

VRCs on EGI and Regional Infrastructures

F. Ruggieri¹

INFN Sezione di Roma Tre

Via della Vasca Navale 84, Roma, Italy

E-mail: Federico.Ruggieri@roma3.infn.it

A. Peresan¹, F. Vaccari, A. Magrin, F. Romanelli, G.F. Panza

Department of Geosciences of the University of Trieste and

The Abdus Salam International Centre for Theoretical Physics, ICTP SAND Group

and S. Cozzini

Consiglio Nazionale delle Ricerche, CNR/IOM uos Democritos

Trieste, Italy

E-mail: aperesan@units.it

Trieste – Italy

A. Fresa¹

INDICATE project

c/o Promoter

Via Boccioni, 2, Peccioli (Pisa), Italy

E-mail: fresa@promoter.it

G. Sipos¹

EGI.eu

Science Park 140, Amsterdam, Netherlands

E-mail: gergely.sipos@egi.eu

D. Scardaci¹

INFN Sezione di Catania

Via S. Sofia, Catania, Italy

E-mail: diego.scardaci@ct.infn.it

M. Reale¹

GARR

Via dei Tizii 6, Roma, Italy

E-mail: mario.reale@garr.it

A. Krenek¹, L. Matyska

CESNET

Zikova 4, 162 00 Prague, Czech Republic

¹ Speaker

E-mail: {ljocha,ludek}@ics.muni.cz

The CHAIN project has organised a workshop at the EGI Community Forum 2012 event in March 2012. The event aimed at reviewing the current status of Virtual Research Communities (VRCs) and their needs in terms of e-Infrastructures. The workshop has been focusing mainly on two topics:

1. An update on the Digital Cultural Heritage and Seismic VRCs status and requirements – including the current VRC vision of the CHAIN, EGI-InSPIRE, EUMEDGRID-Support, INDICATE and GISELA projects.
2. The CHAIN roadmap towards interoperability among e-Infrastructures.

The workshop included presentations about CHAIN, its user communities, its partner projects and of the expected structure and content of the CHAIN roadmap that will be delivered by the end of 2012. Within this roadmap Science Gateways are proposed as sustainable interoperability solution for applications.

*EGI Community Forum 2012 / EMI Second Technical Conference,
Munich, Germany
26-30 March, 2012*

Introduction

The CHAIN project, in collaboration with the GISELA project, has organised a workshop during the EGI Community Forum 2012 event. The workshop focused mainly on two topics:

1. The current Virtual Research Communities (VRC) vision of the CHAIN, EGI-InSPIRE, EUMEDGRID-Support, INDICATE and GISELA projects and an update of the Digital Cultural Heritage (DCH) and Seismic VRCs status and requirements

and

2. The expected structure and content of the CHAIN roadmap.

Two presentations in the workshop were introducing the CHAIN Project and the current status of its activities, while other three presentations were dedicated to the activities performed by the EGI-InSPIRE, EUMEDGRID-Support, INDICATE and GISELA projects to support VRCs.

The workshop has also involved new VRCs (DCH and Seismic) to investigate their requirements, in terms of resources, their actual and potential geographical coverage and understand the possibility for them to exploit the existing regional e-Infrastructures.

Virtual Research Communities (VRCs) could span a large organisational spectrum: from a group of researchers working together, sharing means to carry out common investigations, such as distributed resources: instruments, computing, storage and software tools, up to large international well-structured collaborations with or without large dedicated resources. The Grid infrastructures have been promoted, as it is well known, largely by the High Energy Physics community and their success in the analysis of the data coming from the Large Hadron Collider (LHC) at CERN is undisputable. However these infrastructures have demonstrated to be suited as well for several other scientific domains: from Biomedical to Weather Forecast research. Expanding the support to other VRCs is thus strategic not only towards sustainability, but also to evaluate new requirements and possibly evolve the services to fulfil them.

CHAIN, EGI-InSPIRE, EUMEDGRID-Support and GISELA present here a status report on their main activities to support the VRCs. The contributions show many similarities, as expected, but also some significant differences due to the different constraints in the various regions of the World.

Digital Cultural Heritage and Seismic communities presented their peculiar requirements which are different both in terms of resources and level of maturity.

A true intercontinental infrastructure, based on the existing regional deployments, should provide services and organisational structures that allow VRCs to work seamlessly across the continents.

The following chapters describe the contributions presented at the workshop and in the final chapter conclusions, coming out from the discussion, are drawn.

The CHAIN project

The CHAIN project (Grant Agreement n. 26011) aims to coordinate and leverage the efforts made over the past years to extend the European e-Infrastructure (and particularly Grid) operational and organisational principles to a number of regions in the world. CHAIN uses their results with a vision of a harmonised and optimised interaction model for e-Infrastructure and specifically Grid interfaces between Europe and the rest of the world. The project is elaborating a strategy and defining the instruments in order to ensure coordination and interoperability of the European Grid Infrastructure with emerging e-Infrastructures in other regions of the world (Asia, Mediterranean, Latin America and Sub-Saharan Africa).

The CHAIN consortium consists of leading organisations in all the regions addressed by the project and ensures global coverage, and most efficient leveraging of results with respect to preceding regional initiatives.

The project started on the 1st of December 2010 and has so far achieved important results described synthetically in the following paragraphs.

A questionnaire has been prepared to collect up to date information about the state of the art of e-Infrastructures in all relevant regions, and the information obtained has been integrated with other data in a Knowledge Base that has been made available on the project web site with an attractive graphical interface (Figure 1). The information is further used for project analysis and design activities.

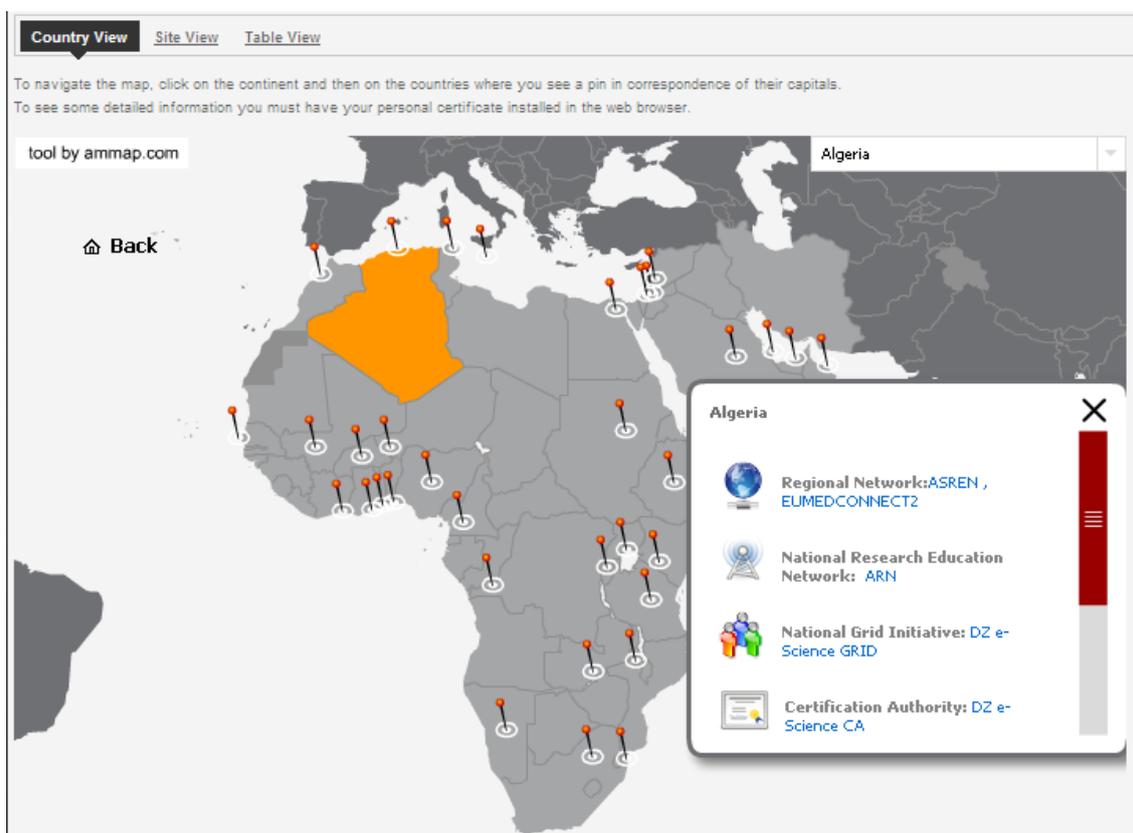


Figure 1 - Country view of the CHAIN Knowledge-Base (<http://www.chain-project.eu/knowledge-base>)

The CHAIN project has assessed the current state of the art of Grid infrastructures in the addressed regions and has produced a set of guidelines that will foster the continuity of e-Infrastructures [1]. The activity promoted the emergence of agreed solutions for interoperation and interoperability across regions and middleware. The project has produced a comprehensive set of recommendations regarding technologies and taking into account organisational and regional specificities [1]. It has as well promoted the uptake of the Science Gateway paradigm adopted by other projects such as: DECIDE², EMEDGRID-Support³, GISELA[16], INDICATE [5] and the CHAIN Application Database (<http://www.chain-project.eu/applications>) has evolved into a “gateway” to the existing Science Gateways.

At the same time, the consortium is studying a coherent scheme of cooperation and interoperation of EGI.eu with external e-Infrastructures taking into consideration the specificities of the different regions. This will provide the assessment of the existing services, and will suggest the necessary services to support interoperation. The most important preliminary outcome of this study was the creation of Africa & Arabia Regional Operation Centre (<http://roc.africa-grid.org>) and of China Regional Operation Centre (<http://www.china-roc.cn>). These two ROCs have been created to support their respective regions/countries deploying new customised instances of the EGI operational tools.

CHAIN is also leveraging on the cross-region communities and applications and promotes inter-regional, e-Infrastructure based, research collaborations. A limited number of reference communities have been chosen but the activity also aims to promote the continuity of support to the large spectrum of other communities, mainly but not only, to those that are already supported by the existing regional projects.

The project is performing a wide-spectrum investigation on the existing and potential trans-continental communities, scientific and technical collaborations, collecting requirements, commonalities, challenges and possible synergies and producing a study of the necessary steps to be made by e-Infrastructures to fulfil the requirements of present and emerging virtual research communities with trans-continental span. As a result of this activity, CHAIN is fostering the use of the Science Gateway (SG) paradigm by those identified VRCs that do not count on a web engine for the submission of jobs. The SG has been developed by the INFN Catania Group and provides two main features that are relevant for the CHAIN and VRCs activities:

- a) an authentication based on Federations of Identity Providers (IdP);
- b) a Grid Engine based on JSAGA standard to submit Grid Jobs to different types of middlewares (MW) (e.g. EMI/gLite/Unicore, GOS, GARUDA, OurGrid, etc.).

A Worldwide Interoperability Demo of a set of applications running on different MW by using the SG is foreseen for the next EGI TF 2012 in Prague.

This way, the aforementioned synergies and commonalities are expected to be easily achieved.

² www.eu-decide.eu

³ www.eumedgrid-suport.eu

Initial Virtual Research Communities (VRCs) are WeNMR (BioMedical), WRF4G (Weather and Earth Science), jModelTest, DC-NET/INDICATE/DCH-RP (Digital Cultural Heritage) DECIDE (Health) and LSGC (Life Sciences), while Climate Change and SuperB (High energy Physics) are under way and other domains have been considered such as Agriculture.

Digital Cultural Heritage

The European amount of digitized material is growing very rapidly, as national, regional and European programmes support the digitization processes by Museums, Libraries, Archives, Archaeological sites and Audiovisual repositories. The generation of digital cultural heritage is accelerated also by the impulse of Europeana [2] that is fostering the European cultural institutions to produce even more digital content. Moreover digital cultural heritage content are complex and interlinked through many relations. European countries are working for the future, in order to create a data infrastructure devoted to cultural heritage research. Currently, Europe have 2 twin-projects (DC-NET and INDICATE) ongoing and a new international coordination action (DCH-RP) is starting in October 2012 to design a validated Roadmap for the preservation of digital cultural content. These initiatives are contributing to smooth the way to the Open Science Infrastructure for Digital Cultural Heritage (DCH) which is foreseen in 2020.

The projects are funded by EC FP7 e-Infrastructures.

DC-NET [3] – Digital Cultural Heritage Network for the European Research Area just completed in March 2012 and produced a Joint Activities Plan to implement the priority services identified by the participating Member States.

The final conference of this project took place in Rome on 8th March and it was the successful conclusion of a very important path across EU Member States to establish an operative dialogue between cultural heritage and e-Infrastructures communities in Europe.

Programme Owners and Programme Managers in the cultural sector worked to agree common perspectives and priorities, and to identify constraints and capabilities in order to establish a plan of joint activities for the implementation and deployment of a new data and service e-Infrastructure for the DCH research.

Coordination and integration among the European Countries are at the core of the DC-NET approach: to contribute to the Joint Programming Initiative on cultural heritage [4], to integrate the research capabilities of the participant Member States, to identify standards, needs and solutions and to valorize the existing projects and results.

INDICATE [5] – International Network for a Digital Cultural Heritage e-Infrastructure provides a concrete approach within an international dimension to the priorities established by DC-NET.

INDICATE is intended for stimulating the international cooperation of e-Infrastructures providers and cultural heritage users in the frame of concrete use-case studies and pilots carried out in the countries of the Mediterranean region and internationally.

The network set up with this project is indeed a community of partners who share common interests and aim to long-term collaboration also beyond the lifetime of the project.

Within this network of common interest, the following activities take place:

- establishing working groups dedicated to specific aspects of e-Infrastructures-enabled cultural heritage which are relevant across national and organizational boundaries;
- holding conferences, workshops and seminars on the convergence of e-Infrastructures and Arts, Humanities and Social Sciences with special focus on digital cultural heritage and the cooperation between European and Mediterranean countries;
- developing on-line tools and resources for the working groups such as use case studies, pilots and other services.

Within INDICATE, the partners explore the potential of e-Infrastructures to overcome key problems which are rising in the e-Culture such as long-term preservation, virtual exhibitions, management of geo-coded cultural content.

The eCulture Science Gateway eCSG [6] is one important result of the project. It is a web portal that provides easy-to-use interface and a single sign-on services. The eCSG is based on Liferay portal framework [7] for which a portlet has been developed to interface gLibrary [8], the digital assets manager for the Grid developed by INFN and COMETA. The Authentication and Authorization supports Federations of Identity Providers enabled by Shibboleth [9]. eCSG is an important progress towards an easier and wider use of the e-Infrastructures by non-expert users and by the DCH sector in particular.

The project is not only Europe-centred, but it has instead a genuine international vocation: Egypt, Turkey and Jordan are partner of the project; cooperation are ongoing in China in liaison with the EPIKH Grid School, in Taiwan in the frame of the TELDAP Conference and in South America in the frame of experiments for live distributed performances in liaison with the CHAIN and GISELA projects.

The INDICATE project started in September 2010 and will last until October 2012.

The DC-NET and INDICATE projects have many partners in common, and are targeted to the same DCH community. The e-Infrastructure programs identified in DC-NET will be at the basis of the sustainability of the results of INDICATE.

INDICATE and DC-NET are part of a wider process, which started 10 years ago among cultural institutions and this process entered in a new phase, by joining the research e-infrastructures.

Time is now ready to start working towards an Open Science Infrastructure for DCH in 2020, and concretely to establish a roadmap for DCH that express clear workpaths, alliances and strategies.

The use of existing research e-infrastructure is an efficient channel for the delivery of advanced services to the digital cultural heritage. It will allow to distribute advanced services for digital cultural heritage research and contents.

To achieve this goal, NREN and NGI facilities needs to be complemented with targeted specific services. DC-NET and INDICATE started the work.

The new initiative DCH-RP Digital Cultural Heritage Roadmap for Preservation, expected to start in October 2012 and lasting 2 years, will continue adding more concrete results in the specific area of the digital preservation.

The main objectives of DCH-RP are:

- i. to harmonize data storage and preservation policies in the DCH sector at European and international level,
- ii. to progress with the dialogue among DCH institutions, e-Infrastructures, research and private organizations,
- iii. to establish the conditions for these sectors to integrated their efforts into a common work,
- iv. to identify the most suitable models for the governance, maintenance and sustainability of such integrated infrastructure for digital preservation of cultural content.

The main outcome will be a Roadmap for the implementation of a preservation infrastructure for DCH, supplemented by practical tools for decision makers and validated through a range of proof of concepts, where cultural institutions and e-Infrastructure providers will work together on concrete experiments, in cooperation with the private sector too.

DCH-RP adopts the eCulture Science Gateway developed in INDICATE for its proof of concept to validate the Roadmap with concrete experiments on the field..

The DCH-RP Roadmap is intended as the first instance of the Open Science Infrastructure for DCH in 2020.

DCH-RP will establish a practical liaison among the participants to the project that can represent a model of cooperation also for the rest of the sector. It will start with its thirteen partners, will then move to the 20 ‘external partners’ from Europe, Taiwan, India, Malawi, USA and South America who have already expressed their intent to participate to the study with the aim, eventually, to become pan-European and global.

Ground motion modelling and seismic hazard assessment

Seismic risk mitigation is a worldwide concern and the development of effective mitigation strategies requires sound seismic hazard assessment. The performances of the classical probabilistic approach to seismic hazard assessment (PSHA), currently in use in several countries worldwide, turned out fatally inadequate when considering the earthquakes occurred worldwide during the last decade (e.g [11]; [15]). Moreover, it is nowadays recognized by the engineering community that standard hazard indicator estimates (e.g. seismic PGA) alone are not sufficient for the adequate design, particularly when dealing with the protection of critical structures (e.g. nuclear power plants) and cultural heritage, where it is necessary to consider extremely long time intervals. A viable alternative, capable of minimizing the drawbacks of traditional PSHA, is represented by the neo-deterministic assessment of seismic hazard, NDSHA ([14] and references therein), based on a sound definition of credible scenario events. Current computational resources and physical knowledge of the seismic waves generation and propagation processes, along with the improving quantity and quality of geophysical data (spanning from seismological to satellite observations) allow, in fact, for viable numerical and analytical realistic modelling of ground motion, which permits the generalization of empirical observations by means of physically sound theoretical considerations.

The NDSHA approach permits to integrate the available information provided by the most updated seismological, geological, geophysical and geotechnical databases for the site of interest, as well as advanced physical modelling techniques, to provide reliable and robust basis for seismic design ([10]; [12] and [13]). A set of scenarios of expected ground shaking due to a wide set of potential earthquakes can be defined by means of full waveforms modelling, based on the possibility to efficiently compute synthetic seismograms in complex laterally heterogeneous anelastic media. In this way strong ground motion parameters can be defined at different scales – regional, national and metropolitan – accounting for a wide set of possible seismic sources and for the available information about structural models. The scenario-based methodology relies on observable data being complemented by physical modelling techniques, which can be submitted to a formalized validation process. The integrated NDSHA approach intends to provide a fully formalized operational tool for effective seismic hazard assessment, readily applicable to compute complete time series of expected ground motion (i.e. the synthetic seismograms) for seismic engineering analysis and other mitigation actions.

The NDSHA methodology has been successfully applied to strategic buildings, lifelines and cultural heritage sites, and for the purpose of seismic microzoning in several urban areas worldwide. Several international projects have been carried out and are still in progress based on the NDSHA methodology, including: the “MAR VASTO” project, with the participation of Italian (ENEA, Universities of Ferrara and Padua, ICTP) and Chilean (University Federico Santa Maria in Valparaiso, University of Chile in Santiago) partners; the UNESCO/IUGS/IGCP projects “Realistic Modelling of Seismic Input for Megacities and Large Urban Areas”, “Seismic Hazard and Risk Assessment in North Africa” and “Seismic microzoning of Latin America cities”; the multilateral-oriented network project “Unified seismic hazard mapping for the territory of Romania, Bulgaria, Serbia and Republic of Macedonia”, supported by the CEI (Central European Initiative).

The use of the NDSHA methodologies has been recently enabled on the EUINDIAGRID2 computational infrastructures, based on GRID computing paradigms. Such e-Infrastructure provides an innovative and unique approach to address this problem. They demonstrated to be an efficient way to share and access resources of different types, which can effectively enhance the capability to define realistic scenarios of seismic ground motion. Intensive usage of these infrastructure may enable scientists to compute a wide set of synthetic seismograms, dealing efficiently with variety and complexity of the potential earthquake sources, and the implementation of parametric studies to characterize the related uncertainties. In this framework, the CHAIN network can play an essential role in facilitating an extensive application of the NDSHA approach in a large number of countries worldwide, characterized by a different seismotectonic setting. This would provide the national and local communities (local authorities and engineers) advanced information for seismic risk mitigation in the study regions and, at the same time would provide a unique opportunity for large-scale validation of the developed methodology.

EGI VRC support

EGI is a federation of national and domain specific resource providers coordinated by a dedicated organisation called EGI.eu, set up as a Dutch foundation (Stichting), and governed by the EGI Council⁴. Technical activity is managed across EGI's three main areas – Operations, User Community, and Technology by a number of groups representing various communities. These groups develop policies and procedures that follow the principles of openness, transparency and consensus and that can be developed internally or in collaboration with external partners, such as the CHAIN project. EGI enables researchers within Virtual Research Communities (VRC) to collaborate, communicate and share resources across international boundaries by offering benefits such as the integration of community resources into EGI through user support and assistance, training and technology specialists and representation. EGI VRCs are self-organised research communities which give individuals within their community a clear mandate to represent the interests of their research field within the EGI ecosystem.

During 2010-2011 the EGI-InSPIRE project co-funded by the European Union under the 7th Framework Program (Research Infrastructures) established services across Europe to provide Human Services (for coordination and community building), Technical Services (for supporting the interaction of user communities) and Infrastructure Services (for securely accessing resource hosted by different organisations) for VRCs. The evolution of these services is driven by the needs coming from the researchers and infrastructure providers within EGI and the organisations they collaborate with internationally. This evolution is driven by a virtuous cycle that includes the prioritisation of their requirements, the fulfilment of these requirements by external technology providers, the assessment of the new technology releases to ensure they meet the original requirements, and then the deployment of new technology into the European Grid Infrastructure.

One of the main challenges for EGI over the coming years is to diversify its user-base by establishing VRCs from diverse scientific disciplines. EGI therefore – together with the National Grid Initiatives (NGIs) that provide the bulk of resources that make up the infrastructure and the VRCs – defined a strategic plan⁵ that integrates the existing services for VRCs and outlines the initiatives that can take place within the EGI community over the next two years, supported through the FP7 EGI-InSPIRE in the first instance and subsequent related projects that will, in the future with further investment, develop EGI's strengths in:

- European-wide coordination and interaction with research communities and national resource infrastructure providers
- Coordination, maintenance, operation and delivery of an open uniform European-wide federated production infrastructure
- Developing and promoting technologies for federating new resources
- Supporting the integration and operation of scalable interdisciplinary Virtual Research Environments personalised to each research community

⁴ European Grid Infrastructure: www.egi.eu

⁵ EGI Strategic Plan: <http://go.egi.eu/EGI2020>

In addition to this strategy and the support services that EGI members – NGIs and various projects – have been offering for the last two years, there now exists a dynamic short-term project framework, called Virtual Team framework⁶ that can be utilised to establish and operate services for new VRCs. Within the framework a number of projects have been already finished, and several other are currently running.

GISELA VRC Support

GISELA [16] (Grid Initiatives for e-Science virtual communities in Europe and Latin America) is a project co-funded by the European Union under its Seventh Framework Programme. GISELA aims at (i) implementing the Latin American Grid Initiative (LGI) sustainability model rooted on National Grid Initiatives (NGI) or Equivalent Domestic Grid Structures (EDGS), in association with CLARA [17], the Latin American NRENs and collaborating with the European Grid Initiative (EGI); (ii) Providing Virtual Research Communities (VRCs) with the e-Infrastructure and Application-related Services required to improve the effectiveness of their research.

In order to extend its user base and make Grid access and use easy for non IT-experts, GISELA has developed the VRC-driven Science Gateway [18] based on a model discussed within CHAIN. The VRC-driven GISELA Science Gateway is a web 2.0 environment which allows users to fully exploit the e-Infrastructure computing (jobs) and storage (data) services through a normal web browser exploiting the model described in the previous sections. Users can obtain access to the Science Gateway using the username/password pair (provided by their Identity Provider) and submit applications simply filling web forms.

The GISELA Science Gateway is considered the key tool to reach the sustainability of the e-Infrastructure in Latin America. Indeed, it has the potentiality to attract a big number of Latin American scientific communities to the Grid world. The number of Identity Federations users is indeed very big (several millions) then, thanks to the authentication procedure we implemented, we could give access to the Science Gateway capabilities to a huge number of potential users. At present, all users belonging to the Identity Federation of Brazil (CAFe [19]), Italy (IDEM [20]), Portugal (RCTSaai [21]), Spain (SIR [22]) and to the eduGAIN [23] inter-federation can access the GISELA Science Gateway. Those who do not belong to any identity federation are enrolled in the GrIDP “catch-all” federation [24]. Moreover, in collaboration with CLARA and the Latin America NRENs, we are promoting and encouraging the creation of Identity Federations in other Latin American countries and their eventual participation in eduGAIN.

After this simple authentication mechanism, users are able to run selected high-impact applications in a very easy way. Then, the distance, in term of competences, between the discovery of the Grid services and the ability to submit jobs will be substantially reduced thus opening the Grid world to a huge community of users.

We decided to organize the GISELA Science Gateway as VRC-driven tool: different sections, specific for each supported VRCs (Life Sciences, Earth Sciences, Industry, etc.), are available inside the portal and can be selected by the users according to the roles and privileges they are granted inside the Science Gateway.

⁶ EGI Virtual Team framework: https://wiki.egi.eu/wiki/Virtual_teams

Currently, we created 4 VRC-specific Science Gateways and integrated 7 applications. They are all listed below:

1. Cultural Heritage Science Gateway:

- a. ASTRA [25]: The ASTRA projects aims to reconstruct the sound or timbre of ancient instruments (not existing anymore) using archaeological data as fragments from excavations, written descriptions, pictures, etc.;

2. Industry Sciences Gateway:

- a. Industry@Grid: The objective of this collaboration is to gridify JobShop Scheduling and Supply Chain models using techniques based on heuristics and branch and bound methods. The achievement of this specific goal can create an important building block of a new industrial grid community, focusing on SMEs that will be able to use the grid as a leading-edge tool for their internal planning process;

3. Life Sciences Gateway:

- a. ClustalW [26]: it is a tool for aligning multiple protein or nucleotide sequences. The alignment is achieved via three steps: pairwise alignment, guide-tree generation and progressive alignment;
- b. GATE [27]: GATE (Simulations of Preclinical and Clinical Scans in Emission Tomography) is an advanced opensource software developed by the international OpenGATE collaboration and dedicated to the numerical simulations in medical imaging. It currently supports simulations of Emission Tomography (Positron Emission Tomography – PET and Single Photon Emission Computed Tomography – SPECT), and Computed Tomography (CT);
- c. Phylogenetics (MrBayes [28]): it is a program for the Bayesian estimation of phylogeny. Bayesian inference of phylogeny is based upon a quantity called the posterior probability distribution of trees, which is the probability of a tree conditioned on the observations. The conditioning is accomplished using Bayes's theorem;

4. Mathematics Science Gateway:

- a. Octave [29]: GNU Octave is a high-level interpreted language, primarily intended for numerical computations. It provides capabilities for the numerical solution of linear and nonlinear problems, and for performing other numerical experiments. The Octave language is quite similar to Matlab so that most programs are easily portable;
- b. R [30]: R is a free software environment for statistical computing and graphics. different seismotectonic setting. This would provide the national and local communities (local authorities and engineers) advanced information for seismic risk mitigation in the study regions and, at the same time would provide a unique opportunity for large-scale validation of the developed methodology.

Africa & Arabia ROC experience

The Africa and Arabia Regional Operation Centre (ROC) (<http://www.africa-grid.org>) has been set up as a joint effort by 5 major projects and initiatives: EUMEDGRID-Support, CHAIN, SAGrid, EPIKH and ASREN. It currently represents both a team of supporting personnel provided by different organization involved in the EuroMediterranean infrastructure EUMEDGRID, with a relevant contribution from Arab countries and the pan-arab consortium of NRENs, ASREN, and a gateway portal to relevant tools for monitoring and supporting the e-Infrastructure.



Figure 2 - Africa & Arabia ROC Web Site

In the transition from the EGEE model for the European Grid to the EGI based era, the centralized model for managing the continental infrastructure has evolved into a Federation of National Grid Initiatives and a coordinating, lightweight central body (EGI.eu).

As a consequence, the overall management of the infrastructure has been structured in a federated model based on NGIs and many previously centrally operated tools have been regionalized to the NGIs (NOC) scale;

SLAs have been enforced to the NGIs, many central tools have been dismissed, and regional tools have been tailored for national needs, to serve the National Grid Initiatives. In this process, some pioneering sites in Africa, previously operated by dedicated ROCs, lost their reference ROC: they became “orphan” in the sense that they missed a corresponding supporting Regional Operation Centre.

Starting in mid 2010, some Grid initiatives and projects dealing with Africa started gathering their efforts to establish a reference support centre for Africa: namely SAGrid

(NGI_ZA), EPIKH (Marie Curie action for know-how exchange and tutoring), EUMEDGRID-Support (Support action for the EuroMediterranean e-Infrastructure).

Initially a Support system was put in place for the EUMEDGRID-Support project related sites, based on an x-GUS instance (kindly provided by KIT, Germany), a customizable and lightweight version of GGUS; this was subsequently embedded into a ROC portal for Africa after the introduction of the idea of a new Regional Operation Center for Africa.

It appeared initially evident that – within the new EGI-production-level model – Africa ROC and its linked sites did not fit the Resource Infrastructure Provider role for EGI, as the SLAs were hard to be accepted and a coordinating overall Legal entity was missing.

Africa ROC did however play the fundamental general role of Support Centre for the Region. It continuously provided tools and a support team to sites in the region, without demanding the commitments from sites required by the EGI.

In some case, like the one of SAGrid, sites started an EGI Resource Provider accreditation process (MoU). In Dec 2010 CHAIN joined the supporting group for Africa ROC.

Step by step, in 2011 all major production tools for EGI Operations have been deployed for Africa ROC. Currently the ROC portal provides access to the following services:

- x-GUS (Support System)
- GSTAT (Grid Monitoring)
- Real Time Monitor (Grid Monitoring)
- GOC DB (Central DB storing sites information)
- NAGIOS (Grid Monitoring)
- MyEGI/Dashboard (Grid Monitoring)
- DGAS/HLRmon (IGI) (Accounting)
- SmokePing (Network monitoring)

This also provided very useful feedback to EGI itself on the portability and flexibility of their tools.

ASREN joined the team later on (third quarter of 2011), and as a consequence, the term ARABIA was added to the official denomination of the ROC: Africa & Arabia ROC was finally established.

The ROC currently supports sites in Sub-Saharan Africa, Nord-Africa, Arabia/Middle East. More specifically, the covered countries are: Morocco, Algeria, Tunisia, Egypt, Jordan, Palestine, Syria, UAE, Senegal, South Africa, Nigeria.

The Africa&Arabia ROC is currently available at <http://roc.africa-grid.org> (or <http://www.africa-grid.org>) and provides support information for users, like:

- ▶ Acceptable Use Policy for VOs
- ▶ User support system (x-GUS ticketing)
- ▶ Mailing list contacts
- ▶ Documentation for Users / Tutorials

and for Site Administrators, like:

- ▶ Gateway to all relevant tools (GOCDB, NAGIOS, GSTAT, SmokePing, DGAS, Dashboard)
- ▶ Links to Site Installation support pages

Starting from the meeting in Lyon at EGI TF2011, EUMEDGRID-Support decided the progressive endorsement of services by African and Arab Countries, in order to ensure sustainability of the ROC beyond the limited lifetime of the EUMEDGRID-Support project itself.

- VOMS (Authentication and Authorisation) (*ASREN, CERIST*)
- Top-BDII (InfoSys) (*CERIST, SAGrid, JUNet*)
- HLRmon/DGAS server (Accounting) (*CERIST, SAGrid*)
- Africa & Arabia ROC portal (ROC) (*JUNet*)
- GSTAT (InfoSys/Monitoring) (*CERIST, SAGrid, CNRST*)
- GOC DB (ROC) (*CERIST*)
- NAGIOS (ROC, Grid Monitoring) (*CERIST, CNRST, SAGrid*)
- MyProxy server (Workload Management/AAI) (*CERIST, CNRST*)

As already mentioned, the Africa and Arabia ROC is both a portal (single gateway to many operational tools) and a team of supporting experts for users and site administrators. At the end of EUMEDGRID-Support an official declaration was signed by partners committing to continue supporting the ROC and the infrastructure beyond the end of the project (Amman, Dec 15, 2011). Currently Ticket Process Manager (TPM) shifts for the Support System are provided on a monthly basis by partners; CERIST currently coordinates ROC related operations globally – ASREN will take over from them. SAGrid committed to provide a Security Officer for the ROC.

ASREN will progressively take over overall coordination of Africa & Arabia ROC from CERIST Along with new partners joining it. A possible evolution which is being considered is splitting the ROC into Africa and Arabia ROCs: it will have to be evaluated if sites grow in number, both to ease manageability and to cover the African and the Middle East regions more punctually

On the long term, CHAIN / CHAIN Reds are the ideal umbrella to plan and manage ROC related changes and transitions. It is certainly vital to keep the community alive and the acquired expertise operational.

The Africa and Arabia ROC is a relevant achievement of the joint effort by SAGrid, EUMEDGRID-Support, CHAIN, EPIKH and ASREN. It's a valuable tool to keep the community alive and continue pursuing the support to e-Infrastructures and their users in the region. It is important to keep investing on it and supporting through it the whole user communities in the region.

The CHAIN Roadmap, first thoughts

One of the most important outcomes of the CHAIN project is a roadmap for existing and emerging e-Infrastructures in the different regions to deploy the organizational approach studied and described in other different project deliverables. The roadmap is expected to describe the basic organizational model and the follow up activities needed to deploy it. The roadmap is to be based on the following major pillars:

- A detailed analysis of the changes that occurred in the European Grid Infrastructure with its new organizational model and its own plans for future (e.g., the EGI Strategy 2020 document and related EGI deliverables);
- the characteristics of the non-European e-Infrastructures, highlighting commonalities and differences including, but not limited to, organization, middleware, security, accounting, and user support;
- the opportunities offered by present and emerging standards with recommendations on the support of new possible standards;
- the emerging paradigms of virtualization, voluntary computing and cloud computing.

More detailed input is provided through CHAIN deliverables, most specifically the following four:

- D2.2 *Interoperability and interoperation guidelines*, that provides an analysis of the data collected by the state-of-the-art-in-regions questionnaire (described above) and proposed solutions for regional organizational and operational interactions;
- D3.2 *Road-map of trans-continental e-infrastructures for virtual communities*, that provides a proposal for a road-map from the point of view that is derived from the requirements and necessities that VRCs have;
- D4.1 *Specificities of the various regional e-Infrastructures*, that highlights and justifies the different activities needed to properly take into account the diversities of the regions;
- D4.2 *Report on the applicability of the preliminary results of the Organizational Study*, that present the results of pilots to demonstrate the applicability of the preliminary result of CHAIN work in this area. The pilots covered a couple of different regional situations, from developing (e.g. Mediterranean or Africa) to well established with its own middleware deployed (e.g. China or India).

The actually proposed CHAIN roadmap structure is represented in the following proposed Table of Content:

- EGI Status overview
This chapter will provide a consistent snapshot of the EGI organization and structure at the time of the CHAIN roadmap publication. It will reflect on-going changes in the development of the EGI, including highlights from the EGI vision and the EGI Strategy 2020.
- Common recommendations on the organizational model
The actual proposal, also studied in the D4.2 deliverable mentioned above, builds on a “ROC” (Regional Operation Centre) like approach. It reflects the limits of the EGI/NGI model that does not scale beyond a size of Europe-like region and proposes a self-organization per world region. Each region represents a unique entity that interacts with EGI and among them in a peer to peer structure. This CHAIN roadmap chapter will also discuss organization of the ROC and a different name may be proposed eventually, in order to avoid confusion with former EGEE ROCs.
- Common technical recommendations

Recommendations on a choice of technology to implement areas covered by the organizational model, providing multiple options when and where appropriate. Specific focus will be given to interoperability at the technical level, to support the peer to peer region-based organizational model. Both standard and emerging technologies will be covered, emphasising their contribution to the interoperability and easy to use for scientific communities.

- Global VRCs

An important part of the CHAIN roadmap will discuss the interaction of the organizational structure with global communities, defining proper interfaces for such interaction. The major expected message from this chapter is a call for synergies, to avoid building parallel infrastructures for each individual VRC and demonstrating the usefulness of a shared organizational model able to deploy different technology. To do so, the specificities of the different applications and codes deployed will be analysed in order to find as much synergies as possible or, at least, propose priorities.

- Region-specific recommendations

A realistic vision of what can be achieved in different regions in a given timeframe (2020) after the end of the CHAIN project. Major contribution by the regional project partners is expected here because they carry the valuable knowledge exclusively. A combination of the region-specific recommendations and each region vision will be provided.

The roadmap will become a strategic document for the upcoming CHAIN-REDS project, used also as a refinement of the CHAIN-REDS description of work.

Conclusions

The presentations have shown that the regional e-Infrastructure projects are investing significant resources in the support of VRCs. The current focus is on enlarging the list of applications supported by possibly discovering new VRCs that can make use of the existing e-Infrastructures. The final goal is of course to increase the number and variety of users to comply with the mission of being “general purpose” e-Infrastructures that can be sustainable.

Scientific Gateways (SG) has been considered and deployed by several projects in order to attract new users trying to lower the barrier of access to the Grid. How much the SG will contribute to the enlargement of the user base is still early to be said but the current experience looks promising. Moreover, the SG could possibly be expanded also in the Cloud domain.

The research community in Digital Cultural Heritage is an emerging VRC that has started to define its requirements through the DC-NET and INDICATE projects and continuing in the next future with DCH-RP. This community has specific needs of data storage and limited requests, so far, for computing resources. The Science Gateway approach has been recently adopted to have an easy access to applications and data. Cloud services are also under consideration for specific requirements of on-line/real-time services. The community recognises to have just started to enter in the concepts of e-Infrastructures and they would be interested to exploit the existing ones if they fit their needs.

The activities in the field of seismology are very relevant for Distributed Computing Infrastructures. The simulations and computational activities are quite heavy and require non trivial resources to produce reliable results on fine graded maps of large territories. A large interest in this field has been shown by several groups of researchers in different regions of the world from Asia to Latin America. The recent destructive earthquakes, including the Japan 2011 and Haiti 2010 events, have raised a vast consensus on the fact that new approaches are needed in order to reliably assess the risks and to develop effective mitigation plans. A true intercontinental VRC can be built on the existing groups and possibly large data challenges can be run to produce alternative maps of risks based on the neo-deterministic approach.

The CHAIN project has worked together with several VRCs to understand their requirements and validate the model that has been proposed for specific issues such as interoperability. The usage of Science Gateways has been accepted as a reasonable short and medium term solution that provides at least two benefits: lowers the barriers for the access and provides interoperability at application level transparently to the users. The Workshop, organised in collaboration with GISELA, has marked a fundamental step forward towards both the understanding of the requirements of the new VRCs, and making them aware of the opportunities offered to them by the e-Infrastructure providers.

The CHAIN Roadmap, under preparation, is taking stock of all the relevant topics discussed and the short-term and long-term solutions proposed in terms of interoperability and VRC support, e.g. Science Gateways and new Regional Operation centres.

References

- [1] CHAIN Deliverable D2.2 “Interoperability and interoperation guidelines”
- [2] www.europeana.eu
- [3] www.dc-net.org
- [4] The Conclusions of the Council of the European Union on the 2nd of December 2008 indicated the Joint Programming Initiatives (JPI) as the instrument to correlate the national investment on the research on the bid societal challenges in Europe with the resources make available on the same themes by the programmes of the European Commission. A JPI has been established among the European Member States, based on a variable geometry, specifically addressed to the Cultural Heritage. Available: http://ec.europa.eu/research/era/areas/programming/joint_programming_en.htm
- [5] www.indicate-project.eu
- [6] indicate-gw.consortio-cometa.it
- [7] www.liferay.com
- [8] glibrary.ct.infn.it
- [9] shibboleth.internet2.edu www.chain-project.eu
- [10] Field, E. H. and the SCEC Phase III Working Group; 2000: Accounting for site effects *in* probabilistic seismic hazard analyses of Southern California: overview of the SCEC Phase III Report. Bull. Seism. Soc. Am. 90: S1-S31.

- [11] Kossobokov, V.G. and A. K. Nekrasova; 2010: Global Seismic Hazard Assessment Program Maps Are Misleading. *Eos Trans. AGU*, 91(52), Fall Meet. Suppl., Abstract U13A-0020.
- [12] Panza, G.F., Romanelli, F. and Vaccari, F.; 2001a; Seismic wave propagation in laterally heterogeneous anelastic media: theory and applications to the seismic zonation. *Advances in Geophysics*, Academic press 43: 1-95.
- [13] Panza, G.F., Vaccari, F. and Romanelli, F.; 2001b: Realistic modelling of seismic input in urban areas: a UNESCO-IUGS-IGCP project. *Pure and Applied Geophysics* 158(12): 2389-2406.
- [14] Peresan A., E. Zuccolo, F. Vaccari, A. Gorshkov and Panza G.F.; 2011: Neo-deterministic seismic hazard and pattern recognition techniques: time dependent scenarios for North-Eastern Italy. *Pure and Applied Geophysics*. Vol. 168 (3-4), 583-607. DOI 10.1007/s00024-010-0166-1.
- [15] Stein S., Geller R. and Liu M. (2011). Bad assumptions or bad luck: why earthquake hazard maps need objective testing. *Seism. Res. Lett.*, 82:5.
- [16] www.gisela-grid.eu
- [17] www.redclara.net
- [18] gisela-gw.ct.infn.it/science-gateways
- [19] www.rnp.br/en/services/cafe.html
- [20] www.idem.garr.it
- [21] <https://rctsaai.fcn.pt/>
- [22] www.rediris.es/sir/
- [23] www.edugain.org
- [24] gridp.ct.infn.it
- [25] www.astraproject.org
- [26] www.clustal.org/
- [27] www.opengatecollaboration.org/
- [28] mrbytes.sourceforge.net/
- [29] www.gnu.org/software/octave/
- [30] www.r-project.org/