

The CHAIN Project

Federico Ruggieri^{1,2}, Angelines Alberto³, Giuseppe Andronico¹, Roberto Barbera^{1,4}, Jorge Blanco-Yagüe³, Gang Chen⁵, Shri P.S. Dhekne⁶, Guillermo Díaz³, Salma Jalife⁷, Kostas Koumantaros⁸, Aleš Krenek⁹, Alberto Masoni¹, Luděk Matyska⁹, Rafael Mayo^{**3}, Tiwonge Msulira-Banda¹⁰, Raquel Muñoz³, Sinha Neeraj⁶, Maragaret Ngwira¹⁰, Luis Núñez⁷, Marco Paganoni^{1,11}, Ognjen Prnjat⁸, Manuel Rodríguez-Pascual³, Antonio J. Rubio-Montero³ and Stefano Troiani³

(On behalf of the CHAIN project)

¹ Istituto Nazionale di Fisica Nucleare (Italy)

² Università degli Studi Roma Tre (Italy)

³ Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (Spain)

⁴ Università di Catania (Italy)

⁵ Institute of High Energy Physics – Chinese Academy of Sciences (CHINA)

⁶ Office of the Principal Scientific Adviser to the Government of India (India)

⁷ Cooperación Latino Americana de Redes Avanzadas (Uruguay)

⁸ Greek Research and Technology Network (Greece)

⁹ CESNET (Czech Republic)

¹⁰ UbuntuNet Alliance (Malawi)

¹¹ Università degli Studi di Milano-Bicocca (Italy)

Abstract. The European Commission has focused in the latest years its efforts in Science development on promoting the research on computational Science. This is so because of the increasing importance of simulating phenomena in almost every field of interest. In particular, a huge investment has been done in Distributed Computational Infrastructures in Europe, but also in other continents such as Latin America, Asia and Africa, mainly in Grid. Thus, different phases for infrastructures projects in all these continents from 2006 to 2011 have produced excellent results and have promoted the use and research on Grid computing. The CHAIN project aims to coordinate and leverage these efforts and 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.

1 Introduction

Over the past 6 years, the European Commission has invested to extend the European e-Infrastructure technology and European e-Infrastructure (and particularly Grid) operational and organisational principles to a number of regions in the world, and reinforcing the close collaboration and exchange of know-how with similar technologies in other areas.

** rafael.mayo@ciemat.es

Thus, although big steps have been made to extend the European Grid principles to other regions, the results obtained so far have to be leveraged and customised so as to provide an overall model for sustainable interoperation between the European Grid Initiative/Infrastructure and external e-Infrastructures. Furthermore, following the evolution path of Grid collaborations between Europe and the rest of the world, it is deemed beneficial to also analyse the issues and models related to coordination between European and external regional e-Infrastructures taking into account not only Grids but also HPC, Clouds, Voluntary computing infrastructures and networking issues

On the technology level, the coordination of different world-wide Grid efforts, listed below, has been restricted to basic operational, organisation and technology know-how transfer/exchange. Moreover, the co-ordination of Grids and HPC has been to some extent neglected, although there is currently a pressure to arrive to some agreement and to a more organised coexistence of the two approaches to scientific computing. Finally, the recent upsurge of other paradigms should be evaluated in the light of a large world-wide e-Infrastructure; these would be the cases of Virtualisation, Voluntary Computing or Cloud Computing.

The extension of European e-Infrastructure to other regions of the world developed on three main tracks:

- Research and Education Networks (i.e. GEANT, ALICE, TEIN, EUMEDCONNECT, etc.);
- Grid Computing (i.e. EGEE, BalticGrid, EELA, EUAsiaGrid, EUChinaGRID, EUIndiaGrid, EUMEDGRID, SEEGRID); and,
- Virtual Research Communities (Health, Biomedical, HEP, Earth Sciences, etc.).

Very little has been done until now to link e-Infrastructures, at intercontinental level, due to their specific requirements.

As aforementioned, a number of different collaboration models have then been established between Europe and the rest of the world, while the projects implementing these collaborations have had impacts typically focused on their regions. The Co-ordination & Harmonisation of Advanced e-Infrastructures project (CHAIN, <http://www.chain-project.eu/>), a Coordination and Support Action with a budget around 1.5 M€, now receives this legacy and intends to coordinate the work carried out by these regional projects as well as optimise an interaction model for e-Infrastructures and specifically Grid interfaces between Europe and the rest of the world.

The project is working on elaborating a strategy and defining the instruments in order to ensure coordination and interoperation of the European Grid Infrastructures with other external e-Infrastructures. This way, the CHAIN consortium, coordinated by INFN and consisting of leading organisations in all the regions addressed by the project, will ensure global coverage, European leadership and most efficient leveraging of results with respect to preceding regional initiatives. In concrete, CHAIN counts on 4 European institutions (INFN, CESNET, CIEMAT and GRNET) and other 4 are from countries in the list of International Co-operation Partner Countries (CLARA, IHEP, PSA, UBUNTUNET).

The different steps to be taken are as follows. First, the project will define a coherent operational and organisational model, where a number of EU countries/regions will possibly act, in collaboration with EGI (<http://www.egi.eu/>), as bridges/gateways to other Regions/Continents. Further, the project will validate this model by supporting the extension and consolidation of worldwide virtual communities, which increasingly require distributed facilities (large instruments, distributed data and databases, digital repositories, etc.) across the regions for trans-continental research. Finally, the project will act as a worldwide policy-watch and coordination instrument, by exploring and proposing concrete steps for the coordination with other initiatives and studying the evolution of e-Infrastructures.

To do so, it is of utmost importance the experience gained through the past by the CHAIN partners, who have participated in GRID project such as the different EGEE phases or the "regional" associated initiatives in Latin America, Asia and or Africa. At the same time, all the European partners also collaborates in the European Grid and HPC Initiatives and its related National Initiatives, so they are up-to-date of the different steps that e-Science is performing.

2 WP1 Project Management

As usual, this WP performs a mandatory activity for ensuring the correct and timely execution of the project. The CHAIN project appears to be challenging from the management point of view due to the large number of regional partners and third parties involved. Moreover, the project has to organize and develop several high level activities and monitor their impact.

All these aspects will be addressed, at the management level, by a correctly articulated structure which will avoid overloads and unnecessary bureaucracy. Moreover, the large geographical coverage of the project poses several challenges, while the work-plan has activity lines, derived from the project's objectives, which span all the regions involved. The managerial structure has thus been designed with a traditional technical management that develops horizontally across the regions and an executive management that vertically addresses the different regions taking into consideration their specificities.

As stated above, INFN leads the project and all the activities which are strictly related to the overall technical and administrative management of the project have been distributed in several tasks. Managerial activities which are related to the other work-packages have been included in the respective work plans. This way, the governance structure is then composed by the following roles and boards:

- A Project Director (PD) nominated by the Coordinator (INFN) to run the project;
- A Project Management Board (PMB): one member for each partner plus a deputy, to specifically address strategic and contractual decisions concerning the project;
- A Regional Executive Board (REB): one deputy Project Director for each Region, to act as an executive board that will be entitled to take urgent decisions between the PMB Meetings. The members of the REB will act as Deputy PDs for the region concerned;

- A Project Office (PO): 3 persons, to deal with administrative and managerial activities and support the PD;
- A Technical Board (TB): made up by the managers of all the work-packages, to take care of the technical management of the project;
- A Technical Manager (TM) to chair the TB and deal with the day-to-day technical discussions ensuring the coherence of all the technical actions in line with the project’s objectives.

WP1 is divided in two Tasks: ‘T1.1 Administrative management’ for dealing with the administrative, financial and overall management of the consortium; and, ‘T1.2 Technical management’, whose aim is to coordinate the overall technical activities, for the achievement of milestones and the delivery of the deliverables according to the foreseen work plan. Typical milestones of this WP are the organization of the project Kick-off meeting and the annual reviews at the European Commission.

3 WP2 Consolidation of existing state of the art

The maturity of non-European e-Infrastructures varies across the world regions, where 3 typical scenarios could be identified when CHAIN proposal was submitted (2010):

- Completely green-field regions which need to be supported from scratch (e.g. sub-Saharan Africa with the exception of South Africa)
- Stable (if small) regional infrastructure established and interoperational with Europe (e.g. Mediterranean, some LA & Caribbean countries, SEE)
- Advanced countries/regions which have autonomously invested in several e-Infrastructures and are willing to interoperate with European ones (e.g. China, India, New Zealand, USA, Japan, some LA countries, etc.).

A set of EU-funded activities have been so far investigating different modes of interoperation and interoperability between Europe and the rest of the world. While strong information exchange has been taking place, there is still a diversity of solutions in place with a wealth of know-how available.

This work-package is working on systemising the current know-how and producing a coherent and structured consolidated picture of the state of the art in the different regions, highlighting commonalities and differences. On the basis of this analysis, a set of guidelines will be produced to guide the interoperability and interoperation among different types of regions. To fulfil this goal, reference will be made to the recommendations developed from international standardization forum as Grid Interoperation Now, part of Open Grid Forum, which is working extensively on this point.

WP2 is leaded by GRNET and its first action has been to elaborate, jointly with the rest of Work packages, a questionnaire with two different versions depending if it is delivered to Regional or National contacts. In both versions, questions related to Infrastructure, Operational and User support services, User communities, EGI

interoperation approaches, Middleware, Network, Related technologies (cloud, for example) and Sustainability appear.

WP2 is structured in two Tasks. ‘T2.1 State of the art analysis’ deals with the in-depth analysis of the current state of the art across the regions and is the main responsible for the questionnaire, which will produce data collection and will gather and systemise the data. The activity of the Task is devoted to gather the accumulated experience and the current best practices in the existing e-Infrastructures with specific focus on the basic and advanced services provided, the scientific communities addressed, the security policies, the sustainability opportunities and plans and the current interoperation level with the European Grid Infrastructure.

Task ‘T2.2 Interoperation and interoperability guidelines’ will analyse the data collected in T2.1, on the basis of which a set of guidelines for interoperability and interoperation will be produced. The data will be analysed under the various topics and a comparison among the different approaches will be done focusing, on one hand, on the areas of possible synergies and harmonisation and, on the other hand, on the specific region-dependant solutions and differences evaluating their impact on the level of interoperations and interoperability with other e-Infrastructures.

Lately, Task ‘T2.3 Long-term sustainability support’ will disseminate the accumulated know-how on NGI and regional sustainability issues and will carry out actions on strengthening National Grid Initiatives and regional structures. The task will examine the current and emerging plans of sustainability, enumerate the opportunities that could influence these and future plans and eventually propose measures of improvement based on the current best practices and new opportunities.

The milestones reflected in this activity are focused on providing the questionnaire to find out what the different regions need as well as produce valid guidelines with the obtained results.

4 WP3 Present and emerging needs of trans-continental scientific communities

There are several communities which are trans-continental by constituency, such as LHC experiments (<http://public.web.cern.ch/public/en/LHC/LHC-en.html>), European Organisation for Astronomical Research (<http://www.eso.org/>), Earth Science, Biomedical applications for neglected and emerging diseases, etc. An inclusive landscape of the virtual research communities accessing or willing to access trans-continental common developments, distributed data and facilities constitute the main motivation for coordination, interoperation and interoperability of e-Infrastructures across several regions, i.e. it also means the CHAIN objective. Some institutions participating in the CHAIN project are already involved in several of the cited communities and are obviously interested in continuing to support them, but this work-package is aiming to go beyond the “legacy communities” and intends to address new fields that have not completely exploited yet the opportunities of such large intercontinental e-Infrastructures.

Thus, the work-package aims at

- continuing to provide (limited) support to the existing and well experienced Virtual Research Communities such as LHC, Biomedicine, Earth Science, etc.;
- providing support for communities that have already approached the Grid technology but are interested to widen their activities collaborating with other continents (they could be the previous ones, but with new applications); and,
- discovering and attracting new scientific communities or merging similar ones actually operating within separate e-Infrastructures worldwide.

Under these premises, WP3 has a main objective: to involve at least a couple of reference communities that will be selected and involved in the project to validate the proposed model jointly developed by the whole project.

To the date, some of these communities could be some of the ones mentioned above, but in any case, the CHAIN project is collaborating with several specific ones:

- We-NMR initiative (<http://www.wenmr.eu/>) in the Biomedical field;
- DECIDE initiative (<http://www.eu-decide.eu/>) in the Biomedical field;
- Phylogenetics applications (<http://darwin.uvigo.es/>) developed at Universidad de Vigo in the Biomedical field;
- INDICATE project (<http://www.indicate-project.org/>) in the Digital Cultural Heritage field; and,
- WRF4G (<http://www.meteo.unican.es/es/software/wrf4g>) in the Earth Science one.

With regard to this, both initiatives already appear in the WP2 questionnaire inside the ‘National’ version so CHAIN is looking for specific contacts of people working on these topics, which will be contacted lately jointly with the We-NMR and WRF4G representatives. Working this way, WP3 expects to fulfill its own milestones, i.e. obtain a shortlist of reference communities from a first call and the questionnaire that will conduct to a study on their required services as well as some Memoranda of Understanding in order to collaborate closely.

WP3 is coordinated by CIEMAT and is divided in ‘T3.1 Scientific communities across the continents’, which is on charge of performing a large spectrum investigation on the existing and potential trans-continental communities, so the requirements, commonalities, challenges and possible synergies could be gathered and analysed, and ‘T3.2 Proposed road-map of services for communities to be deployed on the e-Infrastructures’, the aim of which is to produce a study, updated every 6 months, of the necessary steps for e-Infrastructures to fulfil the requirement of present and emerging virtual communities with trans-continental span. To do so, information will be collected from the existing plans of relevant organisations and/or committees (e.g. ESFRI, e-IRG), the preliminary results of T3.1, the feedbacks received during the workshops and high-level conferences that CHAIN could attend or organise and the collaboration with other relevant projects.

5 WP4 Modelling the cooperation of European e-Infrastructures with non-European ones

On the basis of the results of WP2 and WP3, and in continuous cooperation with them, a model for cooperation and interoperability among European and non-European e-Infrastructures should be studied and proposed; this is the task of WP4, led by CESNET. In addition, a strong collaboration with EGI and EGI-Inspire (<http://www.egi.eu/projects/egi-inspire/>) is a must, so the Work Package makes its work fully transparent to EGI.

WP4 is focusing on having a strong “organizational study” characterization with regional “customisation” applied to a shared model for sustainability. A collection of feedbacks will produce a final version with a road-map for the follow-up of potential extensions to the European Grid Infrastructure. The activities will be based on a series of cyclic processes that will involve in sequence the WP2 results, the meetings with EGI and that eventually will culminate with final reports produced as deliverables.

The timing of those relevant documents (March and September 2012) has been chosen in order to have, on one side, the EGI plans already stabilised and, on the other side, receive the results of WP2 work. The process is not just EGI driven, even when, of course, it will be cyclic and will imply brainstorming Meetings jointly organised with EGI as well as a series of steps that will be put in place in order to maximise the involvement of EGI-Inspire in the definition of the model that will be proposed by the organisational study.

To the date, CHAIN is strongly fostering the development of Africa ROC (<http://roc.africa-grid.org>), which will set the Continent much closer to the Grid deployment that already exists in other regions. As a consequence, Africa ROC GOCDB (<https://gocdb.chain-project.eu/>) is properly working and, beyond it, xGUS regional support system (<https://xgus.ggus.eu/eumedgrid>) has been adapted to CHAIN characteristics and requirements. This work has been of importance in order to shorten the digital divide of Africa with the better established Grid Initiatives in Latin America and Asia.

WP4 main activities are the following. ‘T4.1 Organisational study’ interacts mainly with WP2 and WP3 to study an organisational model that will allow the European Grid Infrastructure to interoperate and cooperate with external infrastructures. It also cooperates with EGI organisation and other Regional infrastructure in order to evaluate the opportunities offered by the evolution of the European Grid Infrastructure, assess the issues that will affect the interoperability with other infrastructures worldwide and evaluate the appropriate measures that will allow a smooth interoperability in the near future.

The second Task is ‘T4.2 Regional Operations Studies’, which deal with the specificities of the various regional operations and will address the specific issues related to interoperability with different middleware and migration to standards, as the ones proposed by GIN/OGF, in cooperation with the EGI planning. The activity involves a couple of different regional situations, i.e. developing regions with European based Middleware and well established regional e-Infrastructure

with different middleware. The Task will, in a second phase, deploy a pilot to demonstrate the applicability of the preliminary results of T4.1.

‘T4.3 Road-Map’ ought to prepare a roadmap for the follow-up of the project which will address the timeline of implementation, based on the work of T4.1 and T4.2. The road-map will be based on the following pillars:

- a detailed analysis of the changes that will have taken place occurred in the European Grid Infrastructure with the new organisational structure;
- the characteristics of the non-European e-Infrastructures, highlighting commonalities and differences including, but not limited to, organisation, HPC, Middleware, Security, Accounting, User Support;
- the opportunities offered by the present and emerging standards with recommendations on the support of new possible standards; and,
- the emerging paradigms of Virtualisation, Voluntary Computing and Cloud Computing.

The outcome of this work will be disseminated not only in Europe but also in the regions where CHAIN is focusing on and beyond. The main milestone of this project is to perform a organisational model about e-Infrastructures to be presented at the CHAIN workshops and any other events of interest. This is so since CHAIN will not provide an own computing infrastructure but those of its associated projects and partners (CLARA, EUIndiaGrid-2, Ubuntunet Alliance, etc.).

6 WP5 Dissemination and Outreach

The CHAIN project, as many other EC funded projects, proposes an intense series of dissemination activities which have as focal point a series of workshops and conferences that will address, respectively, the virtual research communities and the policy makers. It is coordinated by INFN, which is an asset due to its implication in many Regional Grid projects.

Nevertheless, its first step has been to develop and deploy its web-site, which has been migrated from its first Joomla! (<http://www.joomla.org/>) version to Liferay (<http://www.leosys.net/liferay-portal-development.aspx>). The web site was thus designed to achieve all the objectives originally foreseen in the Project’s Description of Work and it makes use of the already available tools such as Agenda, Document Repository, Video Conference, etc. The Liferay Technology allows to plug-in new tools profiled as ”portlets” that can be easily re-used in the web and better addresses the needs and expectations of the Grid communities and offers opportunities to better integrate grid tools. This will of course not only facilitate the maintenance and upgrading of the web site, but also will allow benefiting of portlets developed by third parties (e.g. other projects). The ever-evolving website can be found at <http://www.chain-project.eu>

CHAIN plans to attend and organise thematic workshops and high-level conferences to better address the Virtual Research Communities of relevance and exploit possible synergies with other international events, which is also its milestones. To

the date, it's been present in EGI events, both Technical and User Forums, in Open Grid Forum jointly organised with the International Symposium on Grid Computing and has supported the Conference on "Role of e-Infrastructures for Climate Change Research". This way, CHAIN project also supports the Climate Change community.

The main activities inside WP5 are 'T5.1 Involvement of scientific & technical communities', whose aim is to promote the usage of e-Infrastructures also in cooperation with other projects, 'T5.2 Solicitation of high-level policy awareness', that intends to support the regional communities towards the governments, stakeholders and policy makers getting consensus on the relevant aspects of the analysis of the issues and proposed solutions, and 'T5.3 Dissemination and outreach'.

7 Conclusions

The CHAIN project is a Coordination and Support Action which aims to coordinate and harmonise the Grid efforts currently developed world-wide as well as consolidate the Virtual Research Communities deployments in a wide area. Thus, it intends to address to main impact challenges.

The regions addressed by the CHAIN project are all considered strategic by the European Union, so the collaboration and cooperation with these non-EU International Co-operation Partner Countries (ICPC) is considered a high priority. The cooperation with other countries, which are not in the ICPC list, is also very relevant when a worldwide coordination is required to address common issues such as interoperability, interoperability and standardisation. Moreover, such a wider geographical coverage is mandatory to address the needs of international research communities which have, among their requirements, the transparent access to worldwide resources, laboratories, facilities and data repositories.

e-Infrastructures are indeed widely considered key enablers of scientific and social development. Their widespread usage is rapidly changing the landscape of science and represents one of the most effective answers to problems such as the digital divide and the brain drain. The creation and support of common worldwide e-Infrastructures devoted to research is thus doubly strategic regarding this aspect:

- It aims to speed-up the catch-up process of less developed countries giving them the opportunities to use cutting-edge European technologies and state-of-the-art computing and storage resources; and,
- It allows the extension of the European Research Area to other countries with the opportunities of assessing unique facilities and a large number of talented scientists from Europe as well as outside Europe who will actively and productively collaborate in challenging e-Science research activities.

The second relevant impact that the project aims to achieve is in the field of interoperations and interoperability. CHAIN is uniquely positioned for the large number of key players around the same table, not only for agreeing on the time of adoption of the present and emerging standards, but, more importantly, to promote new relevant ones based on shared policies and best practices. The above

impacts directly correspond with those stated within the Work Programme 2010 - Capacities (Work Programme 2010, Capacities, Part 1, Research Infrastructures, (European Commission C(2009)5905, July 29th 2009), Par. 1.3, p. 20)

Acknowledgements

CHAIN project would like to thank Dr. Alexandre M.J.J Bonvin (Universiteit Utrecht, We-NMR) and Dr. Antonio Cofiño (Universidad de Cantabria, WRF4G) for their support. CHAIN project is funded by the European Commission under its Seventh Framework Programme (FP7-INFRASTRUCTURES-2010-2, contract number 260011).