



Co-ordination & Harmonisation of Advanced e-Infrastructures  
Research Infrastructures – Support Action  
Grant Agreement n. 260011  
[www.chain-project.eu](http://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 3.4***  
***“Road-map of trans-continental e-infrastructures for virtual communities Updated”***

Version 1.2

Due on: 31/10/2012

Submitted on: 31/10/2012

Organisation name of lead beneficiary for this deliverable: CIEMAT

Dissemination Level: Public

Abstract:

This deliverable provides an updated plan for a road-map for existing and emerging e-Infrastructures for Virtual Research Communities (VRC). The first version of this road-map can be found in D3.2 ‘Road-map of trans-continental e-infrastructures for virtual communities’, while the current version incorporates the advances and suggestions received from VRCs already defined in D3.3 ‘Transcontinental scientific and technical communities Updated’ as well as the updated CHAIN information collected during the last twelve months.



---

**Copyrights © The CHAIN Consortium 2011-2012.**

More details on the copyright holders can be found at the following URL: <http://www.chain-project.eu>

CHAIN (“Co-ordination & Harmonisation of Advanced e-Infrastructures”) is a project co-funded by the European Union in the framework of the 7<sup>th</sup> FP for Research and Technological Development, as part of the “Capacities specific program - Research Infrastructures FP7-INFRASTRUCTURES-2010-2”. For more information on the project, its partners and contributors please see <http://www.chain-project.eu>

You are permitted to copy and distribute verbatim copies of this document containing this copyright notice, but modifying this document is not allowed. You are permitted to copy this document in whole or in part into other documents if you attach the following reference to the copied elements: "Copyright (C) 2010. The CHAIN Consortium. <http://www.chain-project.eu>".

The information contained in this document represents the views of the CHAIN Consortium as of the date it is published. The CHAIN Consortium does not guarantee that any information contained herein is error-free, or up to date.

THE CHAIN CONSORTIUM MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, BY PUBLISHING THIS DOCUMENT.

---



---

**Document log**

Issue	Date	Comment	Author
V0.1	15/10/2012	First draft	Rafael Mayo (CIEMAT)
V0.2	26/10/2012	Minor changes	Sijin Qian (IHEP) and Rafael Mayo (CIEMAT)
V0.3	29/10/2012	Some typos	M. Ngwira (UbuntuNet Alliance) and Rafael Mayo (CIEMAT)
V1.0	30/10/2012	A few changes in the text	Roberto Barbera and Federico Ruggieri (INFN)
V1.1	26/01/2013	Revision of the text	Rafael Mayo (CIEMAT)
V1.2	05/03/2013	Final version	O. Prjnat (GRNET), R. Mayo (CIEMAT)

## Table of contents

<b>1. Executive Summary</b> .....	<b>4</b>
<b>2. Introduction</b> .....	<b>6</b>
2.1. Purpose .....	7
2.2. Glossary.....	7
<b>3. Direct actions performed to improve and update the road-map of trans-continental e-infrastructures for virtual communities</b> .....	<b>9</b>
3.1. Precedents.....	9
3.2. Follow-up of WP3 milestone ‘MS07 Agreements with reference communities signed’: new MoUs signed .....	10
3.3. VRCs current status and Science Gateway paradigm .....	11
3.3.1. We-NMR .....	11
3.3.2. WRF4G.....	11
3.3.3. jModelTest / ProtTest .....	11
3.3.4. INDICATE / DC-NET .....	12
3.3.5. DECIDE .....	12
3.3.6. LSGC.....	12
3.3.7. Climate Change .....	13
3.3.8. SuperB .....	13
<b>4. Validation of the reference communities and “Quality metrics” assessment</b> .....	<b>14</b>
<b>5. Proposed road-map (2012 updated version)</b> .....	<b>16</b>
5.1. Current status of DCI.....	16
5.2. Road-map and recommendations .....	17
<b>6. Conclusions</b> .....	<b>20</b>

## 1. Executive Summary

During the first year, CHAIN has performed a study of the state of the art in Grid developments across the world regions and their relationship to the European Grid Initiative. Thus, three different deliverables were submitted: D2.2 ‘Interoperability and interoperation guidelines’; D4.1 ‘Specificities of the various regional e-Infrastructures’; and D3.3 ‘Transcontinental scientific and technical communities Updated’. In the present document, a new release of the requirements of the Virtual Research Communities towards the Distributed Computing Infrastructures is presented.

The first step carried out by CHAIN WP3 towards this deliverable has been to update the information provided in D3.1 ‘Trans-continental scientific and technical communities’ into D3.3 ‘Trans-continental scientific and technical communities Updated’ - about the CHAIN collaborative VRCs. These groups/initiatives were selected because of their potential as sustainable communities in the regions targeted by CHAIN and have provided continuous collaboration and feedback. The list of current VRCs which are working with CHAIN is the following:

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

With all of these, except the last two, CHAIN has signed a Memorandum of Understanding, fulfilling and enhancing in this way the WP3 milestone ‘MS07 - Agreements with reference communities signed’. With the Climate Change (improvements on their characteristics for becoming a VRC jointly addressed with the support of ICTP) and SuperB Communities (further contacts with the Grid related body), conversations have been also held. Since the CHAIN project is close to its end, the different initiatives have agreed to sign a MoU with a CHAIN-REDS perspective. CHAIN-REDS will start on Dec 1<sup>st</sup>, 2012.

In addition to the aforementioned actions, several events have been organised by CHAIN for identifying the list of services currently required by these VRCs as well as the status of the e-Infrastructures, due to the latest advances during the last year. The complete list of these events can be found in the CHAIN agenda webpage, available at <http://agenda.ct.infn.it/categoryDisplay.py?categId=80>, but it is worth mentioning the validation of the reference communities that was carried out by means of the interoperability demo performed by CHAIN during the EGI Technical Forum 2012, where the Science Gateway paradigm was demonstrated.

In addition to the information about the VRC requirements that was obtained by means of the CHAIN survey already described in D2.1 ‘State of the art questionnaire’, WP3 has delivered again during 2012 the VRCs related part of the mentioned questionnaire to the National contacts as well as the first version of the road-map of services requested by the VRCs. Doing so, an updated scenario has been reported to the project.

Last but not least, a specific feedback about the list of requirements already published in D3.2 ‘Road-map of trans-continental e-infrastructures for virtual communities’ has been provided by the coordinators of the VRCs collaborating with. Such information is already present in this document.

Regarding the use of additional documentation, it is worth mentioning the information already collected from ‘EGI’s role for 2020’ and ‘Federated Identity Management for Research Collaborations’.

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 5.2 of this document, but above all, beyond VRCs and DCIs, sustainability must be secured by NGIs and higher political structures.

## 2. Introduction

The effort provided by the EC over the last decade in e-Science has been focused not only on Europe, but also on different world regions, mainly Africa, Asia and Latin America. At the same time, different layers of e-Infrastructures, computing platforms and related human power have been targeted by the EC (Research and Education Networks, Grid and HPC infrastructures and Virtual Research Communities).

Due to the economic status in the aforementioned Regions some years ago, Grid emerged as the main actor for supporting e-Science. Thus, the coordination of these world-wide efforts has been promoted, but even in this case, such coordination has been restricted to basic operational, organisational and technological know-how transfer/exchange. No big advances have been made until now to fully link e-Infrastructures at intercontinental level due to their specific requirements which depend on the targeted region.

The CHAIN project, started on December 1<sup>st</sup> 2010, has been working on coordinating and leveraging 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 – those of core interest to EC. Now that the project is coming to its end, it can be mentioned that this vision has been 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.

Thus, concrete results are available to the community by means of the public CHAIN deliverables D2.2 ‘Interoperability and interoperation guidelines’<sup>1</sup>, D2.3 ‘Regional sustainability report’<sup>2</sup>, D4.1 ‘Specificities of the various regional e-Infrastructures’<sup>3</sup>, D4.2 ‘Report on the applicability of the preliminary results of the Organisational Study’<sup>4</sup> and D4.3 ‘Organisational study for interoperations of EGI with external grid infrastructures’<sup>5</sup>. In these deliverables, an analysis of the questionnaire data launched by WP2 is made, specificities of the various regional e-Infrastructures and their sustainability are described, and recommendations for regional organisational and operational interactions are proposed.

CHAIN WP3, ‘Present and emerging needs of trans-continental scientific communities’, focuses on 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. This work was summarised in D3.1 ‘Trans-continental scientific and technical communities’<sup>6</sup> and its follow-up D3.3 ‘Transcontinental scientific and technical communities Updated’<sup>7</sup>, which focused on the description of the results related to VRCs obtained up to the sixth (May 2011) and to the eighteenth (May 2012) month of the project respectively. In both documents, the communities which are collaborating with CHAIN are described.

In a similar way to that followed by WP2 and WP4 in their deliverables, a study on the requirements and services of interest for the VRCs that had to be offered by the infrastructure

---

<sup>1</sup> CHAIN D2.2, [http://documents.ct.infn.it/record/506/files/CHAIN\\_D2.2\\_V1.8.pdf](http://documents.ct.infn.it/record/506/files/CHAIN_D2.2_V1.8.pdf)

<sup>2</sup> CHAIN D2.3, <http://www.chain-project.eu/status>

<sup>3</sup> CHAIN D4.1, [http://documents.ct.infn.it/record/499/files/CHAIN\\_D4.1%20V1.3.pdf](http://documents.ct.infn.it/record/499/files/CHAIN_D4.1%20V1.3.pdf)

<sup>4</sup> CHAIN D4.2, [http://documents.ct.infn.it/record/515/files/CHAIN\\_D4.2%20V1.pdf](http://documents.ct.infn.it/record/515/files/CHAIN_D4.2%20V1.pdf)

<sup>5</sup> CHAIN D4.3, <http://www.chain-project.eu/status>

<sup>6</sup> CHAIN D3.1, <http://documents.ct.infn.it/record/500/files/CHAIN-D3.1-V06.pdf>

<sup>7</sup> CHAIN D3.3, <http://documents.ct.infn.it/record/525/files/CHAIN%20D3.3-V05.pdf>

providers was documented in D3.2 ‘Road-map of trans-continental e-infrastructures for virtual communities’<sup>8</sup>. Such a road-map, as it previously was the case with D3.1, is now updated and improved in a second study, which has produced this document: D3.4 ‘Road-map of trans-continental e-infrastructures for virtual communities Updated’.

## 2.1. Purpose

The aim of this deliverable is to update and improve the proposal for the road-map for existing and emerging e-Infrastructures for Virtual Research Communities which was originally made one year ago. These new guidelines have been mainly obtained from the continuous liaison with the scientific and technical communities that have official collaborations with CHAIN, from the contacts already established with other different communities in events organised by the project (and events with project participation), and from a new reduced survey delivered by WP3 in 2012. At the same time, actions on forming new communities have been also done during this year.

Specifically, the actions that have been carried out during the latest twelve months of the project in order to obtain the necessary information have been:

- To deliver again the specific part related to VRCs of the WP2 survey to the National representatives.
- To get the feedback about the first version of the road-map of services from the National representatives.
- To maintain a continuous liaison with the identified VRCs fostering the collaboration by periodic contacts, which also include the reception of their feedback about the first version of the road-map. In this way such collaboration has been consolidated by means of new MoUs.
- Two specific workshops have been organised by CHAIN and held within EGI Community Forum 2012 and EGI Technical Forum 2012 in order to know the requirements and necessities that VRCs have and present them, at the same time, the first version of the road-map.
- In addition, CHAIN has participated throughout 2012 in several events where it has also presented the analysis of the status of the e-Infrastructures and the scientific communities, and has been provided with feedback.
- CHAIN has supported the creation of new VRCs.

This deliverable describes all these activities and the resulting updated 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

<sup>8</sup> CHAIN D3.2, <http://documents.ct.infn.it/record/525/files/CHAIN%20D3.3-V05.pdf>

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. Direct actions performed to improve and update the road-map of trans-continental e-infrastructures for virtual communities

In the following section, the actions carried out through the last year to update the road-map of services will be described. In addition, there will be a short summary about the VRCs' status and the developments that have been made for demonstrating interoperability between regional infrastructures.

#### 3.1. Precedents

At the beginning of CHAIN, 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 specificities. These surveys were opened in April 2011 and their analysis is available in D2.2 'Interoperability and interoperation guidelines'<sup>1</sup> and D4.1 'Specificities of the various regional e-Infrastructures'<sup>3</sup>. Both surveys contained a specific section where questions devoted to VRC topics were present; this list can be found in D3.1 'Trans-continental scientific and technical communities'<sup>6</sup>.

During the second year of the project, the VRC-related part of the National survey was circulated again in order to get updated information about the advances made on scientific aspects in the Regions of interest to CHAIN. In addition to the questionnaire submitted to the National representatives, the link to D3.2 'Road-map of trans-continental e-infrastructures for virtual communities'<sup>8</sup> was included in the text of the e-mail and the section containing the specific information about the road-map was attached. This has been the primary source of information for updating and improving the first version of the road-map of services requested by the VRCs.

This first version has also been sent to the coordinators of all the VRCs that collaborate with CHAIN. These actors supported CHAIN by submitting back comments and suggestions about the road-map bearing in mind their experience in their specific domains. This second source of information has improved the quality of the analysis.

Besides, and as it was done in the first year of the project, Distributed Computing Infrastructure providers and scientific communities have met several times by means of the events organised by CHAIN. It is worth mentioning the two workshops organized as part of EGI CF 2012 and EGI TF 2012 which were specifically devoted to this objective. These events are 'VRCs on EGI & Regional Infrastructures'<sup>9</sup>, which was held in March 2012 in Garching (Germany), and 'CHAIN Interoperability Workshop'<sup>10</sup>, which was held in September 2012 in Prague (Czech Republic). Both events have been a great opportunity for sharing again experiences between DCIs and VRCs and have been very useful for updating the road map that is presented in this document, but it is important to note that in the latter, an interoperability demo was delivered by CHAIN (see Section 4 for further details).

Also, a major event during this year has been the organization of the Joint GISELA-CHAIN Conference (<http://indico.ceta-ciemat.es//conferenceDisplay.py?confId=26>), which was held in Mexico City in June 2012. Several sessions were organized and fruitful discussions produced results that are used to draw up this document.

---

<sup>9</sup> <https://indico.egi.eu/indico/sessionDisplay.py?sessionId=74&confId=679#20120329>),

<sup>10</sup> <https://indico.egi.eu/indico/sessionDisplay.py?sessionId=62&confId=1019#20120919>

CHAIN has participated in other events in 2012 and most of them have had a focus on VRCs matters. Thus, the CHAIN Workshop on "Sustainability & Interoperability - Step 2" held as part of ISGC 2012 in Taipei in February 2012, the CHAIN Workshop at TICAL 2012 held in Lima (Peru) in July 2012, the CHAIN Workshop at ICTK 2012 held in Bangalore (India) in July 2012 and the CHAIN Workshop at IEEE Cluster 2012 held in Beijing (China) in September 2012 have been appropriate forums to discuss the analysis of the road-map of services first produced by CHAIN at the end of 2011. Specific information about these events can be found at the CHAIN agenda webpage <http://agenda.ct.infn.it/categoryDisplay.py?categId=80>.

Last but not least, as this document is being submitted to the EC, two last CHAIN events are being held in November 2012: one, in Mumbai (India) as part of the NKN conference (<http://nkn.in/nkn-workshop2012/index.php>) and, second, the UbuntuNetConnect Conference 2012 ([http://www.ubuntunet.net/uc\\_2012](http://www.ubuntunet.net/uc_2012)) to be held in Dar Es Salaam (Tanzania). Detailed information about these as well as the CHAIN WP3 presentations can be found in the previous CHAIN agenda webpage link.

At this point, it is worth mentioning the work done by CHAIN WP5 by implementing in the webpage of the project the 'Knowledge base' and 'Applications' sections. These web pages are very useful for providing information not only to the rest of the project activities, but also to the general public. At the same time, they have been continuously updated with the information provided by the activities carried out by WP3 during 2012.

Regarding bibliographic information search, two documents have been of crucial importance to draw up this deliverable: 'EGI's role for 2020' and 'Federated Identity Management for Research Collaborations'.

### **3.2. Follow-up of WP3 milestone 'MS07 Agreements with reference communities signed': new MoUs signed**

As it can be found in D3.3 'Trans-continental scientific and technical communities Updated'<sup>7</sup>, once the third and last WP3 milestone 'MS07 Agreements with reference communities signed' was fulfilled by the project in September 2011 by signing two (2) MoUs with We-NMR and WRF4G initiatives, the Work Package has been working on improving the quality of the agreed collaborations and has signed until the sixth quarter four (4) more MoUs with jModelTest and LSGC initiatives and with the INDICATE and the DECIDE projects.

During the last semester of the project, there has been an effort for forming a VRC with the Climate Change community as a continuation of the collaboration with WRF4G. The new action has been to contact the Climate Change group of ICTP<sup>11</sup> and to establish a direct collaboration with its associated groups worldwide by the common use of RegCM4 model<sup>12</sup>. As a result, a MoU has been proposed and its final signature will be produced in the next weeks. Since CHAIN is coming to its end and trying not to leave the MoU meaningless, a closer collaboration is foreseen during CHAIN-REDS project (to start on December 1<sup>st</sup>, 2012.)

Same applies to SuperB consortium. Direct contacts with the Grid related body have been taken too, and a MoU draft has been submitted to the collaboration, which has discussed it

---

<sup>11</sup> The ICTP Climate Change group, <http://www.ictp.it/research/esp/research/climate-change.aspx>

<sup>12</sup> The RegCM4 model, <http://www.ictp.it/research/esp/models/regcm4.aspx>

internally within its governing body. Also, the document is focused on a closer collaboration with the CHAIN-REDS project.

The whole list of CHAIN MoUs is available at <http://www.chain-project.eu/news>

### **3.3. VRCs current status and Science Gateway paradigm**

In the following subsection, information about the new advances regarding the originally identified VRCs is 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.

No major advances have been produced since the last information already presented in D3.3<sup>7</sup>. Nevertheless, since the initiative will not have a second phase, CHAIN has agreed to keep providing as much dissemination as possible of the WeNMR infrastructure during CHAIN-REDS.

WeNMR coordinator has provided valuable feedback about the first draft of the updated road-map of services that appears in this deliverable.

#### **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.

The main action carried out during the last months since the submission of D3.3<sup>7</sup> has been the work carried out on developing a SAGA-GridWay interoperability actor so WRF4G and CAM4G could be integrated in the Science Gateway, but the task is hard and no results can be described to the moment.

WRF4G coordinator has provided valuable feedback about the first draft of the updated road-map of services that appears in this deliverable.

#### **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.

From the information provided in D3.3<sup>7</sup>, the most important remark is that both applications can be run on the SG by now (<http://gisela-gw.ct.infn.it/jmodeltest> and <http://gisela-gw.ct.infn.it/prottest>) and have participated as part of the interoperability demo carried out by CHAIN during the EGI Technical Forum 2012. For the near future, improvements on these on-line access versions of the applications are expected in order to enhance their bioinformatics potential, i.e. the number of models that can be executed.

As in the previous communities, a feedback for the road-map of services from the initiative coordinator has been also received.

#### **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 basis.

No major advances have been produced since the last information already presented in D3.3<sup>7</sup>. Nevertheless, by means of the WP3 second release of the survey, new contacts have been given to the initiative in China.

INDICATE coordinator has provided valuable feedback about the first draft of the updated road-map of services that appears in this deliverable, which is of importance because of the different scientific point of view.

#### **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<sup>13</sup> and relying on the Pan-European backbone GEANT<sup>14</sup> and the NRENs. 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.

No major advances have been produced since the last information already presented in D3.3<sup>7</sup>. Nevertheless, by means of the WP3 second release of the survey, new contacts have been given to the initiative in Nigeria.

INDICATE coordinator has also provided valuable feedback about the first draft of the updated road-map of services that appears in this deliverable.

#### **3.3.6. LSGC**

LSGC community is one of the biggest in Grid computing excluding HEP. Its presence is already settled in every region and now is currently running on a best-effort basis. Since LSGC is mature enough, and in accordance to CHAIN aims and objectives, the CHAIN effort on deploying it is more reduced in comparison with other communities: nevertheless, feedbacks from the initiative have been used in the updated road-map of services.

---

<sup>13</sup> NeuGRID project, <http://www.neugrid.eu/>

<sup>14</sup> GEANT network, <http://www.geant2.net/>

### 3.3.7. Climate Change

Climate Change community and its final creation have been one of the most important efforts in WP3. The reader can refer to D3.3<sup>7</sup> for a deeper explanation about the different steps that have been taken so far.

As a result and due to range of scientific fields that can be involved under the Climate Change research, CHAIN has agreed to sign a MoU with the ICTP in Trieste in order to articulate, by means of its current collaborations, a community that could be directly contacted.

Researchers belonging to every region of interest to CHAIN are already linked to ICTP and, on the other hand, all the previous contacts that CHAIN had in what regards this community have been given to the ICTP Climate Change representatives, widening in this way its potential impact.

Since the MoU is going to be signed soon, information about CHAIN-REDS and its further activities has been included in order to count on this agreement as a first action belonging to CHAIN-REDS project.

### 3.3.8. SuperB

The SuperB flavour factory (<http://superb.infn.it/home>) is a major international research centre for fundamental and applied physics to be built on the campus of the University of Rome Tor Vergata. It is an accelerator that will provide complementary information to LHC, looking for rare decays with a very high luminosity electron-positron asymmetric collider.

From the information appearing in D3.3<sup>7</sup>, a new MoU has been agreed to be signed with this initiative in the near future once the community agrees on the final version within its governing body. As in the previous case, foreseen follow-up in CHAIN-REDS is already described in the document.

#### 4. Validation of the reference communities and “Quality metrics” assessment

As a proof of concept, CHAIN has promoted the Science Gateway paradigm on a two-fold basis: on one hand, because it allows demonstrating that any user can access a wide plethora of different infrastructures in a easy and unattended way; and, on the other hand, because its authentication and authorization method does not rely on Grid personal certificates, but Identity Providers and, in some cases, even social networks.

A description of the Science Gateways architecture can be found in D3.3<sup>7</sup> too. Such development work has resulted in the interoperability demo that has been carried out during the EGI TF 2012 in Prague in September 2012. Such a demo has been shown permanently in the CHAIN boot, but also during the organised CHAIN workshop (<https://indico.egi.eu/indico/sessionDisplay.py?sessionId=62&confId=1019#20120919>). As every CHAIN event, the slides presented during this workshop can be found in the project Agenda at <http://agenda.ct.infn.it/categoryDisplay.py?categId=79>.

Since the final WP3 deliverable is still due on Nov 2012, i.e. D3.5 ‘Results of the validation of the reference communities’, the final description of this demo as well as the benefits derived from the CHAIN collaboration with the different VRCs will be described in that document. Thus, the reader should consider this Section as preliminary news related to the demo, which certifies that it was executed satisfactorily.

As part of the validation model, the Quality Metrics<sup>15</sup> assessment is also crucial. With regards to WP3, this is the status as that of October 2012:

Project Key Indicator	Metric	Target (Nov 2012)	Current status (Oct 2012)
PKI WP3.1	Number of VRCs officially supported (MoUs)	2	6 (WeNMR, WRF4G, jMT, LSGC, INDICATE, DECIDE)
PKI WP3.2	Number of applications deployed on the participating regional e-Infrastructures	2 (1 per scientific discipline)	15 (ProfTest, Octave, jModelTest, Astra, Specfem3d, RCSA, R, Me-mls, Industry@grid, Gromacs, Gate, Ga_dppm, Clustalw, Phylogenetics, Cmsquares)
PKI WP3.3	Number of new applications with respect to those already deployed in the Regional projects	1	3 (ProfTest, jModelTest, Gromacs)

**Table I. Quality metrics related to WP3 in the CHAIN DoW. Explanation about the current status information can be found through the text.**

It is important to point out that the applications appearing in Table I are final and definitive ones since they have already been integrated in the SG paradigm. In addition, it is also worth mentioning that the figures appearing in the third column, ‘Current status (Oct 2012)’ have been extracted from the CHAIN applications webpage (<http://www.chain->

<sup>15</sup> CHAIN Quality Metrics, <http://www.chain-project.eu/quality-metrics>

[project.eu/applications](http://project.eu/applications)); those related to PKI WP3.2 are the codes deployed on the participating regional e-Infrastructures and those regarding PKI WP3.3 are the applications that are available through the CHAIN SG portal.

## 5. Proposed road-map (2012 updated version)

In this section, the second analysis devoted to draw up a road-map for existing and emerging e-Infrastructures for VRCs is made. In order to better define the scenario in which such proposal will be made, a brief summary of the current status of the DCIs, as that previously presented in D3.2<sup>8</sup>, will be done.

### 5.1. 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 firstly developed in CHAIN by means of D2.2 ‘Interoperability and interoperation guidelines’<sup>1</sup> and, D4.1 ‘Specificities of the various regional e-Infrastructures’<sup>3</sup>. At the same time that this deliverable is being submitted, updated versions of both documents have been delivered, i.e. D2.3 ‘Regional sustainability report’<sup>2</sup> and D4.3 ‘Organisational study for interoperations of EGI with external grid infrastructures’<sup>5</sup>.

Regarding interoperation and interoperability and taking into account the current status, the following points can be highlighted:

- Establishment of consolidated domestic NGIs, supported by the in-depth documentation (that could be published through CHAIN website and circulated among the NGIs in CHAIN target regions);
- To form VRCs that would group the related researchers of a common field and their liaison with similar communities in Europe and the other regions avoiding duplicated efforts.
- To use the VRCs as a key point for the use of the infrastructure and the best test for demonstrating its sustainability;
- To progress from JRU model on regional level, 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 so;
- 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 support for:
  - 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

About e-Infrastructures, their status has been improved from one year ago, and the main aspects have been kept.

In this way, the most important advance that has been produced is the availability of the Science Gateway paradigm. Prior to it, it was needed to count on a strong group of scientists and developers who would port every application to the Grid or any particular platform and then would train other users in employing it for research. Nowadays, final users only have to take care of the scientific problem they want to solve and use the SG web portal in an unattended way. No personal Grid certificates and no command lines interfaces are needed, only an identification by a certified provider for accessing the web and a web system to upload their input data before running the code. A reduced group of Java and SAGA developers adapt the previous codes to the SG paradigm and make the new version available to the whole community.

## 5.2. 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. Some of these items are already implemented (Oct 2012) but they are listed here anyway because of their importance. Recommendations are the core of the proposal.

Nevertheless and above all, beyond VRCs and DCIs, sustainability of NGIs and higher political structures is important.

### Technical

- Adoption of standards
- Interoperation and interoperability between different e-Infrastructures
  - Grid, HPC, Cloud, Volunteer computing
    - Specific capabilities associated to job submission, credentials, etc.
      - Homogenization of access to heterogeneous DCI for users
    - Initiatives devoted to such an objective (StratusLab, JSAGA, etc)
  - Easy access to the computing infrastructures by means of Identity Providers
    - National ones
    - Catch-all community
- 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
    - Recovery system: resubmission and check pointing
  - Deal with a single job to hundred of thousands per batch
  - Readiness for interactive jobs
  - Job dependency
    - Job events: i.e. 10%, 20% running completion
    - Pre stage and post stage requirements
    - Storage provision
    - Environment preparation
  - Smoothly run those multiple queues
  - Some service (meta-scheduler, pilot job systems) to dispatch jobs among sites
- Computing Elements, Working Nodes and sites should provide
  - 2GB RAM/core minimum
  - Working space larger than 10 GB
    - Shared and non-shared scratch space
  - Availability for MPI clusters

- Access to MPI process management system
  - Increasing request for GP-GPU clusters
  - Remote software management/installation
  - At least 8 job slots.
  - A recommendation for a SE where to store data produced on this site (e.g. in the form of an env variable)
  - A reaction to tickets in less than 10 days
- Storage should provide
  - At least 1 TB (for some VRC this is a minimum)
  - Short Term storage to long term preservation
  - Hierarchical storage management and homogenization
  - Backup capabilities
  - Tools to verify data integrity (e.g., compute checksums on-the-fly and compare with values stored in separate partition)
  - Data cleanup capabilities
- Data
  - Bandwidth
    - Huge data transfers (maximizing use of the current links of 10 Gbps)
    - New visualization tools (VisIt, MayaVi, etc.)
  - Implementation of metadata and search (semantic) engines
  - Files management
    - Collection of files/folders
    - Partial access to datasets
    - Migration, decommissioning, cleaning.
  - Support for server-side encryption (to protect against other users having access to same storage)
  - Support for client-side encryption:
    - Setup of a service to securely store decryption key
    - Setup of a mechanism to define Access Control Lists of subjects/groups authorized to retrieve the decryption key
- Plans for scalability issues
  - Increased demand of computational resources due to the implementation of complex workflows
  - Support for large scale experiments with millions of tasks and dependencies among them
- Accurate information about the sites should be published
  - Counting on up-to-date middleware and certificates
  - Working rules
  - Technical capabilities and characteristics that are supported
- OLA and SLA to be agreed among interested actors
  - In some cases, best-effort agreements are enough
- Simple mechanisms to deal with troubleshooting and monitoring (GGUS)
- 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

- Sustainability of the VO management is a key point due to its cost
- Implementation of user friendly front-ends for the submission of jobs (Science Gateway, gUse, etc.)
- Implementation of library packages and services to access DCI for application developers
- More accessible authentication and authorisation methods
  - Identity providers
  - Robot certificates
  - User's identity repository and management
  - Support to users for obtaining personal certificates
- Drawing up of best practice Handbooks, Technical References and Wikis
- Infrastructures more user community driven
- Agile tools for VO administration, easily allowing creation of new groups and roles
  - Support for multi-level administration, allowing to appoint sub-tree administrators

#### Collaborative scenario

- Promotion of collaborative environments (community building)
  - VRC - VRC
  - VRC - DCI
  - VRC&DCI – Public and Private (funding) agencies
- Agreement on future strategies (objectives, main advances, developments)
- Dissemination and outreach activities
- Public information about applications data base
  - regional and/or national database of applications
  - list of software and services
- Open and periodic calls for promoting and (if possible) funding e-Science

## 6. Conclusions

This deliverable presents 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;
- LSGC;
- 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 work with them has been important for achieving the Quality Metrics described in the CHAIN DoW, but even more, has been crucial for the interoperability demo carried out by CHAIN during the EGI Technical Forum 2012 in which the Science Gateway paradigm was used.

The actions performed by CHAIN in general, and WP3 in particular, 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 of the results provided by the WP2 and WP4 deliverables. 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 5.2.