Co-ordination & Harmonisation of Advanced e-Infrastructures
Research Infrastructures – Support Action
Grant Agreement n. 260011
www.chain-project.eu

SEVENTH FRAMEWORK PROGRAMME
Capacities specific program - Research Infrastructures
FP7–INFRASTRUCTURES–2010-2

Project acronym: *CHAIN*
Project full title: Co-ordination & Harmonisation of Advanced e-Infrastructures
Grant agreement no.: 260011
Start date of project: 1 December 2010
Duration: 24 months

Deliverable 3.2
"Road-map of trans-continental e-infrastructures for virtual communities"

Version 0.3

Due on: 30/11/2011 Submitted on: 30/11/2011

Organisation name of lead beneficiary for this deliverable: CIEMAT
Dissemination Level: Public

Abstract:

This deliverable provides a proposal for a road-map for existing and emerging e-Infrastructures for Virtual Research Communities (VRC). Such a road-map is mainly focused on the VRC previously identified in D3.1 'Trans-continental scientific and technical communities' and their related information updated during the last six months.

Document log

Issue	Date	Comment	Author
V0.1	16/11/2011	First draft	Rafael Mayo and Raquel Muñoz (CIEMAT)
V0.2	29/11/2011	Reviewed text	Federico Ruggieri (INFN)
V0.3	30/11/2011	Final Version	F. Ruggieri (INFN)

Table of contents

1. Executive Summary

As a consequence of the study of the state of the art in Grid developments across the world regions and their relation to the European Grid Initiative performed by CHAIN WP2, two different deliverables were submitted with the achieved conclusions: D2.2 'Interoperability and interoperation guidelines'; and, D4.1 'Specificities of the various regional e-Infrastructures'. In the present document, a third report is made, this time focused on the requirements that are demanded by the Virtual Research Communities to the Distributed Computing Infrastructures.

The first step carried out by CHAIN WP3 has been to update the information provided in D3.1 'Trans-continental scientific and technical communities' about the previously identified VRCs. These groups or initiatives were selected because of their potential as sustainable communities in the Continents targeted by CHAIN, so actions on that direction have been also taken. The list of identified VRCs is the following:

- We-NMR;
- WRF4G;
- jModelTest / ProtTest;
- INDICATE / DC-NET;
- DECIDE;
- Climate Change
- SuperB

With the fist two of them, CHAIN has signed a Memorandum of Understanding, fulfilling in this way the last WP3 milestone in due time: 'MS07 - Agreements with reference communities signed'. With the rest, improvements on their characteristics for becoming a VRC or further contacts have been taken. A special mention should be devoted to the Climate Change group, where the main effort has been focused on consolidating a former community.

In addition to the aforementioned actions, two main events have been organised for identifying the list of services currently required by these VRCs, both of them organised by CHAIN and GISELA initiatives: 'Resource Infrastructure Providers meet VRC' held within the EGI User Forum (Vilnius, April 2011); and, 'Resource Infrastructure Providers meet VRC (II)' held within the EGI Technical Forum (Lyon, September 2011). The first seed for the Climate Change community was established in the 'Conference on the Role of e-Infrastructures for Climate Change Research' (Trieste, May 2011).

In addition, information about the VRC requirements has been also obtained by means of the CHAIN survey already described in D2.1 'State of the art questionnaire'.

The main aspects addressed by the proposed road-map are grouped in three main topics: 'Technical', 'Training, administration and use' and 'Collaborative' recommendations. The whole list can be found in the subsection 4.3 of this document, but above all, beyond VRCs and DCIs, sustainability must be secured by NGIs and higher political structures.

2. Introduction

e-Science has been promoted by the EC over the last decade by the FP6 and FP7 funding programmes. This effort has been focused not only in Europe, but also in different world regions, mainly Africa, Asia and Latin America. At the same time, the aim has been to support different layers of e-Infrastructures, computing platforms and related human power (Research and Education Networks, Grid and HPC infrastructures and Virtual Research Communities).

As a result, probably Grid emerged as the main actor where the coordination of these world-wide efforts has been mostly pushed, but even in this case, coordination has merely restricted to basic operational, organisational and technological know-how transfer/exchange, i.e. no big advances have been made until now to link e-Infrastructures at intercontinental level due to their specific requirements which depend on the targeted region.

The CHAIN project, started on December 1st, 2010, aims to coordinate and leverage the efforts made over the past 6 years to extend the European e-Infrastructure (and particularly Grid) operational and organisational principles to a number of regions in the world, mainly those identified for this purpose by the EC. This vision is structured in several lines of performance: study of the state of the art of the current e-Infrastructures and e-Science status per region; needs and commonalities of these computing platforms and VRCs requirements and services.

In this sense, two main steps forward have been taken by delivering D2.2 'Interoperability and interoperation guidelines' and, D4.1 'Specificities of the various regional e-Infrastructures'. In such deliverables, an analysis of the questionnaire data launched by WP2 is made, specificities of the various regional e-Infrastructures are described and recommendations for regional organisational and operational interactions and suggested developments are proposed.

CHAIN WP3, 'Present and emerging needs of trans-continental scientific communities', devotes its framework to the study of VRCs. As a first step, a coordinated data collection on the existing Grid state of the art across the world was carried out. Such a work was summarised in D3.1 'Trans-continental scientific and technical communities', which focused on the description of the results related to VRCs obtained up to the sixth month of the project, i.e. May 2011, where several candidates communities were identified according to their further sustainability and presence in the continents of interest to CHAIN.

In this document, a similar study to those previously done by WP2 in D2.2 and WP4 in D4.1 is presented. Thus, a road-map of trans-continental e-infrastructures for virtual communities is described.

2.1. Purpose

The aim of this deliverable is to provide a proposal for a road-map for existing and emerging e-Infrastructures for Virtual Research Communities. These guidelines have been mainly obtained from a continuous liaison with the scientific and technical communities that have or could have trans-continental activities and were previously identified during the first semester of the project. At the same time, information on very well established VRCs has been collected and actions on forming new communities have been also taken.

To do so, several actions were previously performed during the first twelve years of the project in order to obtain the necessary information:

- Specific parts (one for regional infrastructures and one for national infrastructures) for VRCs were included in the state-of-the-art questionnaire for e-Infrastructure -and specifically for Grid infrastructures- that was delivered by CHAIN WP2, 'Consolidation of existing state of the art'.
- WP3 contacted and identified VRCs candidates who could provide spread continental presence as well as scientific interest and sustainability. A continuous exchange of information has been maintained in order to enhance the terms of the collaborations.
- Two specific workshops have been organised by CHAIN and held within EGI User Forum 2011 and EGI Technical Forum 2011 in order to know the requirements and necessities that VRCs have.
- A specific High Level conference devoted to Climate Change has been organised and promoted by CHAIN in order to start the consolidation of a related VRC.

This deliverable describes all these activities and the consequent proposed road-map.

2.2. Glossary

CHAIN	Co-ordination and Harmonisation of Advanced e-Infrastructures
DCI	Distributed Computing Infrastructure
DoW	Description of Work – Annex I to the GA
EC	European Commission
EGI	European Grid Initiative
EGI-InSPIRE	European Grid Initiative-Integrated Sustained Pan-European Infrastructure
EPIKH	Exchange Programme to advanced e-Infrastructure Know-How
FP6/FP7	European Commission's Framework Programme Six / Seven
GA	Grant Agreement
HEP	High Energy Physics
HPC	High Performance Computing
JRU	Joint Research Unit
KoM	Kick-off Meeting
MoU	Memorandum of Understanding
MS	Milestone
NREN	National Research and Education Network
OLA	Operating Level Agreement
ROC	Regional Operation Centre
PMB	Project Management Board
SLA	Service-Level Agreement
VO	Virtual Organization
VRC	Virtual Research Community
WP	Work Package

3. Updated list of scientific and technical communities

This section mainly updates the information already provided in D3.1[1] 'Trans-continental scientific and technical communities' as already stated in the CHAIN Description of Work, i.e. in the twelfth month of the project there should be a revised version of D3.1 to be included in D3.2 . Thus, an analysis on the results obtained in the VRC-oriented sections that appeared in the CHAIN surveys (both Regional and National) described in D2.1 'State of the art questionnaire'[2] is present.

Later on, information regarding the different actions performed for the identified VRCs is shown as well as the achievement of the last WP3 milestone, which is directly linked to the scientific communities.

3.1. Precedents

Two types of questionnaires were drawn up for the WP2 survey: a Regional version which aimed to collect the data relevant for the regional infrastructures, their modes of operation and interoperability, etc; and, a National version that focused on existing or nascent National Grid Initiatives and their attributes. These surveys were opened in April 2011 and their analysis can be consulted in D2.2 'Interoperability and interoperation guidelines'[3] and D4.1 'Specificities of the various regional e-Infrastructures'[4]. In this report, there will be a very brief summary of them in Section 4. Both surveys contained a specific section where questions devoted to VRC topics were present; such list can be consulted in D3.1[1].

In addition to this questionnaire, WP3 contacted several groups that fulfilled two main aspects: they were aware of the current sustainability issues and how these should be overcome in order to maintain the community alive in a long term; and, they had a potential presence in the regions targeted by CHAIN. Since these actions carried out by WP3 started from the very beginning, i.e. the project KoM[5], the WP2 questionnaire already contained some questions specifically devoted to collect information related to some of these previously identified VRCs.

At the end, sixth month of the project, six (6) candidates were identified and continuous contacts have been kept since that moment in order to enhance the collaboration between them and CHAIN. The latest advances will be summarised in Section 3.3 per VRC, but the reader can find a deeper explanation of the work carried out during the first semester of the project in D3.1[1].

In order to link Distributed Computing Infrastructure providers and scientific communities, CHAIN and GISELA projects organised two meetings within EGI events in the first year of the lifetime of CHAIN. In April 2011, the workshop entitled 'Resource Infrastructure Providers meet VRC' was held within the EGI User Forum in Vilnius (https://www.egi.eu/indico/sessionDisplay.py?sessionId=19&confId=207#20110413). Six months later and due to its great success, a second round, this is, 'Resource Infrastructure Providers meet VRC (II)' workshop, was also held within the EGI Technical Forum in Lyon (https://www.egi.eu/indico/sessionDisplay.py?sessionId=25&confId=452#20110921) during September. Both events have been a great opportunity for sharing experiences between DCIs

[1] CHAIN D3.1, http://documents.ct.infn.it/record/500?ln=en
[2] CHAIN D2.1, http://documents.ct.infn.it/record/486?ln=en
[3] CHAIN D2.2, http://documents.ct.infn.it/record/506?ln=en
[4] CHAIN D4.1, http://documents.ct.infn.it/record/499?ln=en

and VRCs and have been very useful for the road map that it is presented in this document in the same way that the CHAIN KoM[5] and the CHAIN launch event[6] were the seed for forming the first list of candidates in what concerns scientific communities.

In addition, a special focus was set on the Climate Change community, so the first actions to drive such group of researchers into a VRC were taken by the joint organisation among several initiatives (CHAIN included) of the 'Conference on the Role of e-Infrastructures for Climate Change Research' (Trieste, May 2011, http://users.ictp.it/~smr2238/).

At this point, it is worth mentioning the excellent work done by CHAIN WP5 by implementing in the webpage of the project the 'Knowledge base'[7] and 'Applications'[8] sections, which are very useful for providing information not only to the rest of the project Activities, but also to the general public. Such compilations have been very useful for WP3 whenever it has planned any future action.

3.2. WP3 milestone 'MS07 Agreements with reference communities signed'

According to the CHAIN DoW, the third and last WP3 milestone is 'MS07 Agreements with reference communities signed', the fulfilment of which was scheduled for August 2011 by signing several MoUs. To do so, WP3 worked on a first template that intended to specifically address the necessities and requirements that any scientific community could have in what concerns a project like CHAIN, i.e. this template would have significant differences with the one that was adopted to be ready with e-Infrastructures providers. Later on, of course, this template would be customised by the VRC in agreement with CHAIN.

Conversations with a couple of VRCs were started in early July in order to achieve the milestone and final agreements on the final version of the text were obtained in September 5th with We-NMR and September 7th with WRF4G. A little delay can be then inferred, but bearing in mind the summer break, which affected the European Institutions, and the long round among partners that this kind of agreements require, it can be considered that the milestone was successfully fulfilled.

Profiting from the EGI Technical Forum and the coordinators' presence of these initiatives, the MoUs were effectively signed in September 19th (WRF4G) and 21st (We-NMR). The reader can find both documents in the following links:

- We-NMR http://documents.ct.infn.it/record/503?ln=en
- WRF4G http://documents.ct.infn.it/record/502?ln=en

The CHAIN webpage published the related news in a two-fold basis. The agreement on the final version by the two VRCs and CHAIN was uploaded in September 9th (see Fig. 1) while the specific fulfilment of the milestone was published after the definitive signature in September 21st (Fig. 2).

The whole list of CHAIN milestones is available at http://www.chain-project.eu/milestones

[5] CHAIN KoM, http://agenda.ct.infn.it/conferenceDisplay.py?confId=464
[6] CHAIN Launch Event, http://agenda.ct.infn.it/conferenceDisplay.py?confId=495
[7] Knowledge base in CHAIN webpage, http://www.chain-project.eu/knowledge-base
[8] Applications in CHAIN webpage, http://www.chain-project.eu/applications

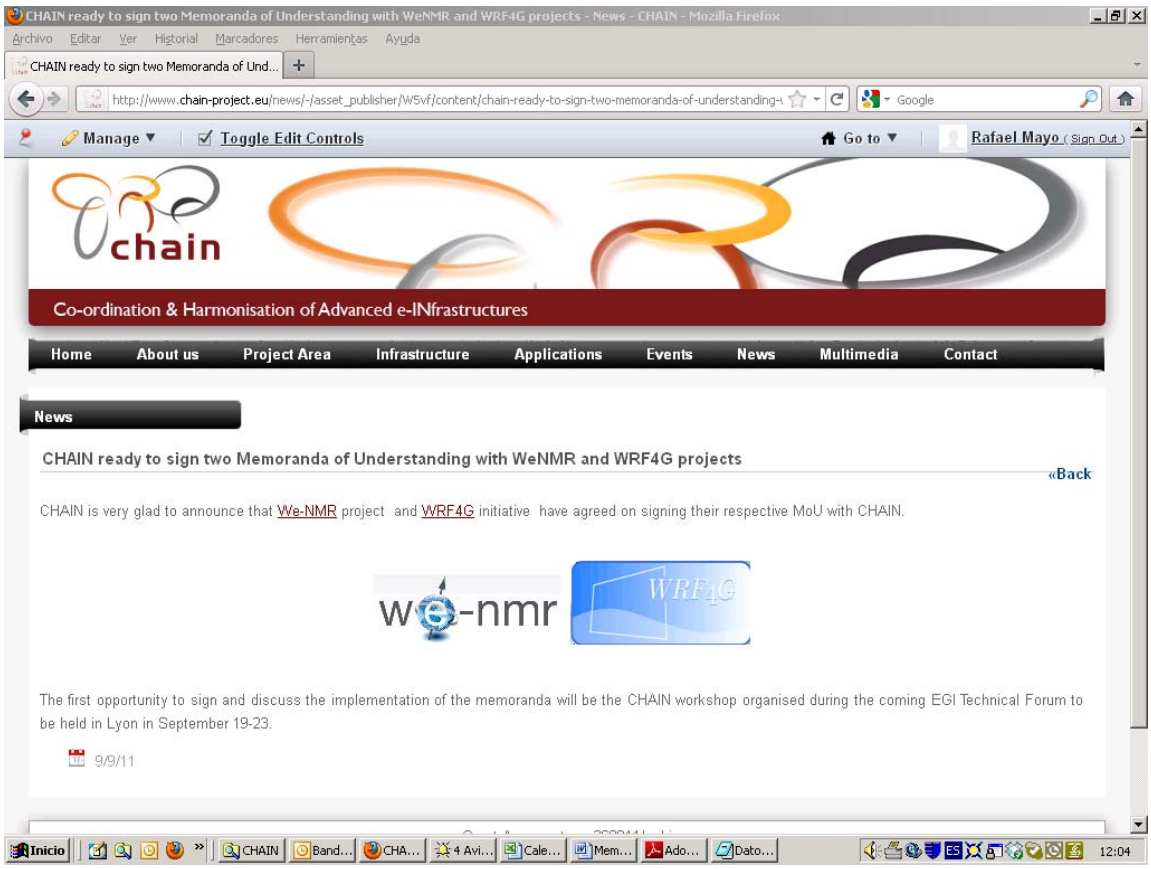

Figure 1. Announcement on the agreement of the MoU to be signed

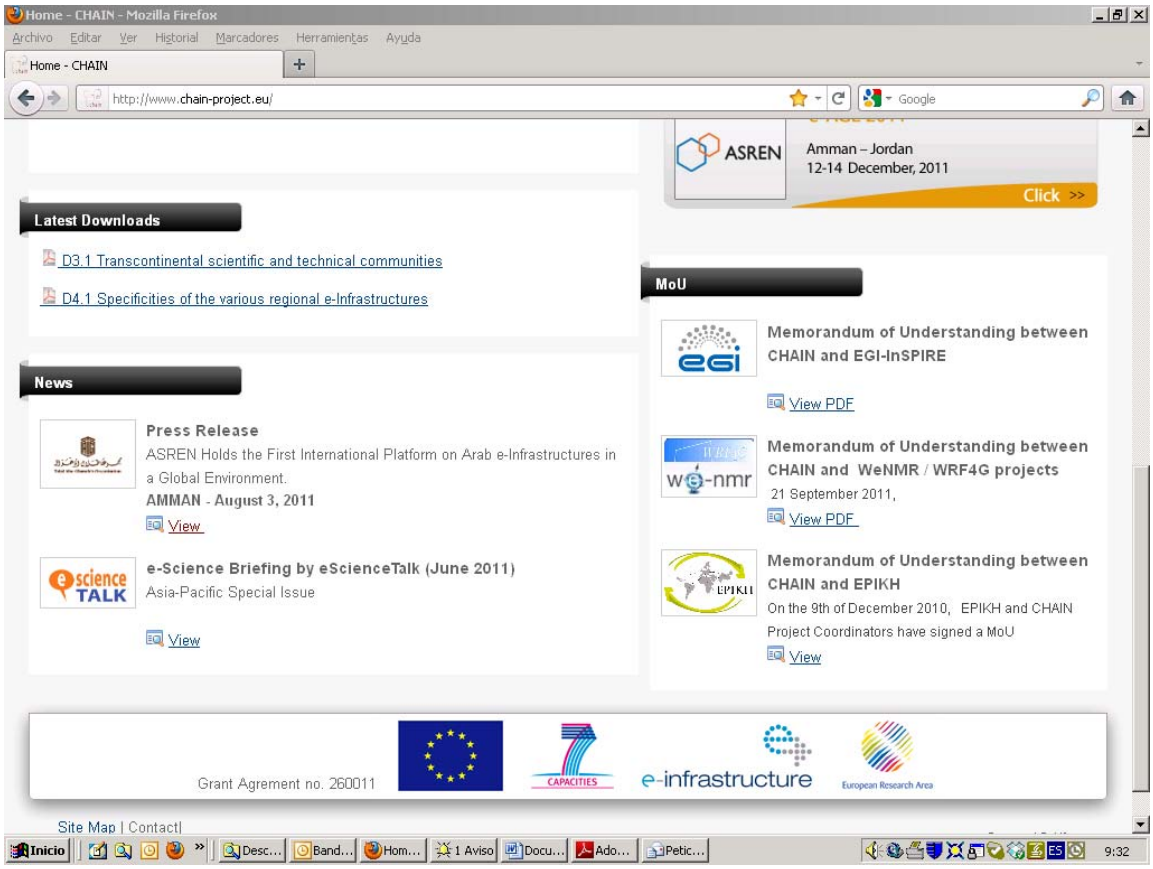

Figure 2. News related to the final signature of the MoU

3.3. Current status

In the following subsections, information about the new advances regarding the firstly identified VRCs will be detailed. As a common basis for all of them, continuous exchange of information between them and CHAIN has been maintained.

3.3.1. We-NMR

We-NMR (http://www.wenmr.eu/) is a project which aims to optimize and extend the use of the NMR and SAXS research infrastructures through the implementation of an e-Infrastructure in order to provide the user community with a platform integrating and streamlining the computational approaches necessary for NMR (Nuclear Magnetic Resonance) and SAXS (Small Angle X-ray Scattering) data analysis and structural modelling.

Since the submission of D3.1, it is worth mentioning the MoU signed between this initiative and CHAIN. Previous to that, a specific question was included about We-NMR and its presence in Asia, Africa and Latin America beyond the current associated or involved researchers. The aim was to know from the National representatives new researchers interested in this kind of studies, i.e. NMR and SAXS. Several answers were positive, i.e. there was an identification of potential users, but most of them were contacts already involved in We-NMR. Nevertheless, two new possibilities emerged: one coming from the Department of Biology of the University of Burundi and the other the from Research Centre for Cellular and Molecular Biology and the Research Centre for Haematology and Related Disorders, both belonging to the University of Costa Rica. Actions for involving those groups in We-NMR have been started jointly by this project and CHAIN.

With regards to the aforementioned workshops organised by CHAIN & GISELA within EGI conferences, We-NMR participated by making a presentation.

3.3.2. WRF4G

WRF4G (http://www.meteo.unican.es/es/software/wrf4g) is a Grid version of the well-known Weather Research and Forecasting (WRF) modelling system application. It is widely used by the meteorological agencies and many other groups in the Earth Science domain. Its Grid-based version has increased the resources where tasks such as Idealized simulations, Regional and Global applications, Parameterization and Data assimilation research or Forecast and hurricane research can be performed.

Since the submission of D3.1, it is worth mentioning the MoU also signed between this initiative and CHAIN. Previous to that, a specific question was included about WRF4G and its presence in Asia, Africa and Latin America beyond the current associated or involved researchers. The aim was to know from the National representatives new researchers interested in this kind of studies, basically regional forecast. Several answers were positive, i.e. there was an identification of potential users, but most of them were contacts already involved in WRF4G. Nevertheless, new possibilities emerged: the Chinese Academy of Meteorological Sciences; the IGEBU and DSS-NBI Centres of the University of Burundi; Cubanergía; and the Laboratory for Atmospheric and Interplanetary Physics from the University of Costa Rica. Besides, the NGI representatives of Ethiopia, Sudan and Congo showed their interest in looking for researchers who could be interested in this application.

With regards to the aforementioned workshops organised by CHAIN & GISELA within EGI conferences, WRF4G participated by making a presentation, in the framework of the Climate Science community in the User Forum and independently in the Technical Forum. Even more,

in this second case, a new application was also presented: CAM4G, which is of importance since it demonstrates that the methodology used in making a distributed version of WRF4G is also valid for other codes such as CAM, broadly used too. The direct consequence of this fact is that other well-known applications employed world-wide, RegModel for example, could follow the aforementioned developments.

In this sense, as it will be commented below, WRF4G has been proposed to be tested and used by the people who are currently interested in the Climate Change VRC and its code owner has offered its support to apply this new methodology used also in CAM4G to other codes. In addition, the owner group of the application, which agreed on disseminating CHAIN plans and outcomes inside the Climate Change CORDEX project[9], maintains this commitment.

3.3.3. jModelTest / ProtTest

These two applications belong to the Life Sciences domain and, in particular, to the Evolutionary Biology. Both are freely available on-line (http://darwin.uvigo.es/) for the statistical selection of best-fit models of nucleotide substitution (jModelTest) and amino-acid (ProtTest) replacement for a given set of aligned sequences. Thus, many researchers interested in molecular systematics, phylogenetics, phylogenomics, molecular evolution and/or bioinformatics use them continuously all around the world. ModelTest has around 30,000 registered users world wide while jModelTest (the Java version) has around 11,000 and ProtTest around 5,000, i.e. almost all countries in the world have scientists that are using one of these tools.

As in the previous VRCs, a specific question related to these codes was incorporated in the WP2 questionnaire. Thus, new interested researchers were located in Ethiopia, Nigeria, Sudan, Democratic Republic of Congo, Panama, Burundi, Taiwan and Costa Rica. All of them will be contacted in the near future.

The reason for that is because CHAIN partner CIEMAT keeps on working with the owners of the codes in order to further spread and facilitate the use of both applications. Once both jModelTest and ProtTest3 had a Grid version[10,11], new heuristics are being applied in order to speed up their performance. Once these developments will be achieved, both applications are intended to be incorporated in the Scientific Gateways portal already implemented by CHAIN partner INFN. This strategy could not be straightforward since Scientific Gateways are based on Liferay, SAGA and specific developed portlets and the Grid versions of jModelTest and ProtTest3 use DRMAA, but a final successful integration is expected anyway. Once this final step forward will be taken, a broadcast announcement of the new easy-friendly release of both applications will be sent by CHAIN and by the code owner to all the registered users.

3.3.4. INDICATE / DC-NET

These projects (http://www.indicate-project.eu and http://www.dc-net.org) are working on coordinating policy and best practices regarding the use of e-Infrastructures for Digital Cultural Heritage. The projects aim at establishing and stimulating a network of common interest made up of experts and researchers in all the relevant fields, whose sustainability will be planned on a long term beyond the project lifetimes.

[9] CORDEX initiative, http://www.meteo.unican.es/en/projects/CORDEX
[10] M. Loureiro *et al.* Studies in Health Technology and Informatics **159**, 244-248 (2010)
[11] M. Rodríguez-Pascual *et al.* Proc. HealthGrid 2011 Conference, H5 (2011)

Humanities have a very different approach to ICT than Science. In addition, their interests and required services are diverse. But Cultural Heritage is an issue of concern everywhere and the CHAIN regions of interest are rich in terms of Archaeology and/or History. That is why, in order to offer a different technological point of view, CHAIN maintains a line of work with INDICATE and DC-NET coordinator from the very beginning, so she has participated in the workshops organised by CHAIN and GISELA within the EGI events in 2011 presenting their specificities. Once a well defined collaboration has been set up by CHAIN with the VRCs that were firstly identified (see subsections 3.3.1 and 3.3.2), focus will be set on new communities such as INDICATE and DC-NET. In this way, conversations with the people who filled the WP2 questionnaire will be established in order to ask for new interested users.

3.3.5. DECIDE

DECIDE (http://www.eu-decide.eu/) objective is to design, implement, and validate a Grid-based e-Infrastructure building upon neuGRID[12] and relying on the Pan-European backbone GEANT[13] and the NREN. Over this e-Infrastructure, a service will be provided for the computer-aided extraction of diagnostic markers for Alzheimer's disease and schizophrenia from medical images.

DECIDE is an initiative with a high social impact that is planned to profit from its own developments and achievements, but also from the CHAIN work in a two-fold basis: dissemination and We-NMR experience. This VRC, that is part of the more generic Health community, has also very specific requirements in terms of privacy and authorisation to access information.
As in the previous subsection case, focus will be set on DECIDE interests too and conversations with the people who filled the WP2 questionnaire will be established in order to ask for new interested users. Again, DECIDE representatives presented the project in the CHAIN & GISELA workshops held jointly with EGI.

3.3.6. Climate Change

In accordance with previous contacts maintained by some CHAIN partners, and the previous Conference on the "Role of e-Infrastructures for Climate Change research" held in Trieste in May 2011 (http://users.ictp.it/~smr2238/), the community of researchers working on Climate Change studies was identified as a potential VRC. It is large and has a presence in many countries around the world.

The Climate Change community members have already exploited large (sometimes dedicated) computing facilities, but their constantly expanding activities require even larger computing resources, which is of importance for CHAIN in order to address new problems and gather new requirements.

WP3 has been working on listing related researchers beyond Europe. Thus, twenty (20) contacts in Asia, Africa and Latin America have been obtained and several conversations have been held. The first objective of such relationship was to define the interests and applications that these scientists had in order to find commonalities. Besides, a search on ongoing projects that could also fit in the CHAIN aim and the Climate Change field was performed; as a result, initiatives such as IS-ENES[14], TRECCAfrica[15], IPCC-DDC[16] and/or agINFRA will be contacted in the near future.

[12] NeuGRID project, http://www.neugrid.eu/
[13] GEANT network, http://www.geant2.net/
[14] IS-ENES project, http://www.enes.org/
[15] TRECCAfrica project, http://www.treccafrica.com/

In the mean time, CHAIN has offered to the already identified researchers to use WRF4G. As it was previously mentioned, this code is well suited to perform regional modelling of usefulness to Climate Change and the methodology that has been employed to port it to the Grid and to other distributed platforms could be adapted in the future to enhance the capabilities of other codes (RegCM, GFS, etc.) that are being currently used by these scientists.

As it has been mentioned through the text, there exists a close collaboration between CHAIN and GISELA; even more, both projects have common partners like INFN and CIEMAT. Thus, GISELA, by means of its dissemination work package, has opened a new line of collaboration with groups researching on Seismology that has been extended to the rest of activities. Being INFN the Manager of the VRC WP in GISELA and CIEMAT its deputy and just the opposite in CHAIN, the possibility of joining this new community and the Climate Change VRC will be weighed up in the near future.

3.3.7. SuperB

. It is well known that the Italian Ministry of University, Research and Education has approved the plan of building a new accelerator (SuperB) in the area of Rome (in the premises of University of Tor Vergata); the plan is to have the accelerator deployed and running with an experiment taking data for 2017SuperB will be a heavy flavour factory that will provide complementary information to LHC, looking at rare decays with a very high luminosity electron-positron asymmetric collider (http://superb.infn.it/). . INFN has proposed this new Research Infrastructure and participates as the major shareholder of a new consortium “Cabibbo Lab”. The new accelerator has awakened the interest of several research groups in Europe and USA and its construction will make use of several parts coming from the dismantled accelerator in SLAC and the experiment will reuse many parts of the BaBar experiment. The facility will thus become a truly international laboratory and will be transformed in a ERIC (European Research Infrastructure Consortium) as soon as other European organisations will join it.

CHAIN is investigating how to promote this new SuperB Virtual Research Community in Europe and other regions since the four LHC experiments already have ongoing collaborations in Asia, Africa and Latin America (India, Taiwan, Morocco, South Africa, Egypt, Mexico, Brazil, Colombia, etc.) and, of course, Europe, CHAIN will work in the following months on fostering the creation of a VRC devoted to this new experiment.

[16] IPCC-DDC, http://www.ipcc-data.org/

4. Proposed road-map

In this section, the first analysis devoted to draw up a proposal for a road-map for existing and emerging e-Infrastructures for VRCs is made. In order to better define the scenario in which such a proposal will be made, a brief description of the current status of the DCIs and the actions performed for getting valuable feedbacks from the VRCs are documented.

4.1. Current status of DCI

In order to assess which new capabilities should be developed by the different DCIs that are operating in the Regions of interest to CHAIN, it is mandatory to know their current status in what concerns their own developments, characteristics, computational and human resources to exploit the previous ones and, finally, their level of maturity with regards to EGI. Also, sustainability has to be borne in mind.

Such a study about the current status of the DCIs has been successfully developed by CHAIN by means of D2.2 'Interoperability and interoperation guidelines'[3] and, D4.1 'Specificities of the various regional e-Infrastructures'[4]. In what follows, an overall summary of them in terms of related topics to VRC is described (particularities to different Regions should be consulted in the aforementioned references).

About interoperation and interoperability and taken into account the current status, it could be documented that it would be an asset:

- the settlement of a consolidated domestic NGI, which provides in-depth documentation (that could be published through CHAIN website and circulated among the NGIs in CHAIN target regions);
- to progress from JRU model on regional level through to establishing a dedicated legal body and financial model;
- to prepare a cook book of recipes on how to set-up and manage fully fledged ROCs, possibly complemented by Wiki pages or community tools;
- to make a sustainability analysis for the ROCs that have not yet done it and a dissemination plan of its existence in the region;
- to ensure a service quality by means of SLA and SOP;
- the integration of the current existing ticket systems with GGUS or the use of a GGUS Regional/National support unit directly (or XGUS) in order to allow users to submit ticket regardless the infrastructure provider/region;
- the organisation of more training activities on application migration (and its associated friendly use, i.e. Scientific Gateways for example) for not only one infrastructure;
- the implementation of standard:
    - accounting mechanism, or at least monitoring information exchange, for as much middleware implementations as possible
    - core services
    - job submission and data exchange mechanisms;
- to coordinate the security related efforts and to participate in the corresponding well established bodies; and,
- to form VRCs that would group the related researchers of a common field and its liaison with similar communities in Europe and the other regions avoiding duplicated efforts.

About e-Infrastructures, their status can be summarised as follows:

- there is of wide interest in the EGI solution, as EGI related services are available in all regions and, in most of them, they are the basis of regional (and national) grid infrastructures; notable exceptions to this rule are
    - China, with its own technology CNGrid
    - India, using adapted Globus services;
- all regions have some interest in voluntary (desktop) Grid services, although the actual deployment varies;
- the further expansion of Grids is related to the national wealth and the innovation (added value) provided by the infrastructure, i.e. the differences in size of the infrastructures (measured through the number of CPUs available and that of registered users, which can be taken as independent variables) usually reflects the economic strength of individual regions
    - CPU cores ranges from 43000 (China) to 900 (Middle-East), being ASREN starting its activities
    - Online storage is large in China and Asia-Pacific (> 1PB), medium in Africa and India and small in the rest (< 100 TB)
    - active users ranges from 2600 (China) to 100 (India)
- there is a link to GÉANT through several specific projects. EUMECCONNECT3, TEIN3, etc.;
- HEP, Bioinformatics/Biomedicine and Earth Sciences are supported in all regions;
- there are huge differences in what respect the organizational structures
    - China and India, being countries with a strong central government, have a formally established coordinating body
    - Latin America, the Mediterranean, Middle East and Gulf countries have a region-wide coordinating body, with some kind of governmental support.
    - Asia-Pacific and Africa Grid infrastructures are built in a bottom up approach, without explicit (formal) coordinating bodies established
- Asia-Pacific and Latin America reports on sustainability plans

4.2. Direct actions performed

In addition to the general overview of e-Science and e-Infrastructures operating in the world, excluding North America, Australia and Japan, which determines the level of interoperation that could be adopted by the different VRCs, it was mandatory to know from the community users the requirements and necessities that they have for performing their studies and research.

In this way, continuous conversations with the identified VRCs, as it was stated in Section 3.3, were carried out in order to get updated information about how they were evolving. But in addition, the following actions were also promoted.

4.2.1. Events

It has been already mentioned that CHAIN has organised jointly with GISELA two specific workshops, the aim of which was to offer to the VRCs the possibility of presenting their requirements to the DCIs and, to the DCIs, doing the same with regards to their capabilities.

In the first event 'Resource Infrastructure Providers meet VRC', held within EGI UF 2011 (https://www.egi.eu/indico/sessionDisplay.py?sessionId=19&confId=207#20110413), the following initiatives made a presentation: DC-NET, Earth Science, Life Science and We-NMR as VRC; and, CHAIN, DEGISCO, EGI and GISELA as providers. Of course, a final round table was also held.

Lately, in 'Resource Infrastructure Providers meet VRC (II)', held within the EGI TF 2011 (https://www.egi.eu/indico/sessionDisplay.py?sessionId=25&confId=452#20110921), there were presentations that corresponded to these initiatives: Agricultural Research, Climate Change, Life Science, South African applications and We-NMR as VRC; CHAIN, EMI, EUIndiaGrid2, GISELA as providers; and, an additional presentation regarding interoperability and standards.

It is important to point out that these workshops have played a key role in the work to be carried out by WP3 and similar events are intended to be organised during the second year of the project profiting from the EGI Community Forum to be held in Gärching in March 2012 and the EGI Technical Forum to be held in Prague in September 2012. The reason for that is not only the benefits that this exchange of information produce in both VRCs and DCIs providers, which would justify its organisation by itself, but also the possibility of attending other parallel session that are offered in such EGI major events. Thus, 'Heavy Users Community', 'VRC' and 'Portal technologies' sessions were outstanding fora for getting aware of the latest developments, necessities and drawbacks of the different initiatives and communities.

Another event organised by CHAIN (with several projects too) that has been useful for WP3 has been the 'Conference on the Role of e-Infrastructures for Climate Change Research' (Trieste, May 2011, http://users.ictp.it/~smr2238/). It was a great opportunity for forming from the scratch a new Climate Change VRC with presence in the targeted CHAIN regions. Besides, it was an ideal event where consolidated projects beyond EGI such as Metafor[17], IS-ENES[18], etc. as well as the DCIs (in what Climate Change refers) could present their status.

4.2.2. Information gathered from the survey

In addition to the meetings in which CHAIN participated, another source of information has been the results related to scientific communities provided in the WP2 questionnaire. The reader can find such questions in D3.1[1].

The goal of this part of the survey was to establish the general e-Science status in the region: main scientific fields (or applications) already deployed; main potential/perspective new communities; impact in the region (number of users); and, existence of a funding programme and/or an ongoing VRC Activity. Besides, a request to find researchers interested in the same area that We-NMR, WRF4G and jModelTest & ProtTest communities are working in was made; the reason for doing so was two-fold: on one hand, these communities were solid initiatives which counts on a support team who maintains and updates the applications they offer and, on the other hand, the topics they address are relevant in any part in the world, i.e. the probability of finding users working on the same area is very high.

Once this overall status would be defined, further actions for acquiring a deeper knowledge on a specific topic would be addressed (to be done in the near future after this deliverable or even started as can be checked by the new users that CHAIN is promoting for WeNMR and WRF4G, see subsections 3.3.1 and 3.3.2 respectively). It should be pointing out that most of these results and their related analysis have been included in D4.1[4] and into this document in subsection 4.1.

Nevertheless, these other conclusions can be made:

[17] The Metafor Project, http://metafortrac.badc.rl.ac.uk/trac
[18] The IS-ENES Project. http://www.enes.org/

- the most widely mentioned new community to be deployed is Computational Chemistry (HEP, Life Science and Earth Science being omitted in this analysis)
- other areas of interest are Cultural Heritage, Astrophysics and Mathematics
- There exists applications databases in China and Taiwan (incorporated in EGI one) and those reported in EUAsiaGrid, EUMEDGRID and GISELA initiatives
- Calls promoting e-Science to users are reported, with specific particularities depending on the country, in Algeria, China, Ecuador, Iran, Peru and Taiwan

4.3. Road-map and recommendations

The proposal for a road-map for existing and emerging e-Infrastructures for Virtual Research Communities according to the requirements declared by the latter and the information obtained by the CHAIN project, which has been summarised previously in this document, is itemized below. Recommendations are the core of the proposal.

Nevertheless and above all, beyond VRCs and DCIs, sustainability must be secured by NGIs and higher political structures.

Technical

- Adoption of standards
- Interoperation and interoperability between different e-Infrastructures
    - Grid, HPC, Cloud, Volunteer computing
- Queues for the submission and scheduling of jobs should
    - deal with execution times from a minute to several weeks
        - Jobs submitted with CPU time requirements
    - deal with a single job to hundred of thousands per batch
    - be also ready for interactive jobs
    - smoothly run those multiple queues
- Computing Elements and Working Nodes should provide
    - 2GB RAM/core minimum
    - Working space higher that 10 GB
    - Availability for MPI clusters
    - Increasing request for GP-GPU clusters
- Storage should provide
    - From 1GB to hundreds of TB
    - Short Term storage to long term preservation
    - Backup capabilities
- Data
    - Bandwidth
        - huge data transfers
        - new visualization tools
    - Implementation of useful metadata and search (semantic) engines
    - Files management
        - migration, decommissioning, cleaning…
- Plans for scalability issues
    - Increased demand of computational resources due to the implementation of complex workflows
- Accurate information about the sites should be published as well as their working rules
    - Counting on up-to-date middleware and certificates
- OLA and SLA to be agreed
- Simple mechanisms to deal with troubleshooting and monitoring

- Provision of intellectual property rights (when some data/information demand it) and secure mechanisms for the data transmission

Training, administration and use

- There are still communities and/or groups of researchers who are not sufficiently ICT skilled
- Community-specific training events are an asset
- Implementation of user friendly front-ends for the submission of jobs (scientific gateways, web portals, etc.)
- More accessible authentication and authorisation methods
    - identity providers, for example
- Drawing up of best practice Handbooks, Technical References and Wikis

Collaborative scenario

- promotion of collaborative environments (community building)
    - VRC-VRC
    - VRC-DCI
- Agreement on future strategies
- Dissemination and outreach activities
- Public information about
    - regional and/or national database of applications
    - list of software and services
- Open and periodic calls for promoting and (if possible) funding e-Science

About how DCIs are giving response to the VRC requirements, the reader can take a look for example at the presentations made in the 'Heavy Users Community', 'VRC' and 'Portal technologies' sessions held within the EGI Technical Forum in Lyon (September 2011) and also to the CHAIN & GISELA workshops mostly mentioned in this document, where several DCIs presented their catalogue of services. The analysis of how services are fitting requirements is beyond the scope of this deliverable, so it will be made in future documents once the VRCs had tested them sufficiently.

5. Conclusions

This deliverable shows updated information about the identified VRCs within the CHAIN initiative, i.e. those communities that can profit from CHAIN activities, but could also help in defining what services and requirements are needed in order to improve their research. These are:

- We-NMR;
- WRF4G;
- jModelTest / ProtTest;
- INDICATE / DC-NET;
- DECIDE;
- Climate Change
- SuperB

All of them are trans-continental scientific and technical communities and the group as a whole represents a variety for what concerns the domains represented.

The actions performed by CHAIN in general and WP3 particularly for gathering information about the current status and services provided or required by the DCIs and the VRCs are also described; these have mainly been the organization of joint events and the analysis on the results provided by the WP2 questionnaire delivered at the beginning of the CHAIN lifetime. All of them have played a key role for drawing up the final proposal for a road-map for existing and emerging e-Infrastructures for Virtual Research Communities that is itemized in subsection 4.3