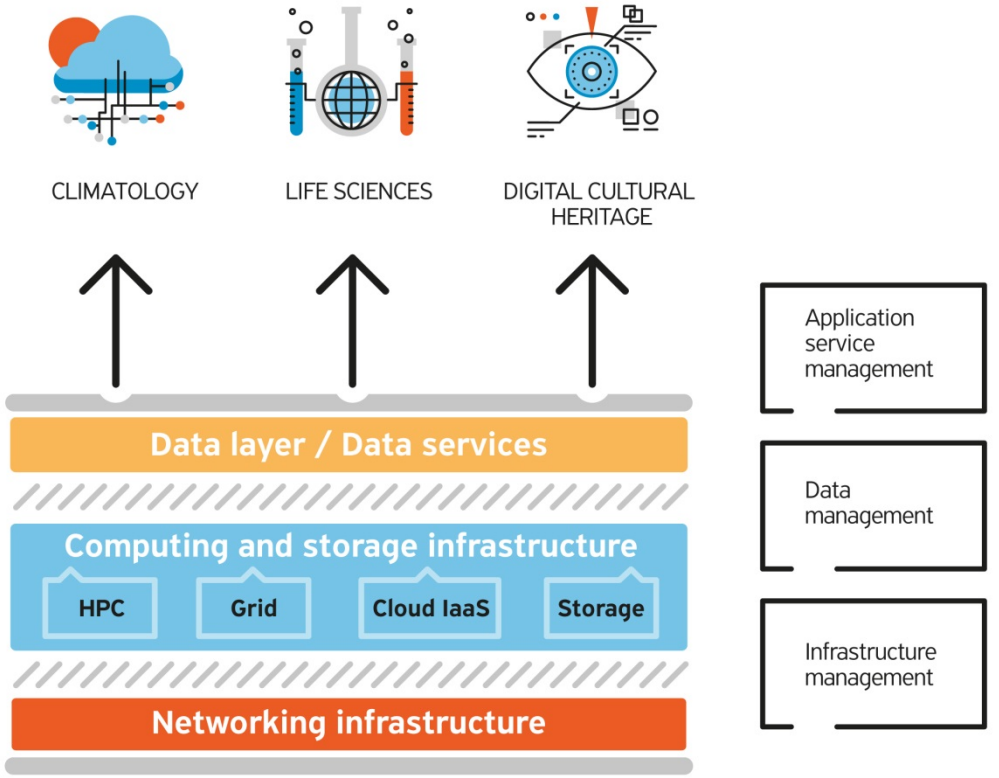
- e-Infrastructure built over last decade
- Merging of SEE and EM regions
- SEE: network SEEREN1-2, Grid SEE-GRID-1/2/SCI, HPC HP-SEE
- EM: HPC LinkSCEEM1-2

Participant no.	Participant organisation name	Part. short name	Country
1 (Coord)	GREEK RESEARCH AND TECHNOLOGY NETWORK S.A.	GRNET	Greece
2	THE CYPRUS INSTITUTE	CyI	Cyprus
3	INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES – BULGARIAN ACADEMY OF SCIENCES	IICT-BAS	Bulgaria
4	INSTITUTE OF PHYSICS BELGRADE	IPB	Serbia
5	NATIONAL INFORMATION INFRASTRUCTURE DEVELOPMENT INSTITUTE	NIIF	Hungary
6	WEST UNIVERSITY OF TIMISOARA	UVT	Romania
7	POLYTECHNIC UNIVERSITY OF TIRANA	UPT	Albania
8	UNIVERSITY OF BANJA LUKA	UNI BL	Bosnia and Herzegovina
9	SS CYRIL AND METHODIUS UNIVERSITY OF SKOPJE	UKIM	FYR of Macedonia
10	UNIVERSITY OF MONTENEGRO	UOM	Montenegro
11	RESEARCH AND EDUCATIONAL NETWORKING ASSOCIATION OF MOLDOVA	RENAM	Moldova
12	INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF ARMENIA	IIAP-NAS-RA	Armenia
13	GEORGIAN RESEARCH AND EDUCATIONAL NETWORKING ASSOCIATION	GRENA	Georgia
14	BIBLIOTHECA ALEXANDRINA	BA	Egypt
15	INTER UNIVERSITY COMPUTATION CENTER	IUCC	Israel
16	SYNCHROTRON-LIGHT FOR EXPERIMENTAL SCIENCE AND APPLICATIONS IN THE MIDDLE EAST	SESAME	Jordan

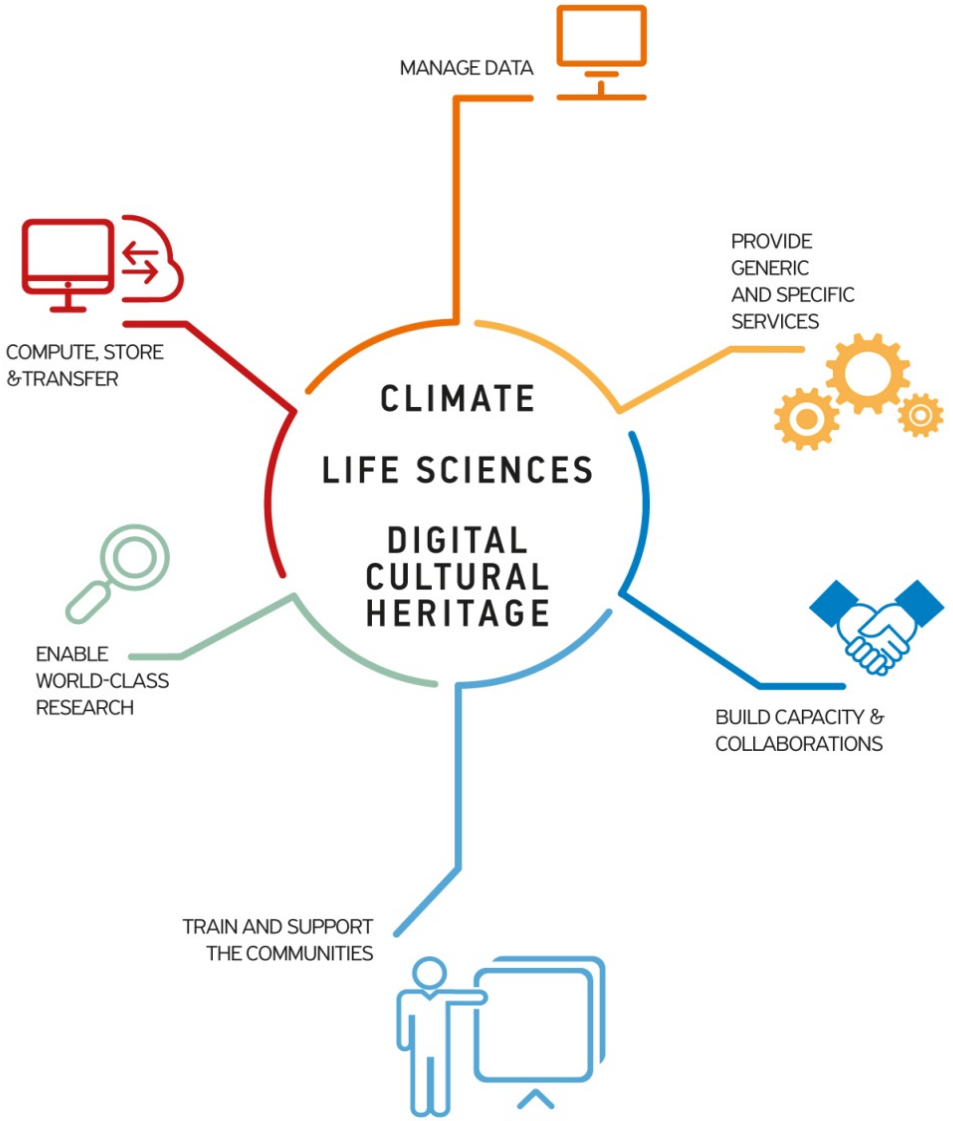
Technologies for open science

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

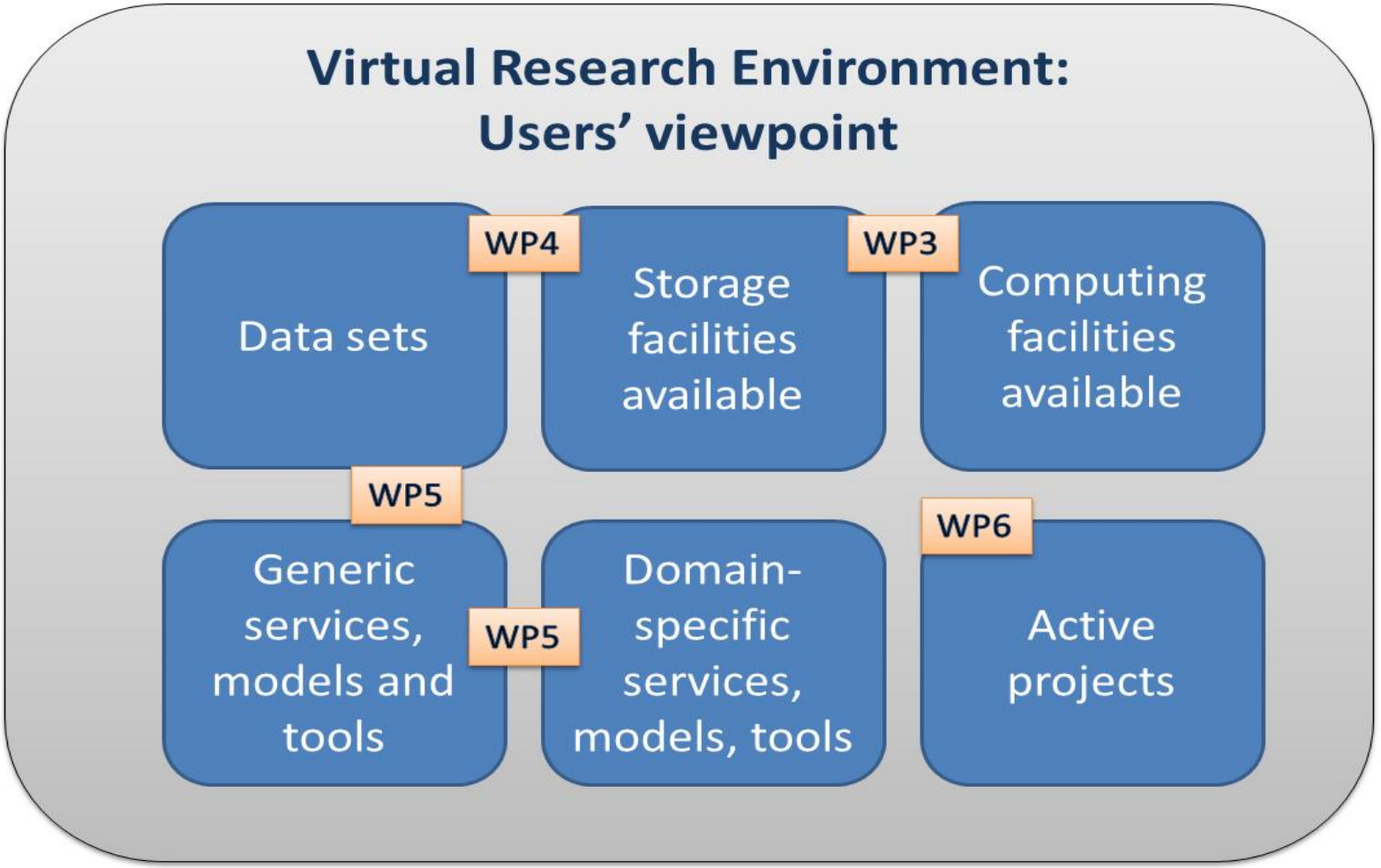
- Diverse computing technologies
- Advent of big data
- Service orientation



Detailed objectives



User-centric approach



Sharing value added services: in numbers

- e-Infrastructure: 21500 CPU cores, 325000 GP-GPU cores and 18500 Intel Xeon Phi cores of HPC, 2900 grid cores, 10500 cloud VM cores, 11 PBytes of storage (of which dedicated 5-15% , 10-15%, 5% and 10% respectively)
- > 10 specific services, 30 data sets and 25 codes
- 39 applications, 45 research teams taking part. 100 publications, 50 presentations
- 17 dissemination events,12 training events

Sharing value added services: the service catalogue

1. e-Infrastructure services: HPC, Grid, Cloud, Code repository

2. Data management services

3. Generic services, models and tools

4. Domain specific services, models and tools

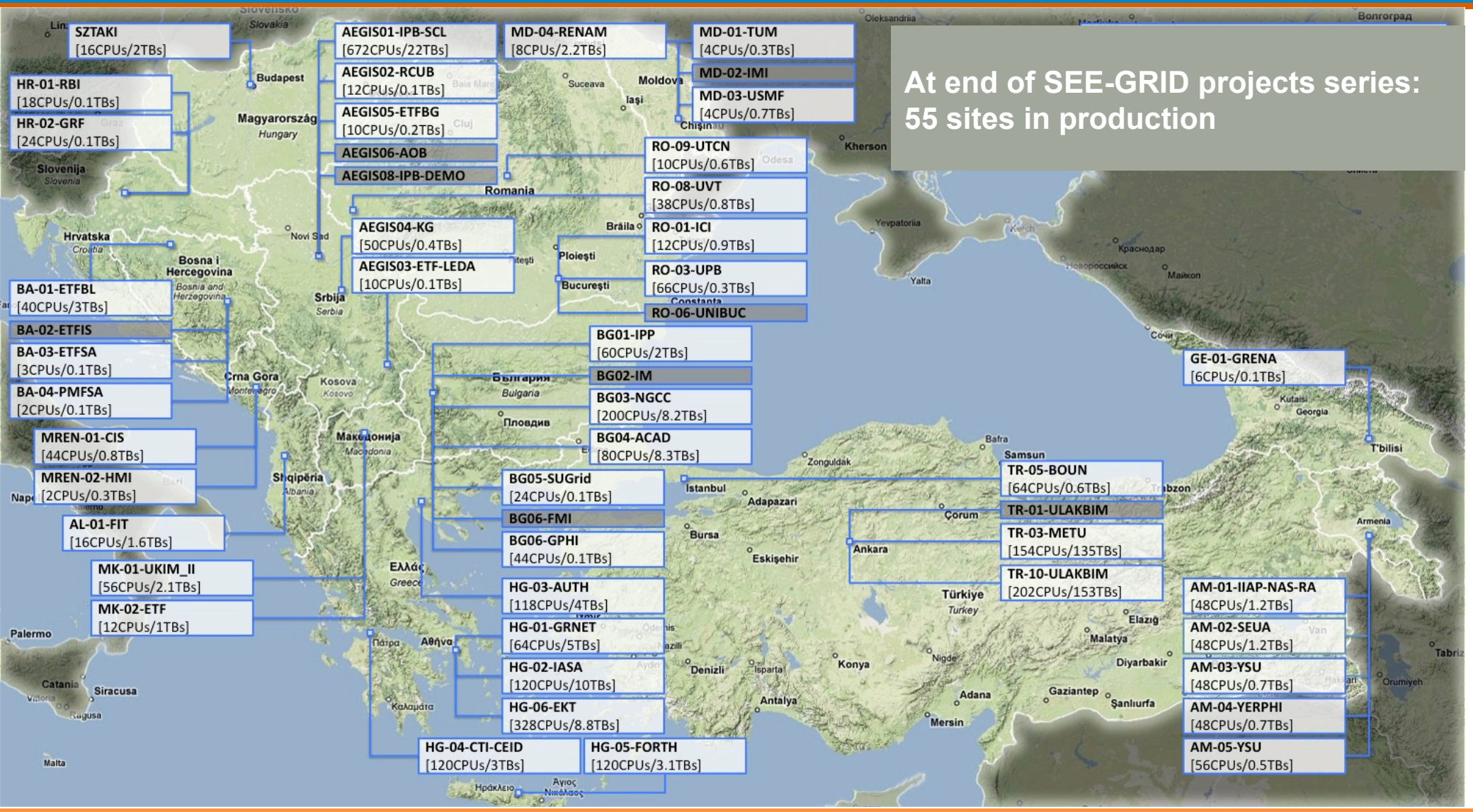
5. Scientific applications and data sets

e-Infrastructure services

Vi-SEEM

Resource Type	Number of countries contributing	Number of sites	Total amount of resources	Total amount of resources dedicated for the project
HPC x86 Compute	8	11	~21500 cores	~17.5 Million core hours per year
HPC GPU	3	3	~325000 GP-GPU cores	~ 18000 GPU cores
HPC Intel Phi Accelerators	1	1	~18500 co-processor cores	~8M co processor core hours per year
Grid Compute	12	25	~ 2900 cores	10-15% of the resources
Cloud Compute	8	9	~ 10500 VM cores	5% of VM cores
HPC Test/Train Cluster	3	4	~300 cores	~50000 cores hours per year
Storage (Disk)	13	18	3.1 PB	515 TB
Storage (Tape)	2	2	8 PB	550 TB

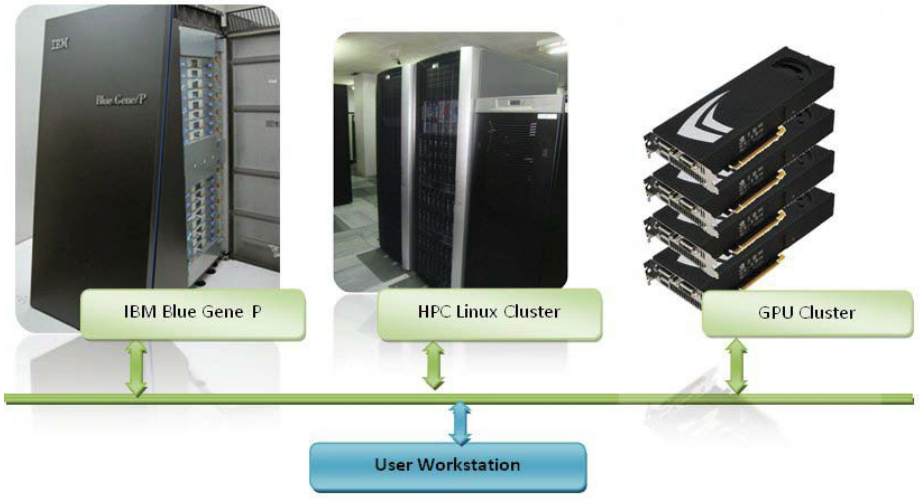
e-Infrastructure services: Grid



At end of SEE-GRID projects series: 55 sites in production

e-Infrastructure services: HPC

- 3 top 500 supercomputers in June 2015 (Bulgaria, Greece, Hungary)
- Heterogeneous infrastructure - Blue Gene/P SCs, large HPC clusters with advanced interconnects. Advantage to users
- To support a substantial number of libraries, toolkits, codes and application software



Data management services

- Functions allowing for data management for selected Scientific Communities, engage the full data management lifecycle:
- Simple data sharing
- Data repositories
- Long-term data archiving
- Working/scratch area
- Data search and catalogue
- Data analytics

Generic and Domain-specific services, models and tools

Vi-SEEM

Generic services, models and tools

- Application environment (optimized applications, codes, libraries), preinstalled in compute resources or provided as VM images. (Climate models, production programs for Life Sciences applications, data management tools for Digital Culture Heritage).
- Workflow, pipeline and software tools repository
- Regional community datasets

Domain specific services, models and tools

- ChemBioServer: a web-based pipeline for filtering, clustering and visualization of chemical compounds used in drug discovery,
- Live Access Server: a web server providing flexible access to geo-referenced scientific data, offering visualization & post-processing capabilities for climate data,
- Clowder: a Digital Culture Heritage repository which also offers integrated interactive visualization tools.

Scientific applications

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

Data sets: climate

Simulation and observation data; global and local sets

Source	Description	URL
Earth System Grid Federation	Data portal for downloading CORDEX and other data (i.e. CMIP5)	http://pcmdi9.llnl.gov/esgf-web-fe/
Climatic Research Unit (CRU)	gridded observations of surface meteorological data and climatic indices	http://www.cru.uea.ac.uk/data
E-OBS	gridded observations of surface meteorological parameters for Europe	http://eca.knmi.nl/download/ensembles/download.php
ECA&D	meteorological station observations for Europe and part of MENA	http://eca.knmi.nl/dailydata/index.php
APHRODITE	Gridded precipitation data for Asia and Middle East	http://www.chikyu.ac.jp/precip/index.html
Global Precipitation Climatology Centre (GPCC)	Gridded global precipitation observations	ftp://ftp-anon.dwd.de/pub/data/gpcc/html/gpcc_normals_download.html
European Centre of Medium-Range Weather Forecasts Data Portal (ECMWF)	Mainly for downloading reanalysis products such as the ERA-Interim	http://apps.ecmwf.int/datasets/
AERONET	federation of ground-based remote sensing aerosol networks	http://aeronet.gsfc.nasa.gov
MODIS	MODIS (or Moderate Resolution Imaging Spectroradiometer) is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites	http://modis.gsfc.nasa.gov
ISA-MEIDA	collection and archiving of Israeli Earth Science data sets	http://nasa.proj.ac.il/
TOMS aerosol index	index that detects the presence of uv-absorbing aerosols such as dust and soot	http://disc.sci.gsfc.nasa.gov/data-holdings/PIP/aerosol_index.shtml

Source	Description
TAU, Israel	Meteorological and hydrological data for Israel, environmental data, such as PM10 and PM2.5 measurements in various locations in Israel
METEO-RO, Romania	a gridded dataset from observations consisting of 9 essential climate variables with daily values (1961-2013) on a spatial resolution of 10 km for the Romanian territory (ROCADA)
IEG, Moldova	data recorded by meteorological stations and posts of the State Hydro-meteorological Service
NIGGG, Bulgaria	NCEP Global Analysis Data with 1º resolution, available since 2000 TNO high resolution emission inventory for Europe Data of the pollution levels, measured by the Bulgarian National Network for Air Quality Control, available since 2000
IGEWE, Albania	data collected from the Hydrological, Meteorological and Seismological networks (such as water flow, temperature, humidity, precipitation etc., and a range of seismic data)

Data sets: life sciences

Vi-SEEM

Simulation and observation data; global and local sets

Country	Available data	Access Policy (Private, Regional, Public)
Greece	Virtual screening results	Data that is published is public and can be shared
Serbia	SNP database of variations in genes involved in Cystic Fibrosis, Thrombosis, Alpha-1-Antitripsine deficiency, Colorectal cancer, Thyroid cancer, Lung cancer, Endometrium. The database contains data on Serbian population.	Data that is published is public and can be shared
Moldova	Investigation Analytics, Research Results, Pathologies case studies	Investigation Analytics and Research Results are shared information. Pathologies case studies (no personal patient info) – available for research community by creating agreement with institution.

Data sets: cultural heritage

Source	Description	URL
Ancient Cypriot Literature Digital Corpus (AKGDC)	Searchable digital library of Ancient Cypriot Literature. Covers the ancient Cypriot literary production from 7th century BC to 5th-6th century AD and examines it through its wide range of literary genres.	Not public yet.
STARC repository	Repository for digital CH objects created over the last 5 years at CyI. Including novel interactive methods for accessing and exploring information.	http://public.cyi.ac.cy/starcRepo/
Electronic Corpus of Karamanlidika	Digitalization and transcription into Latin characters of a corpus of Karamanlidika printed material from different periods and genres with access to graphic, phonetic and morphemic varieties search.	Not public yet.
The digital collection of Bibliotheca Alexandrina	As a library of the 21st Century, the Bibliotheca Alexandrina (BA), along with its affiliated academic and cultural centres, is committed to digitization as a means of preserving, managing, and disseminating information and knowledge. The BA sets out to share them with a worldwide audience via the internet, thus promoting greater understanding and tolerance between cultures. To achieve this goal, the BA has formed partnerships with various cultural, academic, governmental, and corporate organizations for the creation of many significant and compelling digital projects. These projects cover a wide array of cultural and educational themes.	http://www.bibalex.org/Libraries/Presentation/Static/12600.aspx?d=0

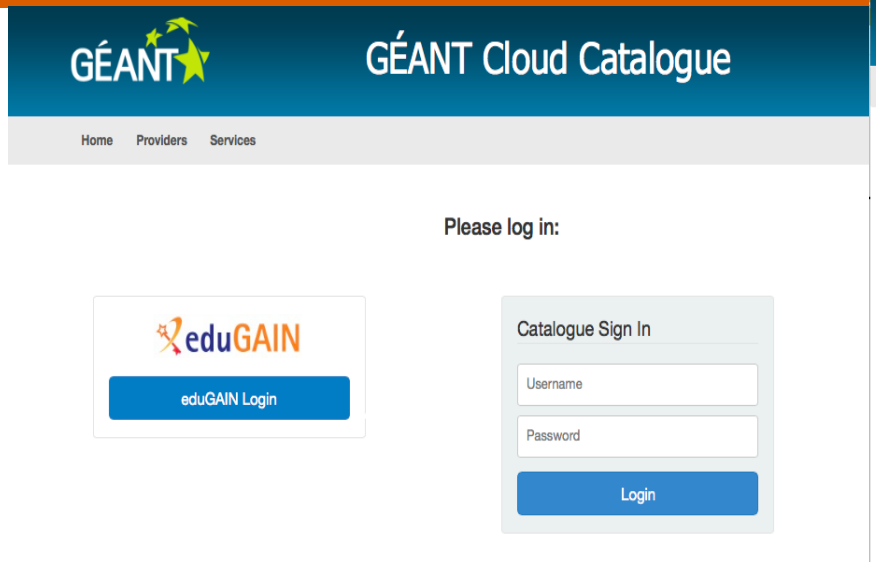
Access to VRE

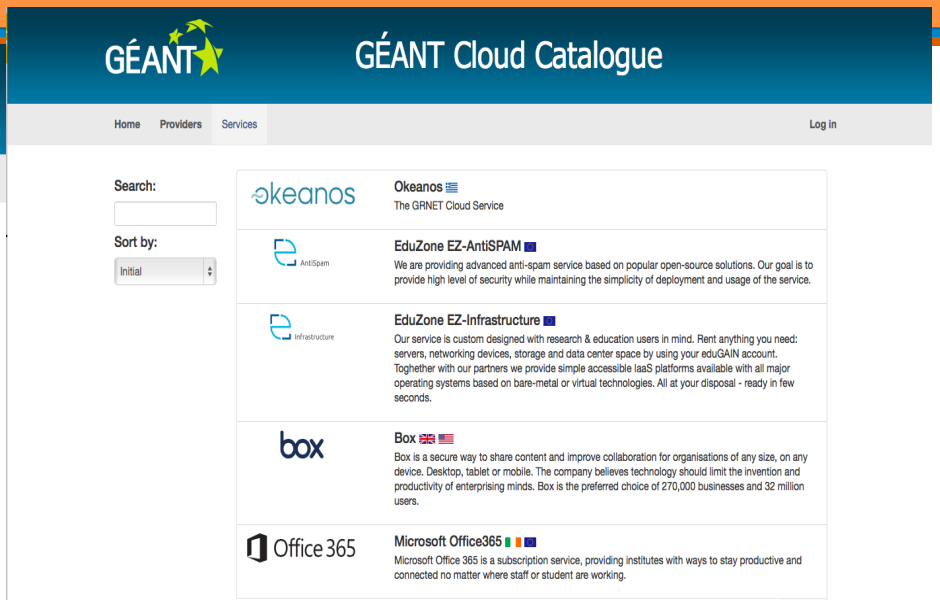
- Through 3 distinct calls
- Pre-selected candidates M12
- Open calls M18 and M28
- Access to scientist outside the target region, including Africa, will be considered

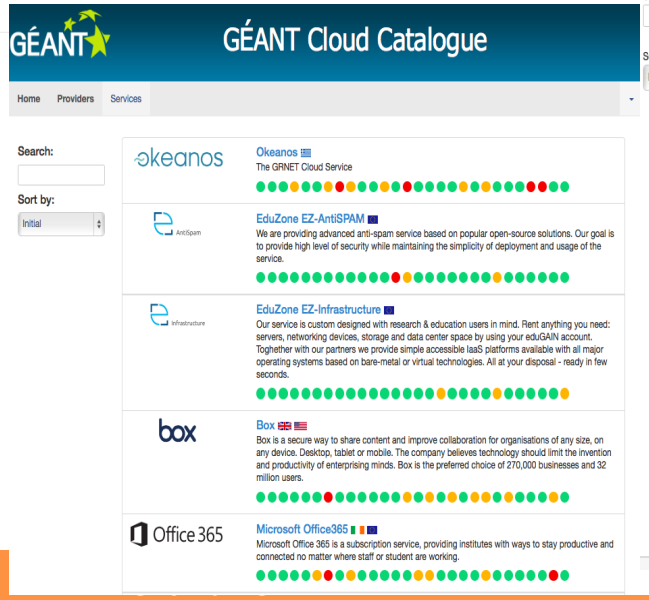
MAGIC project: service catalogue

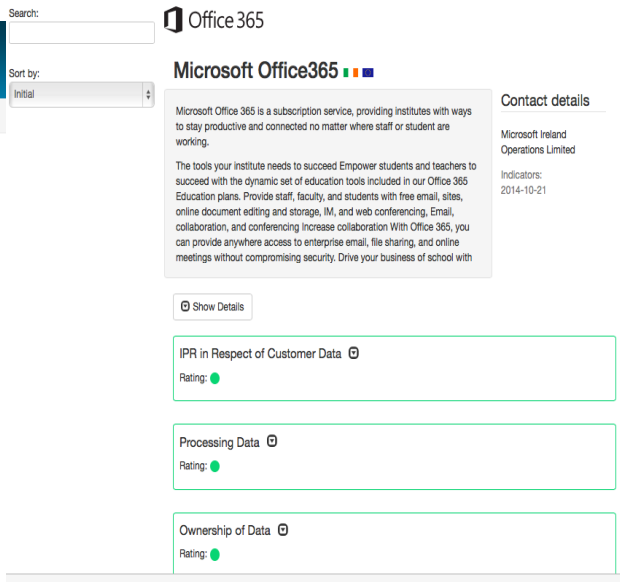
- Middleware for collaborative Applications and Global Virtual Communities
- Will provide application/service sharing platform via a catalogue
- In implementation now
- Current services: colaboratorio, file sender, docuwiki
- Based on GEANT cloud catalogue

MAGIC project: service catalogue









Conclusion

- A case study for open science – sharing value-added services for R&E communities
- VI-SEEM provides a Virtual Research Environment to be provided to the scientific user communities in Climatology, Life Sciences, and Cultural Heritage
- VI-SEEM unifies networking, computing and data management
- Support the full lifecycle of scientific research
- Services will be provided through a service catalogue

- Similarly MAGIC project will establish a global catalogue of applications/ services
- WA can contribute its open services to the global pool, and find services of their interest

Thanks!